Language Documentation and Conservation in Europe

edited by
Vera Ferreira and Peter Bouda
Language Documentation and Conservation in Europe

*edited by*

Vera Ferreira and Peter Bouda
# Contents

**Contributors** iii  
**Foreword** ix  
1. Authenticity and linguistic variety among new speakers of Basque  
   *Jacqueline Urla, Estibaliz Amorrortu, Ane Ortega, and Jone Goirigolzarri*  
   1  
2. Lemko linguistic identity: Contested pluralities  
   *Michael Hornsby*  
   13  
3. New speakers of Minderico: Dynamics and tensions in the revitalization process  
   *Vera Ferreira*  
   26  
4. Kormakiti Arabic: A study of language decay and language death  
   *Ozan Gulle*  
   38  
5. Identity and language shift among Vlashki/Zheyanski speakers in Croatia  
   *Zvjezdana Vrzić and John Victor Singler*  
   51  
6. The sociolinguistic evaluation and recording of the dying Kursenieku language  
   *Dalia Kiseljūnaitė*  
   69  
7. Language Revitalization: The case of Judeo-Spanish varieties in Macedonia  
   *Esther Zarghooni-Hoffmann*  
   80  
8. El árabe ceutí, una lengua minorizada. Propuestas para su enseñanza en la escuela  
   *Francisco Moscoso García*  
   93  
9. Multilingualism and structural borrowing in Arbanasi Albanian  
   *Jana Willer-Gold, Tenes Gnjatović, Daniela Katunar, and Ranko Matasović*  
   103  
10. Reflections of an observant linguist regarding the orthography of A Fala  
    *Miroslav Valeš*  
    115
11. Language Landscape: Supporting community-led language documentation
   
   Sandy Ritchie, Samantha Goodchild, and Ebany Dohle

12. BaTelÈc: A text base for the Occitan language
   
   Myriam Bras and Marianne Vergez-Couret

13. The first Mirandese text-to-speech system
   
   José Pedro Ferreira, Cristiano Chesi, Daan Baldewijns, Daniela Braga, Miguel Dias, and Margarita Correia

14. Bridging divides: A proposal for integrating the teaching, research and revitalization of Nahuatl
   
   Justyna Olko and John Sullivan

15. Brief considerations about language policy: An European assessment
   
   Francisco Carvalho Vicente and Paulo Carvalho Vicente
Contributors

**Estibaliz Amorrortu** is a Professor in the Modern Languages and Basque Studies Department of the University of Deusto (Bilbao). She obtained her Ph.D. from the University of Southern California with a dissertation on attitudes towards different varieties of Basque, including the social acceptance of Basque standard Batua. She works in the field of Basque Sociolinguistics. She is co-author, together with Ane Ortega, of a study about the attitudes of non-Basque speakers toward the minority language, published in 2009. With the co-authors of the article in this volume, she is part of the NeoEusk project, which examines the attitudes, motivations, and identity of New Speakers of Basque.

**Daan Baldewijns** is computational linguist. He started at Lisbon’s Microsoft Language Development Center as linguist for Dutch, before taking on the role of Language Expert Lead.

**Daniela Braga** is senior speech scientist. She worked previously at Microsoft and Voicebox. Currently, she is the chief scientist of DefinedCrowd, a company she co-founded.

**Myriam Bras** is Professor at the Department of Language Sciences of the University of Toulouse. She studied both Computer Sciences and Linguistics. She obtained a Ph.D. on Natural Language Semantics in Computer Science and a Habilitation à diriger des recherches on Discourse Structures in Linguistics. Her research interests include Occitan linguistics. She is the principal investigator of the BaTelÒc Text Base, designed to provide linguists with written Occitan data. She is part of the RESTAURE ANR French project aiming at developing tools and resources for regional languages of France.

**Francisco Carvalho Vicente** is completing a Ph.D. in Communication Sciences at the Faculty of Social Sciences and Humanities of Universidade Nova de Lisboa. He has a degree in Information Sciences from Instituto Superior Miguel Torga (Coimbra). He is co-founder and researcher at CIDLeS and associate researcher at the Political Observatory. His main areas of research are media studies, political communication, and political marketing.

**Paulo Carvalho Vicente** is co-founder and researcher at CIDLeS, integrated researcher at OBSERVARE, researcher at Political Observatory, and collaborator at the Center of History and Culture of the Faculty of Social Sciences and Humanities of Universidade Nova de Lisboa. He has a Ph.D. in Political Science (2012) from Universidade Nova de Lisboa and a degree in History (2004) from University of Coimbra. His main areas of interest are institutions and political processes, European integration, and history and theory of International Relations.
Cristiano Chesi is Associate Professor at the University of Siena and researcher at the Center for Neurocognition and Theoretical Syntax of the Institute for Advanced Study of Pavia.

Margarita Correia is Auxiliary Professor at the University of Lisbon and senior researcher and member of the Board of CELGA-ILTEC R&D unit at the University of Coimbra.

Miguel Dias is Invited Associate Professor at ISCTE - University Institute of Lisbon and the head of the Microsoft Language Development Center in Lisbon.

Ebany Dohe is completing a Ph.D. in linguistics at SOAS, University of London, and is researching the semantic properties of ethnobotanical knowledge of the Pipil language of El Salvador. In the past she has worked with linguistic communities in El Salvador, Guatemala, Nepal, and the Philippines, researching policy for environmental sustainability, bilingual education, and cultural preservation. Other projects include PAW, an effort to encourage interdisciplinary research and collaboration, and the student-initiated Language Landscape, presented in this volume.

José Pedro Ferreira is researcher at CELGA-ILTEC, University of Coimbra. He is currently one of the coordinators of Common Portuguese Word-List, the official orthography-setting resource for Portuguese.

Vera Ferreira is chair of CIDLeS (Interdisciplinary Centre for Social and Language Documentation) and head of its Language Documentation and Typology group. She is lecturer for Language Documentation and endangered languages at the Ludwig-Maximilian-University in Munich. After taking the licentiate degree in English and German Studies at the University of Coimbra (Portugal) and getting an M.A. degree in General Linguistics and Linguistic Typology at the Ludwig-Maximilian-Universität in Munich (Germany), Ferreira has started her Ph.D. in General Linguistics at the Ludwig-Maximilian-Universität and specialized in Language Documentation and endangered languages in Europe. She is one of the Portuguese delegates of the COST Action ”New Speakers in a Multilingual Europe: Opportunities and Challenges” and is currently involved in several projects focused on the documentation, study, and revitalization of minority/endangered languages in Europe. Her main research interests include endangered languages (she is expert on Minderico), language documentation and revitalization, and language typology.

Tena Gnjatović received her Ph.D. from the University of Zagreb in 2014. She is currently working as translator.

Jone Goiri Golzarri Garaizar is a lecturer and researcher in the Social Work and Sociology Department of the University of Deusto (Bilbao). Her Ph.D. dissertation studied language policies and language ideologies of the Basque political parties (1980-2012). Her current research interests include language policy and sociolinguistics in minority language contexts. She is part of the NeoEusk project, together with the co-authors of the article in this volume, which examines the attitudes, motivations and identity of New Speakers of Basque.
Samantha Goodchild is one of the co-founders of Language Landscape (LL). She is co-director and outreach coordinator at LL. She is also currently a Ph.D. candidate at SOAS, University of London, researching multilingual practices in three villages in the Casamance, Senegal, as part of the “Crossroads” Project – “Investigating the unexplored side of multilingualism in Casamance”. She completed her M.A. in Language Documentation and Description at SOAS, University of London, focusing on the use of Mauritian Creole in the UK. Her research interests include multilingualism, sociolinguistics, and language maintenance and revitalization.

Ozan Gulle studied general linguistics, modern Greek and Oriental studies at Universität Leipzig, specializing in language contact. After working at Otto-Friedrich-Universität Bamberg for one term as research associate, he finished his Ph.D. titled “Structural Convergence in Cyprus” at Ludwig-Maxilimilian-Universität München in 2014.

Michael Hornsby is Visiting Professor at the Department of Celtic languages and literatures at Adam Mickiewicz University in Poznań, Poland. He studied French and Welsh at undergraduate level at Cardiff University and obtained his Ph.D. in sociolinguistics from Southampton University in 2009. He has worked on a number of language documentation projects, including “Fisherspeak” (University of Aberdeen, Scotland) and “Heritage Languages of Poland” (Adam Mickiewicz University in Poznań). His research interests include the sociolinguistic developments of minoritized languages under pressure from globalization and he has worked on such developments particularly among Breton speakers in France. He is member of the international COST network on “New Speakers in a Multilingual Europe: Opportunities and Challenges”. He is currently investigating a number of minoritized languages, including Lemko, in order to build up a picture of generalized sociolinguistic trends among speakers of such languages.

Daniela Katunar obtained her Ph.D. at the University of Zagreb in 2015. Her interests include construction grammar, lexical and computational semantics, and semantic typology. She is a research fellow at the University of Zagreb and a Fulbright Alumni from Princeton University.

Dalia Kiseliūnaitė has a Ph.D. in Humanities and is Professor at the Department of Baltic Philology of the Faculty of Humanities at Klaipėda University (Lithuania). Her research interests are Lithuanian and Latvian dialectology, comparative Baltic studies, onomastics, ethnolinguistics.

Ranko Matasović received his Ph.D. from the University of Zagreb in 1995. He is Professor of comparative linguistics at the University of Zagreb and member of the Croatian Academy of Sciences and Arts.

Francisco Moscoso García is Associate Professor of Classical Arabic and Moroccan Arabic in the Department of Arabic, Islamic and Oriental Studies at the Autonomous University of Madrid. He has worked also as Assistant Professor at the universities of Alicante and Cadix. His research interests cover the study of Arabic Dialects, teaching of Moroccan Arabic, and Oral Tradition in the Magreb. Aproximación al cuento narrado en
árabe marroquí (2012), Coplas de la región de Yebala (2014), B chuiya b chuiya (2014), and Diccionario de árabe marroquí (2015) are some of his most recent publications. He has also translated some collection of poems of Mourad Kadir and Ahmed Lemsyeh and two novels of Abdelghani Abou El Aazm.

JUSTyna Olko is researcher at the Faculty of “Artes Liberales” at the University of Warsaw (UW). She obtained a doctoral degree in the humanities in 2005 at the UW’s Faculty of History. She specializes in the ethnohistory, anthropology, and linguistics of pre-Hispanic and colonial Mesoamerica, with a special focus on Nahuat language and culture, Nahuatl linguistics, and issues of European-intercultural communication in a broad sense. In 2013, Justyna Olko has initiated a monolingual editorial series Totlahtol aimed at the restoration of literary culture in Nahuatl. She is also involved in programs for revitalizing Nahuatl and Wymysorys and works with researchers and activists committed to revitalizing dying languages of ethnic minorities in Mexico and Poland. She was the recipient of scholarships from Dumbarton Oaks (Harvard University) and the John Carter Brown Library as well as several major grants for team projects, including the Starting Grant of the European Research Council (2012). She has been awarded the Knight’s Cross of the Order of Polonia Restituta (2013) and a Burgen Fellowship by Academia Europaea (2013).

Ane Ortega is Professor in the Language Education Department of the “Begoñako Andra Mari” Teacher Training University College (Bilbao), where she works in language education in multilingual contexts such as the Basque one, where one of the curricular languages is a minoritized language in the process of revitalization and normalization. Her Ph.D. dissertation (London University) was on language choice and code-switching in bilingual family transmission. She is co-author, together with Esti Amorrortu, of a study about the attitudes of non-Basque speakers toward the minority language, published in 2009. With the co-authors of the article in this volume, she is part of the NeoEusk project, which examines the attitudes, motivations and identity of New Speakers of Basque.

Sandy Ritchie is a Postdoctoral Research Fellow at SOAS, University of London. He works on the AHRC-funded research project Prominent Possessors, which aims to investigate the phenomenon of “prominent internal possessors” from a theoretical and typological perspective. In his Ph.D. research at SOAS, he examined the morphosyntactic properties and discourse functions of prominent internal possessors in Chimane (or Tsimane’), an unclassified language of Amazonian Bolivia. He also helps to run Language Landscape, a language mapping website which aims to raise awareness of linguistic diversity.

John Victor Singler is Professor Emeritus of Linguistics at New York University. He holds an M.A. and Ph.D. in linguistics from the University of California at Los Angeles and an M.A. in African Area Studies from the University of London (SOAS). His interests in contact linguistics include pidgins and creoles, endangered languages; and bilingualism and bidialectalism. He is a specialist in variationist sociolinguistics. His interests extend to the history of African American English, African languages, and language in New York City. He is the author of An Introduction to Liberian English. With Silvia Kouwenberg, he co-edited The Handbook of Pidgin and Creole Studies.
JOHN SULLIVAN is Professor of Nahuatl language and culture at the Universidad Autónoma de Zacatecas, Visiting Professor at the University of Warsaw, director of the Instituto de docencia e investigación etnológica de Zacatecas, and director of the Yale Nahuatl Language Summer Program. He has a degree in elementary education from the Escuela Normal “Manuel Ávila Camacho” in Zacatecas and a Ph.D. in Literature from the University of California San Diego and was a Guggenheim fellow. He is currently working with indigenous researchers at IDIEZ in the preparation of a monolingual dictionary of Modern Huastecan Nahuatl and is a participating researcher in two major projects at the University of Warsaw. His professional activities are all designed to promote international participation in the teaching, research, and revitalization of Nahuatl language and culture.

JACQUELINE URLA is Professor and Chair of the Department of Anthropology at the University of Massachusetts Amherst, specializing in language ideology, the ethnography of resistance, and the place of knowledge in shifting modes of governance. She is the author of Reclaiming Basque: Language, Nation and Cultural Activism (University of Nevada Press 2012) and a member of the NeoEusk project, which examines the attitudes, motivations and identity of New Speakers of Basque.

MIROSLAV VALEŠ completed his Ph.D. in Spanish Philology at Palacký University, Olomouc, Czech Republic. Since 2007, he has been Head of the Department of Romance Languages at the Technical University of Liberec, Czech Republic. His research interests include sociolinguistics, minority languages, and contact linguistics, with a particular focus on Native American languages (Lakota, Shuar). He is involved in the process of standardization of A Fala (Spain). His most significant publication is Observaciones sociolingüísticas del español (2010).

MARIANNE VERGEZ-COURRET has a Ph.D. in linguistics. Her research interests include discourse semantics and computational linguistics. Since 2012, she has participated in a series of projects aimed at developing resources and tools for Occitan.

ZVJEZDANA VRZIĆ is Assistant Professor of Linguistics at the University of Rijeka and Adjunct Assistant Professor of Linguistics at New York University. She is the director of the Center for Language Research at the University of Rijeka. She has an M.A. and a Ph.D. in linguistics from New York University and an M.A. in Romance linguistics from University of Zagreb. Her research focuses on endangered Romance languages in Croatia, Istro-Romanian and Istriot; language documentation and description; the dynamics of language shift; and contact interference, especially in syntax. She is a Co-PI on the NSF-supported language documentation and description project, Documentation of the Vlashki/Zheyanski Language. She also leads two community-centered Croatian-based projects in language documentation and preservation.

JANA WILLER-GOLD received her Ph.D. from the University of Zagreb in 2013. She is research fellow in general linguistics at the University of Zagreb and postdoc in experimental morphosyntax at the University College London.

ESTHER ZARGHOONI-HOFFMANN is a multilingual active internationally in research, translation, and editing. Her expertise extends into the fields of Romance, Oriental, and
Slavic studies, and is based on academic experience in Germany, Portugal, and the Western Balkans. The research project upon which the paper in this volume is based has been ended without completion due to lack of access to material as well as eventually for want of interest in participation. She lives currently in the Republic of Macedonia, where she is engaged in social action and discourse, especially in the field of youth empowerment in Macedonia, Bulgaria, and beyond.
Foreword

Language diversity, its documentation and analysis, have always interested linguists around the world, especially those working on language typology, anthropological linguistics, and sociolinguistics. However, the beginning of language documentation as it is known today goes back to the 1990s. Several factors contributed to the emergence of this “new” linguistic discipline. First of all, technological developments which enabled the recording, processing, and storage of large amounts of linguistic data with high quality portable devices and fewer storage necessities (i.e. by more efficient codecs) opened up new perspectives and possibilities for the work in the field, in and with the language communities. On the other hand, the interest in linguistic diversity and more specifically in endangered languages spread beyond the academic world and became a public issue, mainly through the continuous reports on the subject published by the press and well-known institutions, such as the UNESCO with its Atlas of World’s Languages in Danger.

According to the literature on the domain of endangered languages, complemented with data of Ethnologue and the numbers revealed by UNESCO in 2009, there are around 6700 languages worldwide. 2500 of these languages, which are located mainly in Asia and Africa, followed by the Pacific and Americas, are endangered. Although language endangerment is not a new phenomenon, currently the process of extinction is increasing dramatically. As awareness of these circumstances and numbers (that increase as the linguistic research in different areas around the globe increases) is raising and the linguistic community realized the scientific and socio-cultural consequences of language endangerment, language documentation has established itself as an autonomous research subject within linguistics. As mentioned above, technological developments in recent years have massively improved the means of documentation (collection and storage) and the tools for linguistic data processing and archiving. Moreover, the need to standardize the study and documentation of endangered languages became a subject in academic discussions. Simultaneously, the increased media coverage also contributed to the rise of financial support for the documentation and research of undocumented or poorly documented languages, a fact that fostered research in this domain.

In this context, documentary linguistics emerged as a research field with the aim of developing a “lasting, multipurpose record of a language” (Himmelmann 2006: 1). It is important to refer that in this framework language is understood as much more than an abstract system of rules. Language is language use. It is multimodal, embedded in social, cultural, economic, political, etc. contexts and should be caught in its pragmatic effectiveness. To achieve a “multipurpose record of a language” one needs to apply a holistic perspective on language and document what is commonly known as communicative events, i.e. real situations of language use, the linguistic practices of a speech community. Apart from the interdisciplinary requirements of such a task, this work can and should only be designed and carried out in straight collaboration with the speech community.

It is exactly the collection, publication, distribution, and preservation of primary data of a variety of communicative events that emphasize the difference between documentary linguistics and descriptive linguistics. In this sense, primary data include not only notes (elicited or not) taken by linguists during the work with the language community, but also, and above all, audio and video recordings that reflect the multimodal character of language use in real life situations, as well as photos and text collections. The data are normally transcribed, translated, and they should be also annotated. This task requires linguistic annotations (morphosyntactic, semantic, pragmatic, and/or phonetic) as well as a broad range of non-linguistic annotations (anthropological, sociolinguistic, musical, botanic, gestural, etc.) whenever possible or if important to the language community being documented. Even if the researcher cannot make full annotations (mostly because it is not manageable in the limited timespan of language documentation projects and/or the financial resources available do not permit to build real interdisciplinary teams), the fact of making primary data available presents the advantage that researchers from the same or from other disciplines can use the data for their own purpose and complement it with their own annotations. And the communities, in turn, have real life data that they can use for revitalization purposes.

This new perspective on collecting, analyzing, and distributing linguistic data brought by documentary linguistics has proven to be a very important step towards interdisciplinary research in Humanities and towards the improvement of accountability of linguistic research results. On the other hand, the corpora emerging from language documentation activities represent also a challenge for the development of language technologies.

Nowadays, standard language technology tools (LT) only support major languages like English, Spanish or Chinese. A large number of people is thus not able to use their native language on computers and mobile devices, because hardware or software does not support input of those languages. The situation is even more critical in the case of endangered languages.

Considering that every successful technology can be used to teach, revitalize, and therefore boost the use of minority languages, it is important to rethink the development of LT in the domain of minority languages. This technology should also be able to assist the renewal of local languages and cultures by allowing people to actively study, teach, learn, extend, and spread their language in their community. Examples thereof are the papers by Bras & Vergez-Couret and Ferreira et al. in this volume.

Europe is in fact the continent with less linguistic diversity and the number of minority and endangered languages is reduced in comparison to other parts of the world. Subsequently, Europe is not in the focus of the researchers working on language documentation. Apart from some “major” minority languages in Europe (Catalan, Galician, Breton, Welsh, Basque, etc.), several of the European endangered languages are not known in detail (even in the academia) or documented in a concise and comprehensive way. Primary data on these languages, reflecting their everyday use, is almost non-existent. Moreover, the linguistic diversity in Europe is also unknown to the general public.

In this sense and in order to raise awareness of minority and endangered languages in Europe and to foster the dialog between researchers working on European endangered languages and on language documentation all over the world, CIDLeS – Interdisciplinary Centre for Social and Language Documentation (http://www.cidles.eu/) organized in October 2013 a two-day conference titled *Endangered Languages in Europe (ELE 2013)*. ELE 2013 aimed to provide an interdisciplinary forum in which scholars from language documentation, language technology, and experts on European endangered languages could
exchange ideas and techniques on language documentation, archiving, and revitalization; to further methodological discussions and collaborative research into linguistic diversity in Europe; and to reflect on language policy issues.

European endangered languages and language documentation (cf. the papers by Gulle for Kormakiti Arabic, Vrzić & Singler for Istro-Romanian, Kiseliūnaitė for Kursenieku, and Zarghooni-Hoffmann for Judeo-Spanish), language revitalization (cf. the papers by Olko & Sullivan for Nahuatl and Wymysiöeryš), language technologies (cf. the papers by Ritchie et al. on Language Landscape, Bras & Vergez-Couret for Occitan, Ferreira et al. for Mirandese), multilingualism (cf. the paper by Willer-Gold et al. for Arbanasi Albanian), and language policy (cf. the paper by Carvalho Vicente & Carvalho Ferreira) were some of the topics discussed during the conference. There were two special panels: One focused on endangered languages in the Iberian Peninsula (a region where language documentation as described above is almost unknown) (cf. the papers by Valeš for Fala and Moscoso for Arabic of Ceuta) and a round table dedicated to the topic of new speakers of minority/endangered languages (cf. the papers by Urla et al. for Basque, Ferreira for Minderico, and Hornsby for Lemko).

Besides collecting the papers presented during the conference, which cover a wide range of minority and endangered languages spoken in Europe, this volume represents the first attempt to reinforce the position of European linguistic diversity in the discussion about language endangerment, documentation and conservation.
Authenticity and Linguistic Variety among New Speakers of Basque

Jacqueline Urla\textsuperscript{a}, Estibaliz Amorrortu\textsuperscript{b}, Ane Ortega\textsuperscript{c}, and Jone Goirigolzarri\textsuperscript{b}

\textsuperscript{a}University of Massachusetts, \textsuperscript{b}University of Deusto, \textsuperscript{c}“Begoñako Andra Mari” Teacher Training College

This paper argues that the type of variety learned and used by Basque language learners is a key element in their self-perception as “true” or authentic speakers of Basque. Drawing on focus groups and individual interviews, we find that new speakers are for the most part strongly oriented towards the value of authenticity epitomized by local varieties. While new speakers report the utility of their mastery over the new standard Basque variety, they are not inclined to view this mastery as granting themselves greater authority or ownership over Basque. Rather they strongly valorize the informal and vernacular speech forms indexing colloquial speech and local dialect most identified with native speakers. The new speaker’s sociolinguistic context and motivations for learning Basque seem to be predictive of the strength of this orientation. The findings of this study point to the necessity of further study and documentation of local vernacular as well the urgency for language educators to find ways of incorporating the acquisition of local and dialectal features into language instruction.

1. INTRODUCTION. As we know, since the end of the Franco regime, we have seen impressive gains in staving off the decline of the regional languages in Spain’s historical peripheries, expanding public visibility and incorporating Galician, Basque and Catalan into the educational system, the media, and public administration. Increasing the number of speakers has been a goal in all three language projects. But as the numbers go up the question emerges: will these individuals actually be speakers who use and transmit their acquired language to the next generation? A lot hinges on a group of people we as yet know very little about.

Our project takes this up in the Basque Autonomous Community. In the past two years we have been conducting the first qualitative study of a cross section of new speakers of

\textsuperscript{1}Research leading to this article was made possible thanks to the funding provided by the Bizkailab initiative (Diputación Foral de Bizkaia and University of Deusto) to the project “Euskal hiztun berriak / Nuevos hablantes de euskera”. Funding was also provided by the Spanish Ministry of Economy and Competitivity (FFI2012-37884-C03-03) to the project “Nuevos hablantes de euskera a partir del modelo de inmersión: actitudes e identidad”. The authors would like to thank these institutions, as well as all members of the public who participated in the study. The ideas presented in this article draw on fruitful academic discussion over the issue of new speakers promoted by the European network COST IS1306 “New Speakers in a Multilingual Europe: Opportunities and Challenges”.

\textcopyright Licensed under Creative Commons Attribution Non-Commercial No Derivatives 4.0 International License ISBN: 978-0-9856211-5-5
Basque that are popularly referred to by the term *euskaldunberriak*. Using focus group discussions and individual interviews, our goal is to gain insight into their demographic profiles, self-perceptions, attitudes towards speaking Basque and language revival more generally.\(^2\)

New speakers are a sizeable and therefore critical element to consider in the future of Basque. Language shift towards Spanish was much more severe in the Basque Country than in either Catalonia or Galicia. At the time of Franco’s death in 1975, Basque language speakers in what is now the Basque Autonomous Community hovered at barely a quarter of the population, most of them located in rural areas and small coastal villages. Today, while they are still a minority, the number of Basque speakers is closer to 37.5% – all of whom, of course, are also speakers of Spanish.\(^3\)

Perhaps even more significant than the overall growth in numbers is the shift in the sociological contours of the Basque-speaking population. The once common stereotype of the Basque speaker as an elderly peasant, *baserritarra*, illiterate in Basque but with a rich command of oral tradition and vernacular dialect, has disappeared (Amorrortu et al. 2009). Today, thanks in large part to the popular support and institutional efforts for Basque medium education over the last forty years, Basque speakers are now encountered in both rural and urban settings and literacy in Basque (as well as Spanish) is much more common. In contrast to previous generations, he or she is also likely to be able to read and write competently in standardized Basque. Due to the generalized introduction of Basque into the educational system from the 1980s onward, we see a very noticeable increase in the proportion of young people who know Basque. As Figure 1 indicates, in the last twenty years the group of new speakers has increased considerably and now new speakers outnumber native speakers among youth in the 16-24 year old age group.

Our findings draw upon a preliminary analysis of nine focus groups and nine individual interviews with 63 new speakers between the ages of 18 and 56. We also draw upon an additional focus group with 9 native speakers aimed at assessing their attitudes towards new speakers.

This figure refers to people who describe themselves as being able to understand and speak Basque well. Another 17.3% of the population are “passive bilinguals”, that is individuals who describe themselves as being able to understand, but not speak Basque.

---

\(^2\) Our findings draw upon a preliminary analysis of nine focus groups and nine individual interviews with 63 new speakers between the ages of 18 and 56. We also draw upon an additional focus group with 9 native speakers aimed at assessing their attitudes towards new speakers.

\(^3\) This figure refers to people who describe themselves as being able to understand and speak Basque well. Another 17.3% of the population are “passive bilinguals”, that is individuals who describe themselves as being able to understand, but not speak Basque.
In this paper, we use the term “new speaker” to refer to those individuals who have learned Basque by means other than family transmission. In so doing, our goal is to set aside the ideological connotations and privileging of nativeness that terms like “non-native” or even “second language learner” tend to imply. At the same time, however, our project is very much interested in precisely the question of how the ideology of nativeness is experienced, reproduced or dismantled by new speakers.

We have been examining how new speakers situate themselves vis à vis the popular and widely used folk categories of linguistic identity: *euskaldun* (Basque speaker), *euskaldun zahar* (native speaker) and *euskaldunberri* (new Basque speaker).⁴ We found that new speakers’ categorizations of their own and other’s linguistic identity are strongly shaped by “mother tongue ideology” (Skutnabb-Kangas & Phillipson 1989), that is, they see the language of one’s socialization in the home as giving rise to a single life-long linguistic identity. They tended to also rank these three identity labels – *euskaldunberri*, *euskaldun* and *euskaldun zahar* – along an increasing gradient of authenticity, as shown in Figure 2. Even those who started immersion schooling at the age of 2 or 3, held to a view of native speakers as fundamentally different and more authentic than themselves.

Our study shows that learning the language, even to a high degree of competence, does not guarantee that a person will emerge thinking of him or herself as a true or authentic Basque speaker or be inclined to use it. This leads us to be cautionary, as many of our colleagues have noted, about taking the census figures as an accurate measure of active “speakers”. The way Basque new speakers talk about their identities also showed that, whilst these categories and identities seem to be permanent for some, in many other cases, and with the exception of *euskaldun zahar*, there exist pathways of authentication (Bucholtz 2003) that allow speakers to move on from the *euskaldunberri* category to *euskaldun*: access to Basque speaking networks and opportunities to use the language every day; the degree of competence, and access to the informal register and/or speaking a vernacular variety of Basque. Indeed, new speakers in our study underscored the difference between knowing a language and living in it. “Living in Basque”, that is, speaking it as a language of everyday life, was recognized as a pathway to developing a sense of identification as a Basque speaker. We found that individuals who had ample opportunities to use Basque in their everyday lives, particularly in their immediate social circles of family and friends, and had communicative competence in oral informal registers were those most likely to identify in an unmarked, unqualified way as Basque speakers, *euskaldun*.

In this paper, we explore the values of authenticity and authority in relationship to new speakers of Basque. By authenticity we mean attributions of being genuine, “real”, or credible speakers of a language. Authority has to do with attributions of respect and prestige to one’s way of speaking. We are particularly interested in the degree to which the variety or varieties of Basque spoken influences new speakers’ perceptions of themselves as legitimate speakers of Basque.

---

⁴ For further discussion see Ortega et al. 2015.
2. Language Revitalization and the Restructuring of Value. Pierre Bourdieu’s work (1991) is well known for describing the socio-economic structures that lead to the low prestige that is often accorded to minoritized languages or non-standard varieties. However, in approaching the question of value and legitimacy faced by linguistic communities in situations of language recovery, we have also found the theorizations of Kathryn Woolard (2008) to be highly useful. Woolard has argued that linguistic varieties acquire authority and are legitimated in terms of two competing axes or sets of qualities that she describes as anonymity and authenticity. The ideology of authenticity “locates the value of a language in its relationship to a particular community” (Woolard 2008: 2). In short, value can accrue to a language or speech variety when it is seen as the genuine voice of a particular group of people and place. Anonymity, by contrast, refers to an ideological constellation of values centering on indexical neutrality such that a language is seen to be purely referential, transcending any particular perspective, private interest, or distinctive social identity. Languages attributed this value are regarded as suitable for official use in the public sphere because they belong to everyone and no group in particular.

How these values come to be associated with languages is, as Woolard notes, very much conditioned by histories of power. State-supported dominant languages tend to have a monopoly over the value of anonymity. They are routinely presented as non-ethnic languages identified with citizenship, high culture and public life. Minority languages or vernacular varieties tend, on the other hand, to be limited to claiming attributes of authenticity. They are virtually never seen as suitable for generalized public or institutional use.

One way to understand the language revitalization movements that have been taking place in Spain is as mass-scale efforts to shift this structure of valuation. The preferred use of the term normalization, rather than revitalization, is telling. Minority language advocates see their project to be not simply to preserve or to destigmatize the language, but to enable and make normal the use of Catalan, Basque or Galician in official institutions and public life more generally.

In pursuing more favorable social attitudes and widespread uses for the minority language, language normalization initiates changes in habitual patterns of language use and value that can generate new sorts of tensions and insecurities among speakers about their ways of speaking (e.g. Dorian 1994, Wong 1999, Jaffe 1999, O’Rourke & Ramallo 2013). It is well recognized that corpus planning, sometimes called “modernization”, can further aggravate insecurities on the part of native speakers when confronted with new or standardized forms.5 As Michael Silverstein (1996) has argued, the presumption of the existence of a common set of shared norms is fundamental to Western notions of what it means to be a “real”, “modern” and “public” language. This creates a powerful motivation for linguistic minorities to standardize as they revitalize their language (see also Gal 2006).

In the Basque normalization project, a regularized and single set of norms was seen as indispensable for the new institutional roles Basque was to assume in education, publishing, media and administration. Standardization simultaneously operates as a synecdoche of the unity of the language and the nation. The norms for a Unified Basque, known as Euskara Batua, were developed by the Basque Language Academy in the course of the sixties and seventies and designed explicitly as an amalgam of different dialectal forms. It was intended by its creators to serve as the common written language of Basque speakers,

---

5 Language planners have long been aware of this, warning about the need to attend to gaining social acceptance for corpus reforms. Fishman (1983) famously argued that successful reforms required achieving a delicate balance of tradition and modernity.
occupying the roles that had exclusively belonged to Spanish.\textsuperscript{6} Intentionally de-racinated, values of anonymity and standard language ideology animated \textit{Batua}'s creation.

Although there has been no official policy mandating \textit{Batua} as the variety for official use, it was the form adopted for Basque-medium education and media, as well as public institutions of the Autonomous Community. \textit{Batua}'s creators emphasized the norms were intended primarily for writing, but it quickly became the \textit{de facto} "public" Basque voice. One of the outcomes of the confluence of standardization with these normalization policies is that individuals who learn Basque at school or at adult language academies are taught \textit{Batua}. Unless they have made a special effort or have access to native speaker social networks, new speakers tend to have minimal mastery of the local dialect that native speakers tend to use for everyday talk. We were curious to know if and under what circumstances this state of affairs produced problems for new speakers.

\textbf{3. ANONYMITY AND AUTHENTICITY IN NEW SPEAKER REPERTOIRES.} The socio-linguistic landscape in the Basque Autonomous Community is highly varied and modes of learning Basque have changed significantly in the last thirty years. In smaller towns of the interior or coastal villages, the majority of the population may be native speakers. In the plains of Araba or big urban centers of Bilbao and Vitoria-Gasteiz hearing Basque in the street may be quite uncommon. The sociolinguistic context has been shown to be a significant variable shaping experiences and attitudes towards Basque (Amorrortu et al. 2009).

The design of our research sought to capture this diversity by conducting focus groups from different kinds of sociolinguistic contexts. Similarly, we sought to incorporate new speakers from different age groups and in so doing represent the two main avenues for language acquisition: (1) those who learned Basque in adulthood at Basque language academies; and (2) those individuals, mostly under the age of 35–40, who learned Basque via immersion or partial immersion schooling. While people in the second profile were enrolled in a Basque school by their parents, individuals in the first profile generally made the decision to learn Basque for themselves out of a combination of integrative and instrumental motivations, that is, out of personal conviction (a desire to learn more about and feel integrated into Basque culture), and a sense it would enhance work opportunities.

One of our most notable findings is the strong adherence of new speakers to mother tongue ideology (Ortega et al. 2015). By this we mean that our study participants overwhelmingly felt that someone who learned and spoke Basque in the home possessed a deeper life-long identification with Basque than they felt they could claim. Our data is consistent with the reported widespread tendency of new speakers to perceive native speakers as embodying a more genuine identification with the language (Bucholtz 2003, O’Rourke & Ramallo 2011). With some notable exceptions, most of our new speakers (regardless of their method of learning Basque) said they continued to see themselves as Castilian-dominant because it was the language they “thought in” or habitually spoke.

In focus group discussions, some participants described the utility of knowing \textit{Batua} for obtaining certification to get certain jobs, for education, reading the newspaper and in general accessing the new markets opened to Basque following its officialization. However, it is interesting that while new speakers appreciated the instrumental value of literacy skills

\textsuperscript{6} Villasante (1980) is the key text written for popular audiences explaining the principles on which Euskara Batua is based and its intended use as a common or literary language. \textit{Batua}'s creators gave precedence to striking a balance between intelligibility and preserving distinctive linguistic features. See Urla (1993) for further discussion.
in standard Basque, they did not see this as giving them any particular authority in the language. *Batua* did not enjoy greater prestige as might be predicted by its status as a standard, and none of our study participants described *Batua* as being more correct or proper Basque. Quite the contrary, several new speakers recount being told that *Batua* was “not really Basque”. A sense of artificiality, of being a “book language” or “made up language”, continues to dog *Batua* as it does other minority language standard varieties (Dorian 1981, 1994, O’Rourke & Ramallo 2011).

Moreover, some new speakers expressed a certain envy of their peers who could speak the local variety. Several expressed a desire to be fluent in what some called “local Basque” [*bertako euskera*], or “everyday Basque” [*eguneroko euskera*]. We can see this perspective in the following statement by a young university student who lives in a highly Spanish-speaking area in Greater Bilbao and who has studied in Basque from early childhood:

(1) *Niri gustatuko litzайдake leku bateko euskeraz egitea, lagunekin egunero euskeraz egitea, ba lotsarik gabe euskeraz egitea.* (Gazte-BI-E, 317)

[I would like to] speak a Basque from somewhere, speak Basque everyday with my friends, speak in Basque without feeling embarrassed.

This feeling was more acute, as we might expect, in new speakers who lived in Basque-speaking areas and came into daily interaction with native speakers. The next two excerpts from individuals living in a highly Basque-speaking town show precisely these feelings: the first one (2) is by an older man and the second (3) by a woman whose Basque does in fact display a number of features of the local vernacular.

(2) *Bueno, gero ba kalekoa ikasi behar, klaro, ze hango euskeragaz hemen Arratien, ba bueno, txokantea da, ezta? Egia da, ikasten dozuna euskaltegietan gero erabiltzeko, ba bueno, esparru hau ez da igual egokiena, beraz, ba bueno, kalekoa, eta bueno, ba horretarako be nahiko laguntza nuen ba lagunengandik, ez?* (ZEA-D, 10)

Well, then you have to study street [Basque], of course, since here in Arratia, well, [*Batua*] is shocking, you know? The truth is that what you learn at the language school is not the right thing to use here, so, yes, [you need] the street [Basque] and for that I had a lot of help from my friends, you know?

(3) *Niri gustatuko jatan jakitea bertako euskalkia baina holan, ondo, esaerak eta dana. Niri horrek ematen doste penea ba holan ez dakidala. Nahiz eta ahalegindu, ta entzun, ta hau, ta bestea, baina gero ez dekot hori gauzea, hemen bertakoak iten daunen modua.* (ZEA-E, 151)

I would like to speak in the local dialect, be able to use its idioms and all that. I really feel sorry that I don’t. Even if I try, I try to pay attention, and all that. I just don’t have that, the way the locals speak here.

For new speakers who live in zones where native speakers are in abundance, acquiring mastery of informal and vernacular speech forms emblematic of native speakers was a particular point of pride. Being “mistaken” for a native speaker or receiving praise for being able to speak “our way” [*gure erara*] was a source of pleasure. Such mastery made these participants feel more entitled to call themselves “Basque speakers” rather than “new Basque speakers”. This is expressed in the following statement by a middle-aged school teacher from Bilbao who was posted to a school in the highly Basque-speaking area of Gernika:
New Basque speakers, who have learned Basque for integrative reasons, that is, for participating in Basque-speaking social networks and culture, find themselves in the following paradox: they have indeed learned “Basque” (i.e. *Batua*), but not the variety that is identified with belonging to those networks. They lack the variety invested with authenticity and therefore integrative value. This is a source of frustration for many, but not for all, as will be discussed next.

Further insights into the values attached to Basque varieties are gained through the analysis of Basque new speakers who learn the language purely for instrumental reasons. The following comment comes from one of our interviews with a university professor who said she learned Basque solely for work purposes and expressed no particular interest or desire to learn a vernacular variety:

(5) *Es que mi historia es otra, entonces yo creo que, o sea, poder enriquecer el euskera con elementos sí, pero no tengo que hacer como que soy otra persona, nunca voy a poder ser como esta profesora porque, o sea, esa forma de hablar que tiene, ni de escribir, que es súper de casa, es que mi casa es otra casa. Entonces, no pasa nada, la suya es así, y me resulta bonito, pero la mía es otra cosa, no tengo que hacer que soy otra persona. (...) Yo lo que quiero es hablar el euskera que hable, pero dar mis clases bien, comunicarme bien, poder comunicarme con la gente que habla. […] Pero no, yo no me apuntaría a un cursillo de, pues de otro dialecto, no. ¡Anda que no tengo que aprender cosas para hablar bien el batua!, como para meterme a otra cosa.* (ELK-A, 180, 182)

You see, my situation is a different one, so I think that, in other words, yes, I am happy to enrich my Basque, but I don’t have to become a different person. I will never be like that teacher because… that way she has of talking and writing is very home-like, but the thing is that my home is a different one. And that’s fine, hers is

---

7 This and other individual interviews were conducted in Spanish. Focus groups were conducted in Basque.
like that and I find it lovely, but mine is a different thing; I don’t have to behave like a different person. (…) What I really care about is to speak Basque, any Basque, teach my lessons well, communicate with those who speak it. (…) But I would not sign up for a course to learn a dialect, no I would not. As if I did not have enough to learn in order to speak Batua well, imagine getting into something else.

The above comment about not having to “become a different person” points indirectly to Batua’s qualities of anonymity in contrast to the social indexicality she perceives in vernacular. Speaking vernacular might not only be harder for her, but seems to signal for her an identification she does not feel or want to convey. In other words, the standard Batua suits this speaker just fine, and that is precisely because of its de-racination.

Further insights into how new speakers perceive the differences between standard and vernacular varieties of Basque can be obtained by examining speakers who have access to and use both. Individuals who had schooling in Batua and also competency in the local vernacular allocated different roles to each. They describe themselves as using Batua exclusively for writing, but preferring and feeling more comfortable expressing themselves orally in the local variety.

(6) Idatziz beti batua. Ba, bueno, nik unibertsidadean iten doaz aurkezpenak eta nik in dot nire euskeran, nik eske batua izen badot, ingo dot txarto. Ni hobeto, hobeto azaltan naz Arratieraz. (ZEA-B, 160)
I always write in Batua. Well, in the university, when I do [oral] presentations, I do them in my Basque; you see, if I do it in Batua, I will do it badly. I express myself better, better in Arratiera [the dialect of Arratia].

Interestingly, we found the same kind of division of labor described by youngest members of the focus group we conducted with native speakers:

(7) – Eta batua noiz erabiliko zeunkie zuek?(TO2-M1, 146)
When would you use Batua?
– Inoiz [ez].(TO2-B, 147)
Never.
– Nik batua..., ni batua irakurterako... […] Irakurterako eta idazterako, bestela... (TO2-E, 148, 150)
Batua I would use to read. . . […] Just to read and to write, otherwise . . .
– Hitz egiteko ez, sekulan ez? (TO2-M1, 151)
Never ever to speak?
– Ez, inoiz [ez].(TO2-B, 152)
No, never.

For those individuals who have competency in both dialect and standard, the division of labor described earlier that had once characterized the relationship between vernacular Basque and Spanish is now replicated between Batua and local vernacular. This would seem to point to some degree of success in achieving what the Basque Language Academy had originally set out to do: make a standard literary variety. Batua has had some success in becoming perceived as the unmarked variety for public written communication in Basque.
Schooling opportunities, language policies, campaigns, and standardization have made it possible for Basque to become at least partially de-ethnicized in the sense that it is not seen as a language known or spoken only by people who claim Basque ethnic ancestry. Through Batua, a variety of Basque has emerged that seems to operate as a de-racinated public language for use in particular, restricted, largely formal and institutional contexts. This variety has captured some values of anonymity. But when it comes to the sphere of informal everyday talk, what new speakers aspire to are values of authenticity more than anonymity.

4. CONCLUSION. The type of variety learned and spoken by new speakers has been identified as a key element in the self-perception of new speakers as legitimate speakers. In particular, the knowledge and use of the standard Batua and/or a local variety seems to be crucial for many new Basque speakers. Why this is so cannot be explained without understanding the process of Basque normalization and how the values attached to Basque varieties, including the standard Batua created in the 1960s, have evolved. A full historical contextualization, however, must be left for another occasion.

We have argued that language normalization can be understood as a process of attempting to rework the structure of values attributed to Basque and the majority language, Spanish. Rather than viewing the minority language as having less prestige than the dominant language, we follow Woolard’s suggestion that the value and respect for languages tend to be organized around competing notions of authenticity and anonymity, with non-normalized minority languages being primarily attributed values of authenticity while the majority language has exclusive claim to values of anonymity.

We have seen that a division of labor has emerged between Batua, standard Basque, and the vernacular dialect used for everyday oral communication by native speakers. Because new speakers are taught in the standardized variety, unless they live in Basque speaking areas and socialize in Basque speaking social networks, they do not tend to acquire fluency in the local Basque vernacular that native speakers prefer for everyday interaction. Batua is consequently identified as both the new speaker’s language and as the customary variety for writing and some speaking in formal institutions.

Our research showed that new speakers of Basque were on the whole predominantly oriented towards and valorized acquiring markers of the local vernacular. The new speaker’s sociolinguistic context seemed to be predictive of the strength of this orientation. It was stronger for those living in areas with high proportions of Basque speakers. We did encounter some new speakers who did not express any strong interest in vernacular. These were individuals who lived in predominantly Spanish speaking contexts and had learned Basque primarily for instrumental rather than integrative purposes.

New speakers did not consider that their knowledge of the amalgamated standard Basque made their way of speaking more correct, prestigious or gave them a claim to authority over the language. We might have expected them to claim such authority given the high value Western language ideologies place on literacy and standards. But it is not the case. While in the first years of the standardization process it was common to find Batua described as more “correct” than vernacular, that view has virtually disappeared and most new speakers, like native speakers, aspire to vernacular-like forms in everyday talk.

It would be reasonable to attribute this state of affairs to the as yet incomplete normalization of Basque in general. Spanish continues to function as the dominant language of the public sphere, the economy and the media-entertainment industry. Thus we might say that Batua fails to consolidate a position of authority and prestige over vernacular because
it occupies a very specific and limited niche in the limited public sphere of regional institutions. However, we think it is worth considering that Batua’s relationship to vernacular, the way authority and values are accorded in the Basque case, is not entirely a result of normalization “failure”. Rather it may also be a product of the particularly populist nature of the Basque language normalization movement that has incorporated a large degree of discussion and debate around standardization. A wide base of popular involvement and open debate is responsible for tempering the predilection toward seeing standards as inherently superior to vernacular. Over the years, there has been a steady effort on the part of grassroots language advocates to produce an understanding of standard and dialect as each having their own place and value. The division of labour we encountered between vernacular and Batua among bi-dialectal speakers clearly shows this.

In sum, at the present time of the evolution of Basque, and bearing in mind the increase in new speakers of Basque, the discussions in this paper have helped us understand better the influence that the spoken linguistic variety plays in the self-perception of the speakers as legitimate speakers of Basque. The strong orientation towards authenticity, as well as the importance of integrative motivations for Basque language learning, explain the high value attached to vernacular forms, which are seen as the desired variety for everyday interaction. This explains the dilemma and frustration of many new Basque speakers who, despite their efforts to learn the language, do not quite acquire a sense of legitimacy, due to not having learned the “right” variety.

The value the local variety has for many new speakers must be taken into account in language policy and initiatives to increase language use, but it must still be further explored. More needs to be learned about the profiles of speakers and situations that most require a knowledge of the local variety; the extent to which dialectal features incorporated onto the informal register may be sufficient for a number of new speakers, and, ultimately, we should learn more about the language features and varieties that speakers associate with the value of authenticity.

REFERENCES


Authenticity and linguistic variety among new speakers of Basque


Jacqueline Urla
jurla@anthro.umass.edu

Estibaliz Amorrotu
esti.amorrotu@deusto.es

Ane Ortega
aortega@bam.edu.es

Jone Goirigolzarri
jone.goirigolzarri@deusto.es
Lemko linguistic identity: Contested pluralities

Michael Hornsby
Adam Mickiewicz University in Poznań

In their efforts to organize as a recognized minority within the Polish state, the Lemkos have faced a number of obstacles, both internal and external to the community. This article explores three aspects of self-representation of the Lemko community - group membership, victimhood and “speakerhood” – and examines how these representations are contested on a number of levels.

1. INTRODUCTION. Poland, contrary to popular discourses both within and outside the country, is historically a multilingual and multiethnic state, comprising just under half a million citizens who identify as a member of a minority group, or 1.23% of the total population of Poland (Łodziński 2005: 94–95). These minorities include Silesians, Germans, Belorussians, Ukrainians, Roma, Lemkos, Lithuanians, Kashubs, Russians, Slovaks, Jews, Tatars, Czechs, Armenians, and Karaims (Jasiewicz 2011: 738). Other groups also exist, though lack official recognition, such as the inhabitants of Wilamowice, a town in the south of Poland where speakers of a distinct Germanic variety are still to be found, and which is undergoing revitalization efforts at a grassroots level (Hornsby, in press). Expressions of minority culture were severely hampered by the pre-1989 socialist regime in Poland, and after this date many minorities found new cultural and ideological spaces in which to signal their differences from the prevailing majority culture. Even before the demise of socialism, however, signs were emerging that the cultural homogeneity of modern Poland was less monolithic than it seemed to be. Since the fall of socialism, ethnic diversity (especially in Poland’s borderlands with Germany, Belarus, and Ukraine) has been more openly acknowledged (Hann 1998: 844). This article examines one such minority – the Lemkos – who have been able to benefit from the liberalization of a number of political and cultural spaces in Poland since 1989, and explores how expressions of identity, particularly linguistic identity, can be contested within their own community and more widely within Polish society as a whole. The creation of “new spaces” for minorities to express themselves results from the tensions which exist between “idealized” monolingual nation states and supranational structures and processes (Heller 1999: 339), further leading to the “uniformization of communicative practice” (Heller 2002: 8), which concurrently and conflictingly has “hybridity as a hallmark” (Heller 2000: 10).

2. HISTORICAL BACKGROUND AND CURRENT SITUATION. The Lemkos are a distinct ethnic group from the southeast region of Poland, traditionally from the northern slopes of the Carpathian Mountains (near the Ukrainian border), called Lemkovyna in Lemko, or Lemkowszczyzna in Polish. In the interwar period, both the Poles and the neighboring Ukrainians tried to get the Lemko to identify with Polish or Ukrainian nationalist
causes, but with little success. In 1944, Poland and the Soviet Union agreed upon a series of population transfers that saw Ukrainians, Belarusians, Russians and Rusyns (Lemkos) transferred to the Soviet Ukraine and Belarus. Although these transfers were supposedly voluntary, there was strong pressure to move. Near the end of the war, the Ukrainian Insurgent Army (UPA) sent some of its members to fight in the Lemko region. Whilst few of the Lemkos actively supported the UPA, the new Polish authorities did not see it that way and the Lemkos were perceived as Ukrainian sympathizers. Operation Vistula (Akcja Wisła) ensued as a result. In the spring and summer of 1947 the entire region was depopulated, and the Lemko population was resettled throughout the northern and western territories of Poland. This has resulted in large scale assimilation of the Lemko population into wider Polish society. Aside from the issue of assimilation, there are also major splits within the community, which pose difficult barriers to rebuilding a strong ethnic identity. The most contentious issue dividing the Lemkos is that of religion, one of the main tenets of their identity. While many of the Lemko are Orthodox, considerable numbers of them had converted to the Ukrainian Greek Catholic Church and this can often manifest in different identifications within the Lemko community, with some members identifying more as “Rusyn” (Orthodox) and others as “Ukrainian” (Greek Catholic). During the socialist era, this conflict was submerged due to the liquidation of the Ukrainian Greek Catholic Church, but it has now resurfaced in the post-1989 period (Laun 1999).

As a result of these divisions, it is difficult to enumerate how many Lemkos there are in Poland today. I have discussed elsewhere (Hornsby 2015) how reports on the number of Lemkos, and of Lemko speakers, can fluctuate between over 5,000 to 11,000 persons, and how classifying who exactly is Lemko can be problematic. This can be at least partly explained by the fact that many of the Lemkos, who are loyal to the Ukrainian Greek Catholic Church, are also pro-Ukrainian and identify themselves with the larger Ukrainian community. In contrast, the Orthodox members of Lemko society usually see their identity as being part of a larger Slavonic or Carpatho-Rusyn group and state their ethnicity as Lemko or Rusyn. As Magocsi notes, there is nothing new in a Lemko/Rusyn presence in the Carpathians, which “can be traced as far as the ninth and perhaps even the fifth and sixth centuries. What is new is their status as a nationality” (Magocsi 2009: 6, my emphasis). This article examines this comparatively new status, or recognition, from a number of points of representation: who counts as Lemko, how Lemko as “victim” counts as an identity marker and how the Lemko language confers not only “speakerhood” but also group membership.

3. REPRESENTATIONS. The languages of representation within the Lemko community are very often structured around master tropes such as inclusivity, victimization, and standardization. Inclusivity here implies who can be considered a Lemko and who cannot and which ethnonym (“Lemko”, “Rusyn” or indeed a hyphenated identity) is applicable. The theme of victimization implies a shared (recent) history and events in the 20th century which can be considered to have shaped current Lemko identities are examined below. As far as standardization is concerned, a common cultural heritage among Lemkos appears to be shaped in different “hybrid” formats which nevertheless often conform to transnational trends. However, despite their global applicability, languages of representation do not necessarily imply commensurability at the local level. Thus local conditions sometimes produce very different results for this post-Communist, newly-emergent minority than it does for other minorities in Western Europe, despite the common rhetoric shared with the same minorities. These languages of representation are examined in turn and are
then examined in terms of acts of self-identification among the Lemkos who participated in the research for the present article.

3.1. Inclusivity: Who is in, who is out. Lemkos are sometimes described as “Rusyns” from a political perspective. The Rusyn national movement which emerged in post-1989 central and eastern Europe uses the term “Rusyn” to encompass all East Slavs in the Transcarpathian areas of Poland, Slovakia, Ukraine, and further afield among self-identified groups in Hungary, Romania, Serbia, and Croatia, and with an emerging ethnic awareness among groups in the Czech Republic and Moldova (Baptie 2011: 9-10). The political movement includes those who identify with the ethnonym “Lemko”. The term “Ruthenian” is occasionally encountered principally in non-specialist writings on present-day Rusyn (for example, in English language reports of the Euromosaic program. There are, then, different narratives present (both within the Lemko community and also outside it) over the question of who is a Lemko. As I have noted elsewhere (Hornsby 2015), the increasing tendency among the Lemkos is to identify as Lemko tout court, but with a minority opting for a Rusyn or a Ukrainian identity, but increasingly with a hybrid sense of being Lemko, as in the term “Polish Lemko”.

However, we should again be wary of taking census returns and the results of surveys at face value. During fieldwork1 undertaken in 2012 and 2013, participants talked in much more flexible terms about how they had to make a choice of their identity as Lemkos, and the tendency appeared to be toward a hyphenated identity. However, some hyphenations appear to be more acceptable than others. The current diasporic nature of much of the Lemko population in Poland lends itself to hybridity out of necessity, and research participants reveal varying attitudes to the changing nature of what it means to be a Lemko in 21st century Poland:

(1) Extract 1

*It is all mixed up in our community. This means that all of us, even within one family, depending on individual circumstances, and the direction they take, can feel either more or less Ukrainian or Lemko. I, for example, have a Ukrainian husband from Pidlashsha. I obtained a degree in the Ukrainian language, I know Ukrainian history, I feel a connection with Ukraine. I have a lot of family there. I can therefore say that I feel Lemko above all. But what sort of Lemko? I would say I am a Ukrainian Lemko.* [Research participant 1 (RP1). My translation.]

Thus, marriage seems to be the deciding factor in this particular case – a Ukrainian husband means that RP1, born in Poland, feels more in tune with a Ukrainian orientation and describes herself as a Ukrainian Lemko. Marriage is not always an indicator of ethnic orientation, however, and another research participant (RP2), the sister of RP1, echoes the sentiment that “it is all mixed up” since her own husband’s ethnicity is not a decisive factor in the expression of her individual identity:

---

1 Lemko speakers were engaged in “ethnographic conversations” and semi-structured interviews in western and south-eastern Poland in three separate stages: 1. September 2012 (5 participants, in the age range 20–70; 2 F, 3 M; 4 born in Poland, 1 in the Ukraine; 2 students, 2 Orthodox priests, 1 retired); 2. March 2013 (2 participants, F, 40s, born in Poland, civil servants); and 3. July 2013 (4 participants, 3 late teens [2F, 1M], 1 in his 40s, all born in Poland, 3 students, 1 cultural activist). All interviews were conducted in Ukrainian, with the respondents given the option of replying in Lemko, which they all did. The data used for the present article were drawn from a selected number of research participants, as detailed in the appendix.
(2) Extract 2

I have a Polish husband and we have a child, and we had our son christened in the Orthodox Church. We were married in the Orthodox Church and we attend services [there]. Tradition is the most important thing for me . . . I would say that I am above all a Lemko. I wouldn’t say I am Polish even though I have a Polish husband. (RP2. My translation.)

There therefore appears to be some boundaries in place which prevent Lemko identity being too fluid. Tradition in the form of religious adherence attracts more allegiance than does any sense of conforming to the majority (and her husband’s) identity. As RP1 pointed out during the interview:

(3) Extract 3

My sister has a Polish husband but she wouldn’t say she is Polish. I for example have a Ukrainian husband [and] I can identify as a Ukrainian Lemko. This means that we Lemkos feel closer to Ukrainians than to Poles. (RP1. My translation)

Something prevents Lemkos identifying as Polish Lemkos, it would appear, at least in the narrative of these two participants. This is despite the census returns indicating some 36% of the 11,000 self-identified Lemkos in Poland considering themselves to be Polish Lemko. According to RP1, this is something she has never heard mentioned in her own milieu:

(4) Extract 4

In the environment in which I function among Lemkos in the mountain regions, I don’t know a single person who would describe themselves as “Polish Lemko”. I don’t know if there has been any sociological research which looked at Lower Silesia. Perhaps they say “Polish Lemko” [there]. There might have been greater assimilation there (…) Among the group where I spent my childhood and my teenage years, in the traditional Lemko area, I don’t know a single person who would say that. There were either Lemko-Lemkos or Ukrainian-Lemkos. (RP1. My translation)

So why the apparent increase in this hybrid identity, despite its contested nature? As Omoniyi (2006: 20) points out, “one’s identity isn’t simply chosen from an array of possibilities over the others which are discarded; there is on the contrary a cluster of co-present identities but with varying degrees of salience. The latter depends on the most preferred presentation of self in a given moment”. Thus, when faced with an official document such as a census, Lemkos face a “moment of identification” (Omoniyi (2006: 21) when “verbal and non-verbal communicative codes (…) are deployed to flag up an image of self or perspectives of it”. Given the history of the stigmatization – on all sides – of Lemkos after World War II (“Many Poles made no distinction between Ukrainians and Lemkos, regarding both as traitors and fascist collaborators. Furthermore, being Lemko brought an individual into conflict with the Ukrainian community, whose leadership, like that of Warsaw, assumed the viewpoint that Lemkos are Ukrainians and should identify themselves as such”, Mihalasky 2009: 66), the unease at public declarations of Lemko identification will still be in the memory of some Lemkos and will have been transmitted in part to the younger generation. Identifying with Polish-Lemko hybridity will presumably go some way to assuaging this sense of unease.
3.2. VICTIMIZATION. A sense of unease does indeed pervade the Lemko community and the Lemkos as “victims” is a trope that emerges both within and from outside the speech community. The defining moment for many Lemkos is “Operation Vistula” (Akcja Wisła) which was carried out in 1947 as an act of reprisal, or so the Lemkos claim, against Ukrainian nationalists (the Ukrainian Insurgent Army) who had collaborated with the Nazis in an attempt to secure independence for the Ukraine during World War II. Victimization as “a social object” is part of a wider lexicon that is “learned, used and represented” and is implemented by diverse actors when articulating their claims (Jeffery & Candea 2006: 287). As Ballinger (2003: 129–30) points out, “the end of the Cold War has revitalized and transformed the language of victimhood and ‘genocide’ (…) The powerful moral capital attached to the charge of genocide has nonetheless paradoxically made for the term’s increasingly broad application by groups (…) claiming past persecution.” In the case of the Lemkos, commentators (e.g. Snyder 1999, Subtelny 2001, i.a.) have described the deportations of “Operation Vistula” as “ethnic cleansing”. Weight is sometimes added to the claim by references to the use of Jaworzno (a sub-camp of the Auschwitz complex) for the detention of civilians during the Operation Vistula deportation campaign. As Trzeszczyńska (2013: 441) has pointed out, the Lemkos’ “collective memory is filled with references to the end of WW2 and [the] post-war period which redefined the meaning of being (…) Lemko. I consider their commemorative efforts as [a] peculiar cultural practice which demands [a] deep anthropological analysis.”

For many Lemkos, any accusation of “collaboration” and the collective punishment visited upon them by the Communist authorities was disproportionate to their actual involvement. Current narratives about this event are infused, not unreasonably, with a sense of injustice and the momentum for public recognition of this injustice seems to be gathering pace. When a joint declaration was issued in 2007 by the presidents of Poland (Kaczyński) and of the Ukraine (Yushchenko) condemning the above operation as a violation of human rights,2 the blame was firmly laid at the door of a “totalitarian communist regime” (winowiąc tej operacji był totalitarny reżim komunistyczny).3 It is interesting to note, however, that in this statement issued by the two countries, there is no mention of Lemkos as co-victims with the Ukrainians who were deported, thus seemingly continuing the Communist regime’s practice of conflating Lemkos with Ukrainians (see Mihalasky 2009: 66, cited above).

Transnationally, there are clear signs that the recognition of this act is illegitimate and therefore constitutes victimization. Recent publications sustain the victimization narrative, not just in Poland and the Ukraine, but further afield, such as the USA. The Polish researcher Trzeszczyńska (2013) notes that “the displacements of the Lemko population are undoubtedly the most important turning point which define the post-war history and the modern fate of the members of the group. The Lemkos themselves, together with historians, treat 1944–1947 as a moment in history that forever changed them as a community” (Trzeszczyńska 2013: 76, my translation). The American researcher Reilly (2013) documents, more individualistically, “a personal narrative, but at the same time it represents the experiences of many families from southeastern Poland who were considered to be part of the Ukrainian minority” (Reilly 2013: 8). And the Ukrainian author Shcherba (2011) writes of the “harsh deportation of Ukrainians from their native ethnic lands (…) [and] the terrible atrocities which were suffered by Ukrainians during deportation” (Shcherba 2011: 9, my translation). Note that the transnational nature of the recognition of vic-

---

timization does not imply any sort of commonality – the Polish researcher attempts to distance herself from the events through her stated research aim of “the impact of [an] anthropologist who tries to investigate the traumatic past of an ethnic group which suffered [at the hands of] the nation which the researcher represents” (Trzeszczyńska (2013: 443); the Ukrainian writer adopts a nationalistic stance in upholding the cause of the Ukrainian Insurgent Army as the “only (…) political force [which] acted on behalf of genuine national liberation of the Ukrainian nation and the integrity of its lands in an independent state” (Shcherba 2011: 9, my translation); and Reilly (2013), from an American and personal perspective, recounts her own family history, since one of her research participants was her own maternal grandmother. Thus, “victimization” as a recurrent trope connected to the Lemkos is played out on very different levels and means very different things for both members of the Lemko community and outside commentators. If, as Trzeszczyńska (2013: 443) considers, there is a “heterogeneous nature [to the] collective memory of the Lemkos”, this appears to be the case among commentators outside the community as well.

However, this sense of unease does not remain at just a historical, community-wide level. As Majewicz & Wicherkiewicz (1998) note, attitudes of Poles towards Ukrainians and Lemkos was often hostile after the deportations, with children of Ukrainian/Lemko origin being harassed in schools, and the status of the Ukrainian/Lemko language(s) being considered lower than that of Polish even in places inhabited predominantly by Ukrainians/Lemkos. Fieldwork has revealed that adult participants had experienced such attitudes and less than favorable treatment in their childhood. RP1 reports that when growing up in a traditional Lemko area, her school would schedule very important tests, or recreational excursions on major Orthodox holidays, which would cause considerable inconvenience for herself and her family. Such inconvenience can, of course, be viewed subjectively. Interestingly, however, for the first time in 2013, Polish universities issued declarations that teachers should exercise tolerance and flexibility with students wishing to celebrate Orthodox Christmas on 7th January (suggesting that such flexibility was not always apparent in the past).4

The Lemko language has also come under scrutiny from public officials, revealing a discourse that seeks to delegitimize its presence in public spaces. In December 2012, the town library in Gorlice (near the traditional Lemko-speaking area) issued a greeting card wishing the inhabitants of the town a Merry Christmas and a Happy New Year in Polish and in Lemko. This was commented upon by local councilor Augustyn Mróz (Chairman of the Board of Education and Culture), who questioned the “legality” of including a text in Lemko, since Polish was the official language of the Polish state. Furthermore, he questioned the use of Lemko in preference to “English, French (…) or I don’t know, why not Hebrew?” (http://www.gorlice24.pl (10 December, 2015), my translation). The subtext seems to indicate that Lemko in Gorlice is as foreign – and thus has no place – as English, French or Hebrew (notwithstanding the historic presence of a continuous Jewish community in Gorlice for centuries until the Second World War, let alone the inappropriateness of a Christian festival being commemorated in greeting cards in a Jewish liturgical language). This “othering” of the Lemko language, and consequently of Lemko themselves, seems to indicate a prevalent lack of comfort with difference and diversity since as

---

4 For example, one faculty at Warsaw University issued the following statement: “In accordance with the letter of the Dean no. WLS-44/54/2012 of 14 December 2012, January 7, 2013 is a day off from classes due to Orthodox Christmas” (Zgodnie z pismem Dziekana nr WLS-44/54/2012 z 14 grudnia 2012 r. dzień 7 stycznia 2013 r. jest dniem wolnym od zajęć w związku z prawośawnymi Świętami Bożego Narodzenia).
Janicki has pointed out [an] “essentialist mindset is still widely present in Polish society” (Janicki 2003: 273).5

Interestingly, RP1 made the point, during the interview, of comparing the situation of Lemkos with other, equally less favorably regarded minorities in Poland, such as the Jews. She noted that she had to be cautious over who she was open with about being Lemko, comparing her situation to Poles who had openly identified as Jewish, who then found anti-Semitic graffiti on their exterior walls or a rock thrown through their window. This also reminded her of talking Lemko to her grandmother in public, and her grandmother asking her not to, recalling, for her, past times when using Lemko in the public domain could provoke open hostility. Thus, the language of “victimhood” as employed by Lemkos is commensurable at certain local levels (perhaps because of a certain sense of solidarity with particular minorities) but not at others (such as the inappropriately, in Lemko eyes, conflation of Lemko culture with Ukrainian national aspirations).

3.3. STANDARDIZATION: BUILDING “A” LEMKO LANGUAGE. What language(s) do Lemkos speak? It is noteworthy that the titles of grammars and related items refer not to “variants” but to “language” in those areas/countries where attempts at standardization of the Rusyn/Lemko language have been undertaken. It is possible therefore to analyze Rusyn as a single pluricentric language, or a collection of languages. Baptie (2011: 8) has documented the official names for each Rusyn variant which are: русинський язик (rusín’skij jazyk, “Rusyn language”) in Slovakia; лемківський язик (lemkivskij jazyk, “Lemko language”) in Poland; русинський язик (rusíns’kyj jazyk, “Rusyn language”) in one scheme proposed for Transcarpathian Ukraine and руски єзик (ruski jazik, “Rusyn language”) in Serbia and Croatia (Vojvodina and Srem).

Even the very attempt to standardize the language appears to cause tension. It is first of all recognized by some speakers that there is a need for a standard Lemko language:

(5) Extract 5

_It would be better if they spoke standard Lemko, but even here in the parish, there’s no “standard language” and some people came from Brzoza and they now live here in the parish in Ługi and they speak in their own way and some of their words are their own individual words and that’s how they speak. But the people who came from the east part of the Lemko territory, they use other words, they have different names for some things, they build their sentences differently – there’s simply a difference._ (RP3. My translation)

However, when attempts are put in place to produce speakers of standard Lemko, this can be seen as “unnatural”:

---

5 It should be pointed out that Councilor Augustyn Mróz was relieved of his position as Chairman of the Board of Education and Culture on 25\textsuperscript{th} January 2013 (http://www.gorlice24.pl/, 10 December, 2015).
(6) Extract 6

We know of a young woman who studied Lemko in Kraków and she got married to a man from [the traditional Lemko area] and [I was told that] she would correct her husband’s family’s language, that of his parents, of his grandparents, [telling them], “You speak [Lemko] badly.” Because she knew better, she had studied at university. Because she knows the “new” Lemko speech. And they, the grandparents, don’t know it. It’s funny because you can’t create an artificial language at the university or in a small circle. Language develops naturally and we should let it do so. (RP1. My translation)

Thus, what we are dealing with here is an over-arching ideology of standard language which is becoming evident amongst younger speakers and which was previously less evident among the Lemko speech community. I would argue that this ideology is borrowed from the majority community, where Polish as a standard language is heavily promoted, due to historical attempts to eradicate the language and culture but also, as previously noted, Janicki’s assessment of the Polish mindset as “essentialist” (2003: 273). Language change and standardization is particularly fraught among minority languages, since the speakers are aware of its endangered situation and that attempts in revitalization efforts to make a language uniform for the purposes of teaching and learning can in fact been perceived as breaking down the last vestiges of tradition. One research participant sees the process of change in Lemko as a loss:

(7) Extract 7

Our grandmothers speak lovely Lemko, very nice. One of them has lived in the same place all her life, without shops, without schools, without television, without radio, she speaks a variety (of Lemko) as it was always spoken. When the second grandmother speaks Polish, you can hear that she’s not Polish (…) when we listen to our grandmothers speak, the Lemkos have [the sound] /y/ (…) our parents know this sound, but I don’t hear them use it. And we don’t use it at all. (RP1. My translation.)

4. Cool to be Lemko? Lemko “authentication”.

Culture is becoming increasingly a consumer commodity, to be bought, sold, and exchanged in the new economy. Annual Lemko folk and cultural festivals called Vatra (the Hearth) have been held in a different Carpathian village every year. Thousands from both Poland and Ukraine have returned to their “homeland” in celebration of these festivals and the revived culture has sparked the curiosity of younger Lemko generations wanting to learn about their ancestral heritage. But it has to be modern, if it is to be attractive to younger generations. As Maher (2005: 195) claims: “Cultural essentialism and ethnic orthodoxy are out (…) Metroethnicity is in”. And to be modern, it has to be “metroethnic” or hybrid, as I understand it in a Lemko context. Tropes identified earlier in this article are apparent in acts of Lemko “authentication” and are discussed below, particularly in terms of where such tropes are intermingled with hybrid practices associated with “metroethnicity”.

Lemko festivals aimed at the preservation of the culture are engaged in creating the culture in new forms, forms which are acceptable to the younger generations of Lemkos. Bands such as Lemko Tower and Lemon prove popular among young Lemkos as a result, but while they are based on traditional Lemko music, they also exhibit “glocalized” features
which mark them as different from traditional forms, but which are shared more generally with other musical forms. For example, Lemko Tower is described, in its own publicity, as a band that was formed “to respond to a need in young Lemkos’ hearts, whose grandparents had been forced to western Poland because of the Vistula Action” (Lemko Tower 2012, my translation). Having thus established its credentials as “authentic”, the band plays the hybrid card: “LT gives you music that will not quite remind you about our grandparents but will show you the happiness and energy of young Lemkos who have returned to their roots” (Lemko Tower 2012, my translation). “Not quite remind” suggests that they consider their own music to be progressive, familiar, and yet different at the same time, following wider trends of modernization in folk music elsewhere and linked in with the growing popularity of world music. In a similar way, the text which accompanies one of the CDs produced by the folk ensemble Kyczera states, “Young people in Kyczera sing the songs as they feel and understand them in their own way. Perhaps they sing differently from their forebears but it’s hard to sing in the same way so far away from their beloved mountains” (Kyczera 2011, my translation). The (dis)continuity discourse is thus once again emphasized in metroethic terms, in that these and other cultural practices emerge from specific, local contexts of interaction, while accommodating both fixity and fluidity in output (Otsuji & Pennycook 2009).

The festival Ñîò ïîä Êè÷åðîì (“The world welcomed by Kyczera”) has been organized for the past eighteen years by Lemkos interested in establishing networks with other minorities elsewhere in the world, and it takes place annually in the western Polish town of Legnica, one of the destinations for deportees after the Vistula Operation. This appears to be part of a trend of solidarity with other minorities, of learning from other minorities how to be a minority, but also in learning how to take pride in your local culture in a glocalized fashion. Lemkos’ songs and dances can thus take their place on stage next to performers from around the world and in the words of the festival organizer, Jerzy S, the world will “get to know Lemko culture so that it can be recognized as one of the cultures of the world” (http://www.malopolskie.tv (10 December, 2015), my translation). Such festivals build on the tropes discussed earlier, in that both inclusivity and victimization are part of the festivals’ discourse. Raising the profile of Lemkos in the world in order to be recognized and counted as a minority in their own right appears to be one aim. Another aim, which arose when interviewing the organizer individually, appears to be the refusal of victimization:

When I studied in the Ukraine, I was told by Ukrainians: “Don’t expect us or anyone else to save Lemko for you – only you Lemkos can save your language, your culture, your dances and songs. You will never dance Ukrainian dances better than Ukrainians and we Ukrainians will not dance Lemko dances better than you Lemkos.” I have kept this in mind since my student days in Drohobych. We should preserve our parents’ heritage here in Western Poland. (Jerzy Starzyński, RP4. My translation).

The Lemko community is thus still in the process of adjusting and adapting firstly to post-war and then to post-Communist conditions through discursive means of constructing post-modern identities. Constructing a “minority” identity which mirrors other minority identities elsewhere, finding a voice which recognizes a past where Lemkos were victims but is beginning to reject victimhood in the 21st century, where inclusivity can simultaneously be encompassing or rejecting, with a whole myriad of combinations and permutations in-between, dependent on individual responses, indicates that the main tech-
nique for this construction of identity is largely based on authentication, that is the ways individual assert themselves as “real”, in comparison to other individuals or groups (Bucholtz & Hall (2007). Being “real” thus means having one foot in the past and the other in the present – a passing nod to essentialism is inherent in the discourse of what it means to be a Lemko (with references to ancestry, a homeland and “authentic” (i.e. non-standard) language). Thus, apparent subject (or self-) positioning appears to be an important component of the identity negotiation that takes place among some Lemkos in an effort to negotiate the “self” and the status quo (Doosje et a. 2002, Oakes 2001, Rosenblum & Travis 2006, Schmitt et al. 2003, Turner 1999, Wallace 2001). Identifying as “Lemko” can lead Rusyns in Poland to self-define as members of their heritage language and culture group. On the other, they can also, simultaneously, consider themselves different from their heritage language community and position themselves within the mainstream culture.

Different research participants interviewed for the present article displayed affiliations and different connections to their heritage community and Wallace (2001) describes several types of subject positioning based on self-identification within heritage and mainstream cultures. Of these, two are most applicable to the Lemko situation, I would argue. The majority appeared (RP1, 2, 3, and 4 for example) to fit within his home base/visitor’s base model, where minority language speakers consider one culture (mainstream or that of the minority language) as the home base in which they are most comfortable operating. The other culture becomes a frequently visited environment in which the attachment to cultural practices, including language, is not as strong as in the home base cultural environment. Most significantly for the younger research participant interviewed for the present article, the life on the border model, where minority language speakers position themselves on the edge of the two cultures, sometimes creating a border culture, appears to be the one he most engaged in. This is a challenging process of identity negotiation, since such speakers “exist” on the edge (of the community and of standardized language practices) and are required to engage in, or feel the need for, a balancing act. One of the ways this is manifest, though often not concurrently within the same sub-group or individual, is by a focus on “creative linguistic conditions across space and borders of culture, history and politics” (Otsuji & Pennycook 2010: 244). “Authentic” language can be discursively indexed, by RP1 and others, as non-standard, dialectal Lemko as spoken by the previous generation, but a “hybrid” variety is also seen as representing the current (and possibly the future, but contested) form of Lemko among younger people:

(8) Extract 8

All the time I hear polonized Lemko from my friends and the people around me (. . .) My grandparents correct me frequently because I am not very good at Lemko. When I speak to my grandfather I have to think before I say anything in order for him not to correct me! (RP5. My translation).

The Lemko language does appear to be changing, at least to this research participant, but he appears unhappy with the changes, since he also reported, “I want to speak the same way as my grandfather spoke 50 years ago.” He displays an ideology of “historical authenticity” in that anything from the past must be acceptable whereas present linguistic practices are not, such as the polonized, hybrid variety he mentions. The implications of such a hybrid variety lie outside the scope of the present article, but do indicate an area of sociolinguistic research that needs to be addressed not just within the Lemko community but further afield among minorities elsewhere. Whether such metrolingual practices are seen as “cool” or conversely as “inauthentic” by members of particular speech communities
is a pressing question for minority language policy makers, activists, speakers and/or users alike. See Hornsby (2015) for more details on this phenomenon.

5. Conclusion – Local manifestations of global representations. This article has aimed to explore a number of global languages of representation within the Lemko community in Poland today. Inclusivity, victimization and standardization are themes that are to be found and are well represented among minorities elsewhere but being inclusive in a Lemko sense can mean something very different to being inclusive in another minority language community. The Lemko community has fault-lines running through it in ways that are not imaginable in some other minorities – being a Lemko speaker does not necessarily imply a cohesive identity based on religion, territory or even the very label used to identify the language that is spoken, but rather a multiplicity of identities that can contradict and clash with the ideologies of other members of the speech community and with majority language speakers as well. Indeed, it might be argued that the term “speech community” is less usefully applied to Lemkos than to other linguistic minorities, since the primary identification might be drawn from a religious, geographical or political stance by people who also happen to be speakers of Lemko. As Trzeszczyńska (2013: 443) has pointed out, Lemkos consist of “competitive communities of memory, which are only connected by the ethnonym, cultural heritage and events once lived.” This statement is without doubt true, but needs to be nuanced in terms of contestation – Lemkos do indeed share a common name (in Poland anyway), a common cultural heritage and a common history, but how their representations are increasingly “glocalized” – while following global trends of cultural and historical commodification and hybrid language practices, the Lemkos nevertheless seem to be able to creatively mingle such trends with local conditions to their advantage. The problem arises over their acceptability to the members of the Lemko community themselves and resolving such tensions would appear to be the major challenge the community faces in the 21st century.

References


Michael Hornsby
mhornsby@wa.amu.edu.pl

**APPENDIX**

<table>
<thead>
<tr>
<th>RP</th>
<th>M/F</th>
<th>Age</th>
<th>Where and when interviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>f</td>
<td>40s</td>
<td>Lublin, March 2013</td>
</tr>
<tr>
<td>2</td>
<td>f</td>
<td>40s</td>
<td>Lublin, March 2013</td>
</tr>
<tr>
<td>3</td>
<td>m</td>
<td>40s</td>
<td>Strzelce Krajeńskie, September 2012</td>
</tr>
<tr>
<td>4</td>
<td>m</td>
<td>40s</td>
<td>Legnica, July 2013</td>
</tr>
<tr>
<td>5</td>
<td>m</td>
<td>18</td>
<td>Legnica, July 2013</td>
</tr>
</tbody>
</table>

**Table 1:** Research participants for the present article
New speakers of Minderico: Dynamics and tensions in the revitalization process

Vera Ferreira
CIDLeS - Centro Interdisciplinar de Documentação Linguística e Social

From the sixteenth century on, the blankets of Minde, a small village in the center of Portugal, became famous all over the country. The wool combers, blanket producers, and traders of Minde began to use Minderico in order to protect their business from “intruders”. Later, this secret language extended to all social and professional groups and became the main means of communication in the village. During this process, Minderico turned into a full-fledged language with a very characteristic intonation and a complex morphosyntax, differentiating itself from Portuguese. However, the number of speakers declined drastically during the last 50 years. Minderico is now actively spoken by 150 speakers, but only 23 of them are fluent speakers. More than half of the fluent speakers are new speakers of the language. New speakerness is a relatively new phenomenon in the Minderico speaking community and a direct result of the revitalization process which was initiated in 2009. This paper examines the role of the new speakers in the revitalization of Minderico, considering issues of authenticity and socio-linguistic legitimacy.

1. INTRODUCTION. With the increase in popularity of the blankets of Minde since the sixteenth century, the wool combers, blanket producers, and merchants of village began to use Minderico in order to protect their business from “intruders”. Later, this “secret variety” extended to all social and professional groups in Minde and became the main means of communication in the village. During this process, Minderico turned into a full-fledged language with a characteristic intonation and a complex morphosyntax. Today Minderico risks becoming extinct, more than ever before in its history. Due to economic, social, and educational reasons, the number of speakers declined drastically during the last 50 years. This situation led to a revitalization process, which started in 2009 inspired by a DoBeS project to document Minderico, funded by the Volkswagen Foundation.

The phenomenon of new speakers as described by O’Rourke & Ramallo (2011), O’Rourke & Pujolar (2015), O’Rourke et al. (2015), Jaffe (2015) or Hornsby (2015b), and, in the case of Minderico and many other minority language communities, a direct result of the revitalization process, is thus relatively new for the Minderico speech community. Six years of continuous revitalization activities are not enough to draw clear conclusions. However, there are some tendencies that can already be observed, for instance a discrete increase in the number of active speakers (among them several new speakers) and the emergence of a (de-)legitimization discourse, which opposes on the one hand, different generations of
New speakers of Minderico: Dynamics and tensions in the revitalization process

speakers and, on the other, “traditional speakers” (Grinevald & Bert 2011: 49) and “new speakers” (O’Rourke et al. 2015: 1).

After a description of the history and development of Minderico, focusing on the linguistic peculiarities that detach it from Portuguese, on the second part of the paper I will present a characterization of Minderico new speakers and discuss the tensions that are emerging within the speech community as a result of this new linguistic constellation.

2. MINDERICO: FROM SECRET LANGUAGE TO EVERYDAY LANGUAGE TO ENDANGERED LANGUAGE. Minderico (ISO code /drc/), locally known as piação and piação dos charales do Ninhou, is an Ibero-Romance language spoken mainly in Minde (Portugal) by a community of 150 active speakers (a group composed by “fluent speakers”, “semi-speakers” and “neo-speakers” following the terminology by Grinevald & Bert (2011: 50)) and approximately 1,000 passive speakers (those who understand the language but do not speak it – “terminal speakers”, “rememberers” and “ghost speakers” in Grinevald & Bert (2011: 50–51) terms).

2.1. GEOGRAPHIC CONTEXTUALIZATION. Minde, the village of Minderico, is a small town in the center of Portugal which belongs to the municipality of Alcanena, district of Santarém, and lies 115 Km north from Lisbon and 240 Km south from Oporto. But Minderico was originally not only confined to Minde. Due to private, economic, and professional relations, Minderico extended also to two adjacent villages: Serra de Santo António where the language is already extinct and Mira de Aire where there are still four speakers with almost only passive knowledge of the language. Both villages worked on the same economic branch as Minde – textile production and commercialization. Moreover, Mira de Aire belonged administratively to Minde until 1709 and Serra de Santo António until 1918 (Martins 2010: 37). The three villages where also connected administratively, contributing to the dissemination of the language.

In Serra de Santo António and Mira de Aire, Minderico developed particular lexical and phonetic features. Following the strategies of vocabulary development in Minderico (cf. Section 2.2), some lexemes were adapted to the social reality of these two villages. For instance, in Minde francisco vaz is the most frequent lexeme for “priest” which derived from the name of one of the most important priests in the village in the eighteenth century. In Mira de Aire, the word raso (a Portuguese-based word meaning “full”) is used instead, because the priest in Mira de Aire was known in the community for being almost always drunk, i.e. “full with alcohol”. On a phonetic level, the closeness of vowels and monophthongization are some of the features that characterize the Minderico variety of Mira de Aire (Minde terrazinha vs. Mira de Aire tirrazinha “girl, little girl”). However, it was in Minde where Minderico developed most and maintained the status of language of everyday communication until today, although with less prominence and much fewer speakers.

Minde was (and still is) a monoindustrial village of textile and wool artifacts (Martins & Nogueira 2001: 147–159, Martins 2010: 86–90). Due to its strong and prosperous textile industry, Minde had more than 7,000 inhabitants until the end of the 1970s. According to the census undertaken in 2011, the population decreased considerably, to 3,293. One of the main reasons for the accentuated decrease was the crisis in the textile industry: Several people had to leave the village in order to look for job opportunities in the Portuguese urban centers (Lisbon and Oporto) but also abroad (mainly United States, Canada, France, Germany, and Switzerland). As expected, this socio-economic development had also a clear impact on the vitality of the language and its use, as we will see in Section 2.3.
It must be emphasized that the development and maintenance of Minderico in Minde is also intrinsically related to and can be explained by the geographic position and the geological features of the village. Minde lies in a close depression between the plateau of Santo António and the plateau of São Mamede. On the west side, the village is surrounded by a polje (a large flat plain in karst territory that inundates during rainy winters and spring seasons). Therefore, the access to Minde was till recently very difficult (Martins & Nogueira 2002: 213–214). The geographical isolation not only contributed to the evolution and preservation of Minderico but also reinforced its development as an independent language with its own system and particularities, unintelligible to Portuguese speakers (cf. Section 2.2).

The coat of arms of Minde (Figure 1) reflects the intrinsic relation between the geological features, the economic activity, and the language – three elements that strongly formed the identity of what can be called the Minderico community. The green nest on the top of the coat of arms metaphorically represents the geological depression in which Minde lies and, simultaneously, the autochthone name of the village - Ninhou (“big nest”), an augmentative derivation of the Portuguese word ninho “nest”. The needles with wool in the middle show the importance of the textile industry and the blue waves at the bottom are again related to the geological specificities of the village, representing the polje.

**Figure 1:** Coat of arms of Minde (photo by Vera Ferreira)
2.2. FROM SECRET LANGUAGE TO EVERYDAY LANGUAGE. In order to protect their business from intruders in the markets they visited all over Portugal, the wool combers, blanket producers, and merchants of Minde created a special language, based mainly on Portuguese. It allowed them to negotiate the prices for the blankets among each other in front of strangers and/or customers in an unintelligible way.

The first written documents (mainly personal letters, wills, and church registrations) appeared at the end of the eighteenth century, some of them reporting on histories in the past. Thus, according to the sources available, one can assume that Minderico emerged at the end of the seventeenth century as a sociolect, a secret language of a professional group. This is, in fact, the well-known and most admitted explanation for the emergence of Minderico (Furriel 1996; Martins & Nogueira 2002: 133–136; Martins 2004: 4–6; Martins 2010: 229). Following Ferreira et al. (2015: XIX–XXIII), the lexemes related to textile production and commercialization are, however, almost inexistent in the Minderico lexicon. The only clear reminiscence of this possible sociolectal origin is its high complicated numerical system. Instead, lexemes related to everyday communication and everyday needs (food, drinks, human body, means of transport, animals, etc.) are the most frequent ones. This can be explained by the fact that Minderico, contrary to the normal limited lifespan of secret languages (Siewert 1999, Klepsch 1996, Geipel 1995), has evolved from a secret language to an everyday language, being used not only for commercial reasons but also
and mainly in all daily social contexts, becoming the main means of communication in the village and a unifying identity element.¹

During this process, Minderico expanded its vocabulary continuously, adapting it to the needs of the community and reflecting simultaneously the technological and socio-cultural developments of the society in which it was integrated. Vocabulary enlargement was (and still is) intimately tied to the socio-cultural experiences of the xarales (the inhabitants of Minde). For example, names and nicknames of well-known persons from Minde and the neighboring areas were used as lexemes to express physical or psychological characteristics, as these characteristics were salient for those persons. Being a small and close-knit community, where everyone knows each other, this method of vocabulary formation did not represent an obstacle to effective communication.

Thus, apart from loanwords such as ganau and ambria from Spanish hambre “hunger” and ganado “cattle” respectively, naifa from English knife and jones “hat” from the English anthroponym John, or French père and mère (modern forms for videiro and videira) for “father” and “mother”, metaphors and metonymies are the two main recurrent strategies of Minderico vocabulary enlargement (Ferreira & Bouda 2009).² The words piar “to speak, to talk”, pataeira “watermelon, breast”, and a do aníbal “bicycle” are examples thereof. Piar, from Portuguese piar “to cheep”, is based on the metaphorical projection of the sound produced by the birds to the human domain. In pataeira, the meaning “watermelon” derives metonymically from the toponym Pataias, a place known in the region for its watermelons (the suffix –eira reinforces the idea of origin); “breast”, on the other hand, is a metaphorical development based on the similarity of form between watermelons and breasts. A do aníbal represents a metonymy based on the anthroponym Aníbal which was the name of the owner of the first bicycle repair shop in Minde.

In the process of becoming the everyday language in Minde, Minderico not only enlarged its vocabulary but turned into a full-fledged language with a special intonation and a complex morphosyntax, a language with a system of its own, unintelligible for Portuguese speakers. This development (from secret language to everyday language) is not unknown to linguists and is well discussed in the literature on secret languages.³ In this context, for instance, Heinz Kloss (1967: 29) talked about Abstandsprache or language by distance (“the reference being of course not to geographical but to intrinsic distance”).

Minderico belongs without doubt to the group of Ibero-Romance languages. It is diachronically related to Portuguese, a fact that obviously determined its structures which show clear Ibero-Romance characteristics (Ferreira & Bouda 2009: 100–101). However, it also developed features that detached it clearly from Portuguese.

For instance, in the domain of morphology, “elliptical partitive constructions” (Ferreira 2011, Ferreira & Bouda, Ferreira et al. 2015), exemplified in (1) os do noé “animals”

¹ The work by Ferreira et al. (2015) opens up a new hypothesis for the origin of Minderico, which questions its sociolectal background and needs further research. Maybe the language existed in Minde before the boom in the textile industry and was already used at that time for everyday communication in the village, a fact that could be easily explained through the geographic isolation of the village and corroborated by Mozarabic influences in the lexicon (Ferreira et al. 2015). Its subsequent use for business protection could be seen, therefore, as a logical consequence, considering the unintelligibility of Minderico outside of Minde. Because of the lost of communicative domains over time, its business function remained prominent in the memory of the community, which may have influenced the explanation of its origin.

² A study I carried out in 2008 and presented in Ferreira & Bouda (2009: 103) shows that more than 60% of the Minderico vocabulary is based on metaphors and metonymies.

New speakers of Minderico: Dynamics and tensions in the revitalization process

(1) os animais do [M.PL] Noah
(2) o pássaro do [M.SG] padre Faria
(3) a alface da [F.SG] morcela

could be mentioned as a typical and active morphological strategy for noun formation in Minderico. I called them elliptical because after the article (os, o, as) one expects a noun that is actually missing. The missing noun is essential for the meaning of the construction. We could easily reconstruct (1) as “the animals of Noah”, but the reconstruction is not always so straightforward and most of the time it depends on a profound knowledge of the socio-cultural, economic, gastronomic, and political traditions in Minde. We are not able to reconstruct (2) as “the parrot of priest Faria”, if we did not know that the priest Faria, who lived in Mira de Aire, was known for having a parrot in the balcony of his house. The same happens to (3). Its meaning is only comprehensible if we consider that blood sausage in Minde is always prepared with parsley.

Contrary to Portuguese and other Romance languages, Minderico shows traits of nominal incorporation, as it is described by Mithun (1984, 1986) and Mithun & Corbett (1999), a phenomenon which has effects on verb valence and syntactic structure. Verbal constructions with nominal incorporation comprise a light verb, empty of meaning or with a very general meaning, and a nominal element which lost referentiality. The noun is integrated into the light verb and builds with it an indivisible unit, specifying its meaning. In Minderico, there are three light verbs – gâmbiar “to do something with the hands”, pôr “to put”, and jordar “to do, take, bring, go, …”, being the latter the most frequent one.

(4) Ali o covano jorda-as as perneiras

“He puts on the socks” (lit. “There the man clothes-puts the socks”)

In example (4), the noun as do mestre-grosso “clothes” is part of the verb jordar, specifying its meaning to “to dress”. It is not possible to add any other word between the light verb and the nominal element without losing the original meaning, nor can as do mestre-grosso be pronominalized – Ali o covano jorda-as as perneiras would mean “he take/bring/throw them, the socks” in which as perneiras would be simply a reiteration for emphatic purposes of the pronominalized direct object of jordar. As do mestre-grosso is, thus, an intrinsic part of the verbal complex, which as a whole shows a transitive pattern.

Example (4) also shows another clear difference between Minderico and Portuguese, namely in the pronominal system. The personal pronouns in Minderico are based on proximal (aqui), medial (aí), and distal (ali) deictic adverbs which are combined with the nouns covana “woman” and covano “man” for feminine and masculine: aqui a covana/o covano “1SG.F/M”, aí a covana/o covano “2SG.F/M”, ali a covana/o covano “3SG.F/M”, aqui as covanas/os covanos “1PL.F/M”, aí as covanas/os covanos “2PL.F/M”, and ali as covanas/os covanos “3PL.F/M”.

---

4 Abbreviations: DART – definite article; M – masculine; F – feminine; SG – singular; PL – plural; 1 – first person; 2 – second person; 3 – third person; LV – light verb; PRES – present.
5 See note 4.
6 See Ferreira & Bouda 2009 and Ferreira forthcoming(a) for a further discussion on nominal incorporation in Minderico.
7 Or a charal/o charal with the same meaning but only used by and for people who were born in Minde.
2.3. MINDERICO TODAY. As mentioned before, the number of speakers of Minderico declined drastically during the last 50 years and Minderico risks becoming extinct more than ever before in its history. Intergenerational transmission was interrupted and Minderico is no longer passed to children at home. All speakers of Minderico are and were always bilingual, speaking Portuguese along with Minderico. Minderico is almost only used in familiar contexts (when talking to older members of the family and to friends), but even in this context the pressure of Portuguese as the language of education, administration, economy, etc. is clear. Thus, bilingualism with clear diglossia characterizes the speech community.

Moreover, the knowledge of Minderico is not homogeneous among its speakers. Currently, there are 150 active speakers, but only 23 of them are fluent speakers. Curiously, 15 of the fluent speakers are new speakers, i.e. they did not learn the language at home but through the revitalization process or through the contact with other fluent speakers; they learnt it in a later period in their lives, by their own decision. Moreover, Minderico has about 1,000 passive speakers, who understand the language but have very limited productive skills reflected for instance in some frozen expressions. Intensive code switching between Portuguese and Minderico characterizes the speech of the majority of active speakers.

Additionally, the almost inexistent presence of Minderico in the media and in new digital domains, together with the lack of official recognition as a minority language and the consequent lack of official support and prestige, contribute to its current endangered status.

The Volkswagen Foundation in the framework of the DoBeS program, 8 funded a documentation project 9 which allowed the collection of data 10 necessary for the production of Minderico teaching materials that had previously been non-existent. Thus, in 2009, a group of members of the community, with the support of CIDLeS – Interdisciplinary Centre for Social and Language Documentation, 11 developed and initiated a long-term revitalization process. The activities that are being carried out are varied and target a large audience, not only the younger generation, aiming at the involvement and commitment of the whole speech community, in different domains of its everyday life (see Ferreira forthcoming(b)).

Six years of continuous revitalization activities are not enough to draw clear conclusions. However, there are developments and tendencies that can be observed already, namely more digital writing in the language (SMS, E-mails, blog entries, etc.) and a discrete increase in the number of active speakers (almost all new speakers of the language). Simultaneously, a (de-)legitimation discourse unknown before the beginning of the revitalization is also emerging. It opposes on the one hand different generations of speakers and, on the other, traditional speakers and new speakers, a topic that will be discussed in the next section.

3. NEW SPEAKERS OF MINDERICO. The category of “new speaker” is not new and, of course, not only confined to minority language contexts. In the literature, it is commonly examined under labels such as “non-native speaker”, “second language speaker”, “foreign language speaker”, “L2” speaker or “learner”. However, these labels focus more on language proficiency and the assumed / expected quality of the linguistic knowledge. “New
New speakers of Minderico: Dynamics and tensions in the revitalization process

speaker", instead, is more neutral, focusing less on the linguistic knowledge and more on the social elements and motivations behind communicative practices (Costa 2015: 128). Thus, the characterization of Minderico new speakers proposed in this section follows the definition developed by the researchers of the COST Action New Speakers in a Multilingual Europe: Opportunities and Challenges and further specified for the context of minority languages, namely

The "new speaker" label is used (…) to describe individuals with little or no home or community exposure to a minority language but who instead acquire it through immersion or bilingual educational programs, revitalization projects or as adult language learners. (O’Rourke et al. 2015: 1)

Considering that the revitalization of Minderico started six years ago, it is important to highlight that the characterization of Minderico new speakers aimed at in this paper represents the first approach to the topic in the Minderico context.

The phenomenon of new speakerness is relatively new for the Minderico speech community and was mainly fostered by the revitalization process. In this sense, three groups of new speakers can be distinguished according to speakers’ backgrounds and relation to the language:

1. People who came to Minde at some point in their lives for personal (marriage) and/or economic reasons and only got acquainted with the language as they moved to the village – this is the most prominent group;

2. People who always lived in Minde but were not exposed to Minderico in traditional familiar language acquisition contexts – this group comprises almost all passive speakers and the children learning Minderico at school now;

3. People who do not live in Minde but in some way (mainly though their ancestors) feel connected to the village and its culture.

In face of the shrinking number of speakers, the “incapacity” of the “traditional” speech community to reverse language shift, and the highly endangered status of the language, members of these three groups have acquired an extremely important role in the revitalization process, mainly because they are the ones that foster it. For the new speakers of Minderico, learning and using the language is a way not only of preserving a cultural identity with which they identify, but also of (re)integration in a community to which they now actively belong (or in some way belonged before). One can undoubtedly affirm that the future of Minderico clearly depends on its new speakers and their positive attitude towards the language.

3.1. Observable Tensions. The emergence of new speakers brought, in turn, some tensions within the speech community. To understand them and the reasons behind them, it is important to remember that the Minderico community is, in general, a very closed one.

---

12 For the history and theorization of the “new speaker” concept and its application to different European minority language contexts see among others Hornsby (2015b) for Breton, Yiddish, and Lenno contexts and the issue no. 231 of the International Journal of the Sociology of Language, which is a special volume on New speakers of minority languages: the challenging opportunity edited by Bernadette O’Rourke, Joan Pujolar, and Fernando Ramallo in 2015.

in which it is really difficult to get in and be accepted as “one of them”. The semantic
differentiation between xarales (people who were born in Minde and may or may not live
in village) and covanos (people who were not born in Minde but moved to the village –
this category applies even to the children of the incomers who were already born in Minde)
corroborates the social characteristics mentioned above. Consequently, charales have more
(implicit) legitimacy over all communitarian societal aspects (language, culture, traditions,
etc.) than covanos.

In this context, two different types of tensions could be observed: a) between different
generations of speakers and b) between “traditional” speakers and “new speakers”.

Intergenerational tensions that oppose old and young speakers, and which are in fact
common to all other languages, in the case of Minderico go in opposing directions and are
related to shifts in authority and language legitimacy. On the one hand, the older speakers
do not accept easily the way the younger generation speaks Minderico, mentioning that
they speak a kind of “modern Minderico” and not “pure Minderico”, delegitimating at the
same time their knowledge of the language by classifying what the young people speak
as “invented, artificial language”.14 They are very critical above all about the enlargement
of Minderico vocabulary to modern contexts of daily life, such as contexts related to new
technologies, mainly because for them “authentic Minderico” is connected to a concrete
difficult period in their lifetime (when they went to the markets to sell the blankets) and
characterized by communicative practices with clearly delimited diglossic boundaries.15
They do not associate Minderico with modern life – for them, this role is played by Por-
tuguese. This attitude obviously influences the way the younger generation uses or decides
not to use Minderico. Some of them feel that they are not proficient enough and do not feel
confident in using the language in the presence of elders – a fact that sometimes leads to
avoiding the language at all.

On the other hand, some members of the older generation underestimate their knowl-
dge of Minderico just because they do not speak “modern Minderico” which they in-
herently associate with higher levels of education, social status, and consequently more
prestige – some of the features that characterize the younger speakers of the language.

But the main tensions are the ones that oppose traditional speakers (most of them very
old and with a good proficiency) and new speakers. Traditional speakers of Minderico are
those that have acquired the language at home, in Minde, and always used it in everyday
communication. The new speakers, in turn, as mentioned in Section 3, are mainly people
who acquired the language in “artificial” language contexts, during revitalization activities.

Traditional speakers do not recognize the speech of new speakers as authentic Min-
derico – curiously, this opinion comes inclusively from several passive speakers, the majority
of them charales. The main argument in this tension is prosody, following, thus, the trends
encountered in other minority language communities with similar tensions, as discussed
for instance by Hornsby (2015a: 110) and Costa (2015: 133). Traditional speakers argue
that new speakers are not able to produce authentic Minderico prosody and accuse them to
deteriorate the language by using Portuguese prosody when speaking Minderico – a case
of common “blurring of linguistic boundaries” (O’Rourke & Ramallo 2013: 18). They
explain the lack of authenticity in the speech of new speakers with the fact that most of
them were not born in Minde or are not from Minde or were away from the village for too
long. At this point, it is important to emphasize that prosody is one of the most endangered
aspects in Minderico; even traditional speakers are losing it, using gradually more Min-

14 This attitude is comparable to what Costa (2015) has observed in the Occitan context.
15 For a comparable study on Breton see Timm 2010.
New speakers of Minderico: Dynamics and tensions in the revitalization process

Contrary to traditional speakers, new speakers see Minderico as a necessary element of their modern life and not something related to the past that cannot be “updated”. Even though they are aware of their “insufficient” proficiency, they see themselves as the revitalizing forces and are conscious that the future of the language depends on them, a consciousness and recognition also encountered among new speakers of other minority languages in Europe, as for instance among the new speakers of Galician (O’Rourke & Ramallo 2013: 29), Scottish Gaelic (McLeod & O’Rourke 2015: 169, O’Rourke & Pujolar 2015: 147) or Occitan (Costa 2015). They auto-legitimate their speech on another level with the argument that if they do not use the language as they can and transmit it at home, in their families, then it will die out in a very short period of time, an argument very much in the lines of what O’Rourke & Pujolar (2013) describe in “cases of extreme language shift”:

[I]n cases of “extreme language shift” (...) linguistic legitimacy and authenticity can no longer be linked to the seemingly inherent characteristics of its speakers. Instead, legitimacy comes from those who claim authority and construct such legitimacy. (O’Rourke & Pujolar 2013: 58)

And Costa (2015) adds that “legitimate language is not a given. It is constantly negotiated among users, and what constitutes legitimate language may vary according to the setting in which it is used” (Costa 2015: 129), and the Minderico case is a clear example thereof.

4. CONCLUDING REMARKS. The revitalization of Minderico is a recent process. It has started in the year 2009. Despite its limited lifespan, one could already observe the emergence of a new phenomenon within the social structure of the speech community, namely the phenomenon of new speakerness and the connected discourse about locality, authenticity, legitimation, and language ownership that has started to question the position and power commonly ascribed to “traditional speakers” of the language.

Considering the highly endangered status of Minderico which manifests itself not only in the reduced number of speakers, the lack of official recognition, support, and digital presence, but also in the lack of prestige and economic value as well as the prevalence of negative and conservative attitudes towards the language influenced by the Portuguese public and academic opinion, I believe that the maintenance of Minderico really depends on the new speakers and their attitudes towards the language with which they socially and culturally identify. They are in fact playing a decisive role in a process that aims at reversing language shift.

Bearing in mind my experience with the community, the success of that process will depend on the capacity to find the balance between the linguistic, cultural, and historical knowledge of traditional speakers and the energy and engagement of new speakers. That can only be achieved by involving both groups and considering their different needs in the revitalization process.
New speakers of Minderico: Dynamics and tensions in the revitalization process

References


Ferreira, Vera, Ilona Schulze, Paulo Carvalho Vicente & Peter Bouda. 2015. Dicionário Bilingue Piação-Português. Minde: CIDLeS.


Vera Ferreira
vferreira@cidles.eu
Kormakiti Arabic: A study of language decay and language death

Ozan Gulle
Ludwig-Maximilians-Universität München

Kormakiti Arabic (also called Cypriot Maronite Arabic) is a language with approximately 150–200 speakers in Kormakitis, a village north-western Cyprus. Kormakiti Arabic is highly endangered, not only due to its low number of speakers but more importantly because younger Maronites with their roots in Kormakitis do not acquire Kormakiti Arabic naturally any more. Kormakitis itself is almost only inhabited by elderly Maronites who lived there before the separation of Cyprus in 1974. This paper is on language death and language decay of Kormakiti Arabic. Several historical sources are used in order to illustrate the historical and socio-linguistic environment this language survived until today. The linguistic evidence is then compared with the theory of Gaelic-Arvanitika-Model Sasse (1992a) in order to show parallels, as well as the differences between Arvanitika and Kormakiti Arabic.

1. INTRODUCTION. Kormakiti Arabic is a minority language spoken by Maronites in the villages Kormakitis, in north-western Cyprus. The language is tightly knitted with Cypriot Maronites (which is why it is sometimes called Cypriot Maronite Arabic, though I shall discuss why this term is not correct) and in order to understand Kormakiti Arabic and its current status as a highly endangered language, it is crucial to understand the socio-linguistic history of Cyprus.

The research this paper is based on is part of a Ph.D. project on structural convergence in Cyprus among Cypriot Greek, Cypriot Turkish and Kormakiti Arabic. All of these three languages have their respective corpora. The corpus of Kormakiti Arabic consists of approximately 5000 tokens and includes texts from Borg (1985), as well as from own field research. Borg (1985) is the most detailed grammar on Kormakiti Arabic which has its own text collection (unfortunately without any audio support). It is, however, not the first or only work on this language. The first research on Kormakiti Arabic was Newton (1964) “An Arabic-Greek Dialect”, followed by Tsiapera (1969) “A Descriptive Analysis of Cypriot Maronite Arabic”. Apart from these descriptive works, there are also a few papers on Kormakiti Arabic such as Jastrow (1977) “Gedanken zum zypriotischen Arabisch”.

This paper illustrates how endangered Kormakiti Arabic is and how its endangered language status came to be, using historical as well as linguistic data. The historical and socio-linguistic data will be used to support these claims, as I assume that socio-linguistic

1 I wish to thank Ludwig-Maximilians-Universität München and its Graduate School Language & Literature Munich for their support during this project.
parameters are the main decisive factors behind language contact. The main focus of this paper is on language death and language decay, which shall be elucidated in their respective chapters.

2. Cypriot Maronites. Maronites are Eastern Syriac Christians belonging to the Eastern Catholic Church. They are named after Saint Maron who lived near Mount Taurus, thus the name “Maronite” actually refer to a religious community rather than to an ethnic/national one (Hourani 2007: 1). There is a Maronite community in Lebanon today of a considerable size and a Maronite diaspora in various countries. Throughout the history, many Maronites migrated to Cyprus in different waves due to different historical events.

According to Hourani (2007: 1–5), the first appearance of Cypriot Maronites on Cyprus begins around the 7th or 8th century, mainly due to Islamic conquest of Maronites homeland (which was mainly today’s Lebanon and Syria after their dispersion from Antioch) and the inter-Christian rivalries between the Jacobites and Byzantines. The destruction of Saint Maron’s monastery in Apamea around 938 caused another wave of immigration, followed by a third one upon the purchase of the island by Guy de Lusignan from the Knights Templar at the end of 12th century (Hourani 2007: 5, Kyrris 1985: 212). The last wave happened after the defeat of the Crusaders in Tripoli and the Holy Land. In 1121 and 1141, two Maronite monks were appointed by the Maronite Patriarchs to be abbots of the St. John Chrysostom monastery at Koutzoventi. Hill (1948: 3) mentions a lack of record of Maronites on the island between 1141 and the end of the 13th century. Reportedly, they chose not to settle in cities but in the mountains north of Nicosia. They must have had a “chief centre” called Tala or Attala in Karpass but Hill writes that this place was then “no longer traceable”. Hill also notes that under the last years of the Lusignan rule, the number of Maronites must have been around 7000 and 8000 (Hill 1948: 4). Tsoutsouki (2009: 209) notes that Maronites indeed came from different regions. Thus, inhabitants of Kormajit originate from Kur in north of Lebanon, the inhabitants of Asomatos from Shmat in Byblos, and the people of Ayia Marina from Qanoubine area. The ancestors of people from Karpasha are allegedly from a village close to Tripoli.

Kyrris (1985: 206) counts the villages Kormakitis (originally Krommyakites), Asomatos, Karpassa, and Kambylia among the Maronite villages of Cyprus. Apart from these, Ayia Marina (also called Santa Marina in Gemayel (2009: 139)) also used to be a Maronite village according to my Maronite informants. It is not known, how many villages the Maronites had during the Lusignan and Venetian rule, but according to Gemayel (2009: 137), right after the Ottoman conquest, the Maronites did not have more than 33 villages. Tsoutsouki (2009: 203) reports that by 196, there were 19 Maronite villages.

2.1. Social status of Cypriot Maronites. The beginning of the Lusignan and thus Catholic reign on Cyprus marks one of the most important turning points in history for Cypriot Maronites. During the Lusignan rule in Cyprus, the Maronite community in Cyprus saw an increase in its social status. Maronites were Catholic just like their Lusignan rulers on the island, and, following Hourani (2007: 8), they received extensive freedoms and exemptions, though she does not explicitly give examples. Consequently, the Cypriot Maronites began losing their social status after the Ottomans conquest of the island from the Venetians. According to Hourani (2009), “[w]hen the Orthodox Church regained its
power, which it had lost during centuries of Catholic rule, its members remembered the oppression of the Catholics and since most of the Catholics who were in the island were the Maronites, they began their retaliation against them.” (117–118). Catholic churches were confiscated by the Orthodox Church and the members of the Catholic church were accused of working against the Ottoman rule. As a result, many members of the Maronite clergy were imprisoned or killed and the believers of the Catholic church were forced to convert to the Orthodox faith. As a result of being under the control of the Orthodox church, the Maronite churches were closed on major religious holidays in order to punish the believers of the Catholic faith for belonging to it (Gemayel 2009: 140). The reason for the Catholics being under the Orthodox church was the Ottoman ferman (royal decree) of October 1571 forbidding Catholics to live or own property in Cyprus, including the churches (Kyrris 1985: 254). The remaining Catholics were forced to belong either to the Orthodox church or the Muslim community. The Catholic church was re-established on the island after the peace treaty with Venice on 07.03.1573, but the survivors of the Ottoman conquest were already forced to leave their religion in this brief period and the property damage was done.

Today, Maronite Cypriots have more or less the same social status as Greek Cypriots. Although they would go to a Catholic church on Sunday and not to an Orthodox one, they exclusively speak Cypriot Greek in public and have Greek (or at least hellenized) names. Tsoutsouki (2009: 194) refers that the Maronites are “Greek in public and Maronites at home”. Since 1974, the hellenization of Maronite Cypriots accelerated as most of them were forced to leave their home villages, thus also their cultural centres. Tsoutsouki (2009: 205) mentions that the inhabitants of the four Maronite villages made up more than 97% of the Maronite population on the island. After 1974, they enjoyed a privilege granted to no other community in Cyprus, namely crossing the Green Line. After the opening of borders in 2003, this has even become easier, allowing Maronites to visit their villages more often.

3. KORMAKITIS AND KORMAKITI ARABIC. Although it is usually called Cypriot Maronite Arabic, the language is not spoken by Maronites outside of Kormakitis, except for the Kormakitis Maronites who migrated to cities after 1974. It is, thus, not the language of Cypriot Maronites but the language of Cypriot Maronites from Kormakitis. Obviously, the village Kormakiti is the centre of Kormakiti Arabic and of the culture. Inhabitants trace their roots back to Kur in today’s Syria where their ancestors migrated to Cyprus. Kormajit allegedly comes from the phrase Kor ma-jit, meaning “Kor did not come”. As to why Maronites from other villages do not speak the language, there are no written records. Some inhabitants from Kormakiti stated that the Maronites in other villages never spoke Arabic and speculate that they might have been hellenized linguistically before arriving in Cyprus. The inhabitants of Kormakiti often refer to their language as lũğa tel-deʔa (“language of the village”) or simply as aravika (“Arabic” in Greek).

Today, every speaker of Kormakiti Arabic is bilingual in Greek. Most of its speakers are above 40–50 years old and there is little chance that the language will survive in a few generations, unless the language revitalization efforts prove to be successful. There is also no written standard for the language, except for the systems developed in the last decades by linguists who documented the language and its speakers when trying to teach the language to younger generations of Maronites who did not learn it from their parents. Although the Maronites of Kormakitis (and Cypriot Maronites in general) went through difficult times concerning their social and religious situation, these did not cause the speakers of Kormakiti Arabic to shift to Greek. According to the elderly informants, the shift to Greek started happening in the 1950s, possibly due to the increasing ethnic tensions in Cyprus, though the
informants did not mention it specifically. The ultimate completion of the shift process was 1974 when Cyprus was divided into two parts and many working Kormakitis Maronites left their village and moved to southern Cyprus, while the elderly stayed in Kormakitis which is now in the northern part of the island.

4. BETWEEN LANGUAGE MAINTENANCE AND LANGUAGE DEATH. Language contact scenarios are usually divided into two, namely “language maintenance” and “language shift” (Thomason 2001, Thomason & Kaufman 1988). Roughly, the speakers keep their native language in a language maintenance scenario, whereas they shift to another language in a language shift scenario. Language maintenance can range from casual contact to very strong cultural pressure (Thomason & Kaufman 1988: 74–6). The main idea behind the whole theory is that the stronger the social interaction between two speech community is, the more intense the language contact phenomena are going to be. Thus, causal language contact only leads to lexical borrowings, whereas stronger social interaction may lead to morphological and syntactic borrowings.

Language shift, on the other hand, has its own mechanics. Changes caused by language shift are usually found in the “target language”, i.e., the language to which the speakers shift. What is more relevant and interesting for this paper is the indirect effect language shift might have – “language death”. The term language death is generally used to describe the situation when a language does not have any speakers left. As Crystal (2002) puts it, a language is practically dead when the second last speaker dies, as there is no one else left for the last remaining speaker to speak the language to, except for maybe a few semi-speakers who understand it but cannot speak.

As to how language death occurs, Sasse emphasizes that it is first external factors that affect the whole situation, causing the speech behavior to change, which in turn then leads to changes in the linguistic structure (Sasse 1992b: 13). Language death is caused primarily by language shift (excluding a handful other completely non-linguistic factors which may also cause it, such as for instance genocide). Sasse defines language shift as an interruption in “language transmission”, i.e., the speakers stop passing on the language to the younger generations. In a process which he calls “primary language shift”, the speakers make another language “primary language” and their native tongue “secondary language”. This may be followed by speakers giving the abandoned language a negative value; or sometimes considering it to be positive, when it is seen as a feature of identification, and in other cases to be negative, e.g., as a tool for everyday communication (Sasse 1992b: 14). At this point, the linguistic changes in the abandoned language can be observed as it goes under a process called “language decay” (sometimes also called “language attrition”). This stage is named thus as certain features of the language begin to disappear and a new kind of speakers arise. These speakers, called “semi-speakers”, have an “imperfect knowledge” of the language, since they were never fully exposed to it. Sasse (1992a) categorizes the semi-speakers in two types. The first group of semi-speakers are the ones with good language proficiency but who never became full speakers due to lack of a regular use of the language. They are called “forgetters” or “rusty speakers” who have gaps in the lexical repository but otherwise know the language. These type of semi-speakers can be found in a situation where a rapid language shift took place or the language is dying, and many of the speakers ceased transmitting the language whereas some of them still do. Type 2 constitutes the group of semi-speakers who are also called “semi-speaker proper” who grew up in families with no language transmission to its children. The children can pick up some of the language by listening to the elderly speaking among themselves and maybe occasion
ally speaking it themselves with the elderly (Sasse 1992a: 62). Thus, there is a continuum of semi-speakers with different proficiencies.

As long as a language does not suffer from a sudden death, i.e. all of its speakers dying out due to war, disease or another non-linguistic reasons, there is always a certain time period in which the dying language undergoes certain changes. Naturally, the mechanisms involved in language death are different than those in language maintenance. Although it is difficult to define “the” mechanisms of language death, as it is with language contact in general since “anything goes” (Thomason 2001), we are able to observe similar changes in dying languages. Sasse (1992a) exemplifies certain changes in Arvanitika, the Albanian dialect in Greece (its very name stands for “Albanian” in Greek), which died out due to its speakers’ shift to Greek – a situation very similar to that of Kormakiti Arabic. By observing the changes in Arvanitika, it is maybe possible to draw some similarities to the changes in Kormakiti Arabic. The model presented by Sasse (1992a) is called Gaelic-Arvanitika-Model (GAM) and it is a model of gradual death.

1. The first change was the loss of subordinative mechanisms. As Sasse explains it: “Arvanitika semi-speakers do not use gerund forms of the verb in spite of the fact that there is an exact parallel form in Greek. The most frequent type of subordinate clause is the short relative clause. Adverbial clauses are avoided except for those introduced by ‘when’ or ‘if’. At the phrase level, modifiers are rare; genitives and adjectives are not frequently used.” (Sasse 1992a: 70).

2. The second change was the loss of systematic integration, i.e., the Greek lexical items were not integrated into Arvanitika phonetical and phonological system and were used in the exact same form as in Greek. For example, λίτορας is the Greek borrowing for “television” in Arvanitika, but the semi-speakers were using the original Greek form λίτορας.

3. The third change was the breakdown of grammatical categories, as the whole TAM system was coming apart and the semi-speakers did not differentiate between forms like the future particle do and the subjunctive particle tα, even inventing mixed forms such as de or do (Sasse 1992a: 70).

4. The final change was “agrammatism”, as Sasse calls it. He defines it as “the total dis-integration of the morphological system” with effects such as the loss of suppletive forms in paradigms, the mixing-up of personal markers in verbs, the regularization of plural noun forms, and the syntax getting mixed up. Sasse also notes that although semi-speakers were making these mistakes in their language, there were utterances which they were perfectly producing. This is due to the fact the speakers know certain utterances by heart and can repeat them. The mistakes listed usually happen when these semi-speakers were asked to be creative and form spontaneous utterances (Sasse 1992a: 72).

5. Semi-speakers also had problems finding lexical items.

6. Extreme phonological variation and distortion can be observed in the speech of semi-speakers.

7. Lastly, semi-speakers use phonological hypercorrections. In the case of Arvanitika, a lexical item like herα became çerα under the influence of Greek phonology, where
the same phoneme is /x/ before back vowels and /ç/ before front vowels. Then, in an “attempt to imitate the ‘something different’”, the semi-speakers pronounced it as xer (Sasse, 1992a: 72).

Although these may not be the exact changes in Kormakiti Arabic, the Arvanitika examples show a general tendency towards both generalization/standardization, confusion and loss in language usage. These examples and tendencies are going to be the basis for a comparison between language death in Kormakiti Arabic and Arvanitika. In general, the aspects of language decay are easy to observe and I follow Sasse in his notion that “the bulk of typical decay phenomena, especially agrammatism, syntactic reduction, and extreme variability, is so different from what happens in normal contact-induced change, that it can be clearly set off from the latter.” (Sasse, 1992a, 75).

This list of features of language decay by Sasse is very similar to what other scholars discussed on this topic. Dressler (1988) lists several structural and functional changes during language decay, namely: 1) Borrowing of several lexical items without necessary phonological and phonetic integration. Dressler calls these lexical items “Gastwörter” (lit. “guest words”) (Dressler 1988: 1552); 2) Loss or change in the phonology and intonation of the language by semi-speakers; 3) The native “productive processes” of the language are lost and replicated with those of the contact language; 4) Too much phonological/lexical variation; 5) Loss of stylistic registers, which leads to a monostylistic language.

4.1. Structural Borrowings in Kormakiti Arabic. This paper focuses more on language death and language decay regarding Kormakiti Arabic, and not so much on the borrowings. The reason for including structural borrowings is to prove two important points: 1) Language maintenance and language shift need not be mutually exclusive (Thomason & Kaufman 1988: 45), and language contact is a dynamic process. Two languages can be in a stable contact situation where both speech communities maintain their languages during a certain period, while the whole situation tips over in another period, where changes in socio-linguistic situation cause the speakers of one speech community to shift to the other language. 2) Following the first point, not every language contact phenomenon in Kormakiti Arabic can be attributed to language decay. There are certain cases which most probably have happened while the speakers of Kormakiti Arabic were still maintaining their language. Due to space restrictions, I will only mention two such examples.

4.1.1. Directive and Locative Marking. Kormakiti Arabic seems to lack any marking for directive and locative, very similar to spoken Greek. In written Greek, locative or directive is usually marked with an inflected form of stos (which is probably a grammaticalization of the locative/directive preposition se as a proclitic on the article) followed by a noun in the accusative. In spoken Greek (in Cyprus and on the mainland), this preposition is usually missing and the directive or locative is only marked with the accusative case. Since Kormakiti Arabic lacks any case, these constructions are not marked in the language. This separates Kormakiti Arabic from other varieties of Levantine Arabic which do not only mark directive and locative, but also differentiate between these two.

\[
\begin{aligned}
apn \quad smn-t \quad oti \quad eprp \quad ta-r \quad strayt
\end{aligned}
\]

PRN.1SG hear-PST.1SG COMP NEC SUBJ-go.1SG army

“I heard that I was supposed to go to the army.”

b. Locative (Borg (1985): “Stories about my youth.”)

\[
\begin{aligned}
koft-n \quad ext \quad sa\at- \quad k-kaf
\end{aligned}
\]

sit.PST-1PL one hour-DUAL ART-coffee.house

“We sat inside the cafe for about a couple of hours.”

However, I heard and noted speakers of Maronite Arabic using fi to mark locative (as in other Arabic varieties) on various occasions. As I discussed my notes with the speakers, however, they insisted that fi is sometimes used due to the influence of Levantine Arabic, but in “true” Maronite Arabic locative is not marked. There are few cases of evidence, though, which speak for a different or “Greek” usage of fi, i.e., for directive and locative purposes. Example (2) shows the locative use of fi in Kormakiti Arabic.

(2) Usage of fi in Kormakiti Arabic

\[
\begin{aligned}
pn \quad t\llll \quad xad \quad p \quad ta-r \quad bar \quad fi-l-llk
\end{aligned}
\]

for SUBJ-find wood NEC SUBJ-go.1PL outside in-ART-field

“In order to find wood we go out to the field(s).”

Thus, it is possible to conclude that this category of Kormakiti Arabic has been influenced by both Standard Greek and colloquial Greek. The fact that all Maronites are bilingual in Greek and their education is in Greek, enabling them to master the colloquial variety as well as the standard language, makes this conclusion even more plausible. It can also be observed that this usage of fi is a newer development, as it was only found in the speech of one younger speaker. There is a third possibility that the usage of fi is related to influence of Standard Arabic or Lebanese Arabic. Some of the active members of the Maronite community (especially the ones from Kormakitis) learn some Lebanese Arabic due to their contacts with other Maronites in Lebanon. According to the metadata collected during fieldwork, the speaker in example (2) knows some Lebanese Arabic and when I inquired him about the usage of fi, he admitted that it is actually a Lebanese Arabic “word” and not a Kormakiti Arabic one, thus admitting that he formed this clause under the influence of Lebanese Arabic. This fact serves as an evidence that Kormakiti Arabic is not entirely isolated from the Levantine varieties, especially from Lebanese Arabic.

Furthermore, the influence of the Greek locative/directive marker se (usually has the clitic form [st + case/person/number inflection] and the whole construction is basically [se + article]) seems to have also influenced Kormakiti Arabic with its locative usage. Example (3) shows the directive usage of fi. It is an example from Borg (1985) produced (probably) by a native speaker of Kormakiti Arabic, i.e., not a semi-speaker.

(3) Borg (1985): “A story”

\[
\begin{aligned}
ist \quad vka \quad fi \quad \zre-y \quad ad \quad l-llshm
\end{aligned}
\]

then fall.PST.3SG in feet-1SG that ART-sergeant

“Then the sergeant fell at my feet.”
Since the Levantine Arabic varieties do not use fi for directive purposes, the usage of this morpheme for locative and directive marking can also be attributed to Greek. This phenomenon is interesting as both the informal and formal structures of Greek have affected Kormakiti Arabic, which can be explained due to long time bilingualism. One could also argue that these are effects of language death, i.e., the whole locative/directive system has collapsed and the speakers are using the Greek structures depending on whether they feel that they are speaking more formally or completely informally.

4.1.2. Imperative Inflection of “to come”. A very interesting case of borrowing are the imperative forms of the verb “to come” as described in Borg (1985: 91). The verb “to come” in Kormakiti Arabic can be expressed through the radicals žyy. Thus, for example, third person masculine past tense is inflected as žɑ (“he came”). The imperative forms of this verb are, however, borrowed from Greek, whilst incorporating them in the Arabic system. The imperative forms of “to come” in Greek are ela in singular and elate in plural. Ela is also the imperative of “to come” in Kormakiti Arabic, while its feminine singular form is elί, and its plural form is elu (without gender specification).

4.2. Language decay. Kormakiti Arabic is not only a moribund language; every piece of knowledge we have on this language originates from last three-four decades, when the language has already started dying out. Here, I shall illustrate those features of the language which show parallels with language decay of Arvanitika and discuss how these features are in fact related to language decay.

4.2.1. Loss of Systematic Integration. This is one of the most visible signs of language decay in Kormakiti Arabic. There are still lexical items which stand as a testament of a previous stage of language contact when the lexical borrowings were integrated into the Arabic system, e.g. the plural forms of nouns. Thus, the speakers borrowed the lexeme ksınari (“axe”) from Greek as kšınar by integrating it in an Arabic singular pattern and then derived its plural form using the same Arabic pattern, kšenir4 (Borg 1985: 69). Another example for this kind of lexical integration of nouns is the Greek lexeme kammini which became kammin in Kormakiti Arabic and the plural form is kmemīn. This kind of integration is apparently not the case anymore, as can be seen in example (4).

(4) Borg (1985): “A story”

\text{allik p-petrokopi, n-tammet l-ispiridkya ta}
\text{those ART-stonecutters PASS/MED-end.PST.3SG.F ART-matches REL}
\text{kan-yišelu fayyes}
\text{PROG.PST-set.fire.3PL dynamite.holes}

“When those stonecutters were igniting sticks of dynamite, the matches got used up.”

The lexemes petrokopi, ispiridkya and fayyes are all Greek and are embedded in the Kormakiti Arabic clause much like code-switching. All three lexemes are inflected in Greek with the appropriate plural morphemes. This clause is not from a recent field work but from Borg (1985), it means that it must have been recorded before 1985. If such examples should be considered a feature of language decay, then we must assume that language

---

4 Although Borg provides the Greek origin of the lexeme as ksınari, in Cypriot Greek it is in fact kšınari which explains the /ʃ/ phoneme in Kormakiti Arabic.
decay in Kormakiti Arabic began well before 1980s. This should not come as a surprise as the speakers whom I conducted interviews with stated that the language shift began in the 1950s. We do not know how proficient this speaker was, whether he/she was fully proficient or a rusty speaker in Sasse’s terms.

A weak point of this feature as an argument for language decay is the difficulty to distinguish between loss of systematic integration due to language decay and casual (or even systematic) code-switching. One can see various other signs of language decay in Kormakiti Arabic which would speak for accepting these examples as loss of integration, but this kind of argumentation is somehow circular. It is certainly important to consider whether these lexemes never had any counterparts in Kormakiti Arabic or whether their counterparts are loss. Another feature of language decay mentioned by Sasse could be referred at this point in order to explain the situation mentioned above, namely the fact that semi-speakers have difficulties in finding lexical items in their language.

The remaining speakers of Kormakiti Arabic often complain that many modern words such as car and computer do not exist in their language, which they see as a problem for the revitalization of the language. Should we also assume that the language never had the words for “stonecutter”, “matches” and “fire”? Even if we do, example (5) should make it rather clear that we can assume a systematic loss of lexical material in the language as well as loss of systematic integration.


\[
\begin{align*}
\text{kun-na-li-u} & \quad \text{d-dikasti} \quad \text{l-istoria} \quad \text{šait-na} \quad \text{u} \\
\text{say-PST-1PL-IO-3SG} & \quad \text{ART-judge} \quad \text{ART-story} \quad \text{POSS.F-1PL} \quad \text{and} \\
\text{kanyitxak} & \quad \text{u-v} \quad \text{uo} \\
\text{PROG.PST-laugh.3SG.M} & \quad \text{too} \quad \text{PRN.3SG.M} \\
\text{“We told the judge our story and he too had a good laugh.”}
\end{align*}
\]

It is quite understandable for the speakers of Kormakiti Arabic to have borrowed the lexeme dikasti since it is part of the administrative register that in the history of Cyprus never existed in Kormakiti Arabic. It is difficult to judge whether there is a phonological integration. The original Greek form in nominative singular would be dikastis, so we can observe a slight change in the lexeme, which is not always the case, since many examples from my own recordings do show other lexemes with the final /s/ intact. Concerning the lexeme istoria, one can state without doubt that it provides evidence for loss of lexical material, since it is difficult to imagine that Kormakiti Arabic lacked the word for story.

4.2.2. LOSS OF SUBORDINATIVE MECHANISMS. There are evidences of several unusual features in the subordinative system of Kormakiti Arabic and language contact must have had an influence on this aspect of the language. Whether one could speak of a loss in this case, requires further investigation of the language. For this reason, I will only present a short overview of clause subordination in Kormakiti Arabic.

Sasse mentions that the subordinated clauses in Arvanitika are usually short relative clauses, and the only adverbial clauses are the conditional and temporal ones. Relative clauses in Kormakiti Arabic are introduced by the relativizer ta and are quite productive. A quick search for the relativizer shows that relative clauses were used 63 times in my corpus of Kormakiti Arabic. I am not aware of any universal ratio of relative clauses per main clause and therefore it is difficult to measure whether this number is high, normal or low. In any case, relative clauses are being used in Kormakiti Arabic. The adverbial clauses are a
different case due to their formal nature. Though these clauses are often used in Kormakiti Arabic, they are often introduced by Greek adverbials, as can be seen in example (6), where the adverbial clause marker molis is used.

(6) Borg (1985): “A story”

\[
\text{molis} \quad \text{rka\-'at} \quad \text{i\-'ri} \quad \text{u} \quad \text{rux-t} \quad \text{ta-la\-'a} \quad \text{parra}
\]

as.soon.as hit-PST.1SG feet-1SG and go-PST.1SG SUBJ-get.out out

“Just as I stamped my foot and was about to go out.”

This is not an individual case or spontaneous code-switching as molis is used systematically in Kormakiti Arabic. There are in fact various Greek adverbials being used systematically in Kormakiti Arabic such as istera (“afterwards”) and amma (“when”), alongside with native adverbials. Even though many of the native adverbials were replaced by Greek ones, I argue that this is a case of massive borrowing and not of language decay, as subordination is still being used widely in Kormakiti Arabic speech. Even the complement clauses are widely used, although they are introduced by the Greek complementizer oti. As for the loss of subordinative mechanisms on the phrasal level, i.e., loss of modification and possessive constructions, this is simply not true for Kormakiti Arabic.

4.2.3. Breakdown of Grammatical Categories. Contrary to Arvanitika, the breakdown of grammatical categories is not the case in Kormakiti Arabic, since the TAM system is surprisingly almost completely intact. The tense and aspect consist entirely of native Arabic morphemes, whereas there are a few cases of loss in the modal system. One example is the necessitative modal verb prepi which is Greek and is inflected as in Greek when used for the third person. According to my research, it seems that the necessitative is only expressed this way.

The same can be said for agrammatism, since we could not observe the radical lack of proficiency and “the total disintegration of the morphological system” in the case Kormakiti Arabic.

4.2.4. Phonological Variation, Distortion, and Hypercorrections. This feature of Arvanitika cannot be observed directly in Kormakiti Arabic, as I could not observe any variation or distortion of any kind in my corpus. What could be considered a similar change is the complete adoption of Greek phonology, except for the retention of the Arabic ayin phoneme /\(/. The loss of its own phonetics and phonology can be interpreted as a kind of distortion, but it is a much more radical one.

4.2.5. Summary. The comparison above shows that language death in Arvanitika and in Kormakiti Arabic constitutes itself in different ways, although the context of language contact shows certain parallels. This summary illustrates that language shift and language decay in Kormakiti Arabic must be new phenomena; or at least we must assume that there was a period of language maintenance before the language shift set on. The main reason why these two dying languages are behaving differently from each other during language death and decay could be the semi-speakers. Sasse places the semi-speaker in the middle of his language decay theory. The semi-speaker is the reason why the language decay occurs in the first place and without the semi-speaker we would have language death without decay. The lack of semi-speakers could be the reason why we do not observe much
more language decay in Kormakiti Arabic. There are certainly speakers who can be categorized as rusty speakers, i.e., speakers with full proficiency who did not or do not speak Kormakiti Arabic for long periods of time due to several reasons, such as moving away from Kormakitis due to work or family. We would indeed expect from these rusty speakers to have problems retrieving the lexical items from their Kormakiti Arabic repository and using more and more Greek borrowings in their speech. There are, however, almost no type 2 semi-speakers of Kormakiti Arabic. The language shift happened very sudden and changed the balance of power for the languages in Kormakitis. The following quote is from an informant who went to school before the language shift has started. He explains his school years:

When we first went to school, we could not speak Greek. We did not know anything and we used to speak only Arabic. We could not speak Greek and our teacher was getting tired of us. He was strenuous with us until we learnt to speak Greek. So at home, with my family, we used to speak only Arabic. So it was very hard and every morning when we went there [to school], the teacher asked us – to teach us Greek – he asked us “what did you eat today?” And we had to answer in Greek. But we could not speak Greek and answered in Arabic.

The speakers of Kormakiti Arabic were becoming bilinguals after the school age, as they were going to the Greek schools. The language was being transmitted to the young speakers who were speaking it in the village. When language shift came and the language was not transmitted at home, the children did not have any other possibility of learning the language. This shift somehow hindered the emergence of semi-speakers. There are few passive speakers I met, who claim that they can understand a few Kormakiti Arabic but they cannot speak it. These passive speakers (and also other full speakers) usually perceive Kormakiti Arabic proficiency in a binary style. One is either fluent in it or one cannot speak it at all.

5. LANGUAGE REVITALIZATION EFFORTS AND DIFFICULTIES. After mentioning language death in the case of Kormakiti Arabic, I would like to briefly describe the efforts to revitalize the language. Since both government authorities in Cyprus are quite indifferent to the vitality of Kormakiti Arabic, Maronites from Kormakitis are organizing revitalization activities themselves. Their status under the European Charter for Regional or Minority Languages help them in getting financial support from the European Union for projects and for networking with other linguistic minorities in Europe, such as the Sámi people in Norway.

Today, Kormakiti Arabic is only taught in one private school – St. Maron’s school in Nicosia – in the form of afternoon classes. Last year, around 80 children were learning the language at this school. Although Kormakiti Arabic is recognized by the Republic of Cyprus under European Charter of Regional or Minority Languages, the classes cannot be included in the normal school curriculum, because the government requires the school to present learning materials for every subject.

There is a project called xki fi sanna (“speak in our language”) which is organizing language summer camps every year in Kormakitis. Children visiting this summer camp are divided in different age groups in which they learn and practice the language, and then do

5 Original in Kormakiti Arabic, translation by author
some performances to the village, which include for instance singing songs in Kormakiti Arabic or playing theatrical pieces which do not only portray the language but also the culture of the village.

These language revitalization efforts face numerous challenges. The largest obstacle is the fact that the young people with their origins in Kormakitis are not exposed to the language any more. They live in large cities in the southern Cyprus, which are predominantly monolingual in Greek, and they are raised in this environment. Sometimes even their parents cannot speak Kormakiti Arabic. Moreover, there are almost no economic or social advantages in speaking Kormakiti Arabic. Thus, many young people are not motivated to learn the language. It is also crucial to note that the language is not important for the speakers’ identity either. Maronites define themselves through their religion, Eastern Catholicism. This feature is enough the separate themselves from every other linguistic or religious group in Cyprus, and it is enough to mark their identity. Finally, the lack of institutional support for Kormakiti Arabic proves it even more difficult to revitalize the language.

6. RECAPITULATION AND FURTHER RESEARCH. Linguistic evidence and historical data on Kormakiti Arabic point out to a rather quick language death scenario for the language. The attitudes of the authorities and the speakers of the language pose many difficulties for language revitalization efforts. As this minority language is most probably not going to survive for much longer, it is necessary to act rapidly in order to fully document it. Although there are several books and papers on Kormakiti Arabic, audio recordings are rare and there is no linguistic documentation based on a multimedia (audio and video) collection of primary data. The researches done until today have not yet exhausted the interests of linguistics. There is, for example, very little research on language acquisition and speaker behaviors in Kormakiti Arabic. As stated on the website of the NGO xki fi sanna, Kormakiti Arabic is still a “goldmine” for ethnographic and linguistic research.6

REFERENCES


Ozan Gulle
Ozan.Gulle@lip.lmu.de
Identity and language shift among Vlashki/Zheyanski speakers in Croatia

Zvjezdana Vrzić and John Victor Singler

"New York University, b University of Rijeka

The language Vlashki/Zheyanski, spoken in two areas – the Šušnjevica area and Žejane – of the multilingual, multiethnic Istrian peninsula of Croatia, evinces strong loyalty on the part of its elderly speakers, yet in both areas a language shift to Croatian is well underway. Vlashki/Zheyanski is a severely endangered Eastern Romance language known in the linguistic literature as Istro-Romanian. In order to study the domains and frequency of use of the language and equally to examine speaker attitudes about language and identity, we administered a questionnaire to speakers in both locations. Our sample included responses from individuals in four age groups. Our discussion here focuses on 16 men and women from the two older groups, 51–70 and 71-and-older. In Žejane, speakers saw knowledge of the language and family lineage as defining components of being a “real” member of the community. The name for the language, Zheyanski, comes from the village name. Hence, someone who speaks the language asserts that village belonging and village affiliation are at the core of speakers’ identity. In terms of national identification, whether Croatian, Italian, and/or Istrian, Zheyanski speakers by and large showed little enthusiasm for any of the three choices. In terms of language use, all respondents continue to use the language on a daily basis but report that they speak mostly Croatian to their grandchildren. In the Šušnjevica area, people used the same criteria, language knowledge and family lineage, to define group membership and feel close affiliation to their home village. Unlike in Žejane, the name of the language, “Vlashki”, does not correspond to a unitary group name accepted and liked by all. In terms of larger identity, villagers embraced identities that they share with their Croatian-speaking neighbors: Most felt “extremely Istrian”, and at least “fairly Croatian”. The language shift to Croatian is also more advanced here: All the speakers report speaking mostly Croatian to their children. While speakers in both Žejane and the Šušnjevica area endowed their language with a critical role in their identity, this attitude toward Vlashki/Zheyanski does not manifest itself in their communication with younger generations where other social forces have caused the shift to the use of Croatian.
1. INTRODUCTION. Vlashki/Zheyanski is a severely endangered Eastern Romance language, spoken in two different locations on the Istrian peninsula in Croatia, known in the linguistic literature as Istro-Romanian because of its historical connections to Romanian. The terms Vlashki and Zheyanski are the speakers’ own names for two geographically distinct varieties that are fully mutually intelligible but also easily identifiable as different by their speakers.

In this paper, we report on the results of a sociolinguistic questionnaire study designed to examine the degree of language endangerment and shift in Vlashki/Zheyanski-speaking communities and speaker attitudes about language and identity.

1.1. LOCATION OF THE Vlashki/Zheyanski-SPEAKING COMMUNITIES. The speakers live in the northeastern part of Istria, a historic region located in the northwest of the Republic of Croatia, between the Gulf of Trieste and the Kvarner Gulf in the Adriatic Sea. Speakers of the Zheyanski variety live in the isolated mountain village of Žejane close to the Croatian border with Slovenia. Further south and across the mountains from Žejane, speakers of the Vlashki variety inhabit five villages a short distance from one another, lining the northern portion of the Čepić Valley at the foot of the Učka Mountain. The largest village of the five is Šušnjevica. The two places, Žejane and the Šušnjevica area, are now divided by an administrative provincial border.

The mountainous area in the north has always been difficult to access directly from the valley further south. The Šušnjevica area and Žejane are around fifty kilometers apart and there are now good roads around the mountain connecting them. However, the two groups of speakers seem to have had little if any interaction in the course of their history in Istria. Today, as in the past, residents of the Šušnjevica area and Žejane are oriented toward different urban centers for work, business, and schooling.

1.2. DEMOGRAPHICS. According to the 2011 census, there are 406 people living in the villages where Vlashki/Zheyanski is spoken. Of the 406 residents of the villages, according to our estimate, around 120 are fluent and active speakers of the language.

In Žejane, there are 130 inhabitants, including 45 active and fluent speakers, close to 35 percent of the village population. In the Šušnjevica area, 276 inhabitants live in the five villages, including roughly 75 fluent and active speakers, a bit more than a quarter of the population of these villages.

Among the speakers, the great majority are over fifty years old. These older speakers are fluent, balanced bilinguals in Vlashki/Zheyanski and Croatian. Many speakers who started school in the 1950s, and even the 1960s, report having been monolingual in Vlashki/Zheyanski before school. Younger speakers are typically Croatian-dominant, and many learned Vlashki/Zheyanski as a second language.

There seem to be few, if any, fluent and active speakers of the language in the population under the age of 25, hence few if any among preschool and school-age children. In Žejane, according to the count done in 2009, there were only six speakers between the ages of 25 and 50 years old and there were no fluent and active speakers under the age of 25.

We estimate that there are additional 450 speakers of Vlashki/Zheyanski elsewhere in Croatia, primarily in neighboring towns and cities such as Matulji, Opatija, Rijeka, Kršan, Labin, Pazin, and Pula. Outside Croatia, there may be another 400–500 speakers, primarily in the United States, especially New York City, and in western Australia. In all, there are roughly 1,000 speakers in the world, and the language is severely endangered.
1.3. External social causes of language shift. Historically, the number of people in the Vlashki/Zheyanski-speaking villages has been small, with village size similar to the size of other villages in the area. In the half-century before World War II, most village families lived on subsistence agriculture and sheep herding but complemented these traditional occupations through some family members engaging in small businesses, industrial work, and service jobs in local towns and cities, such as Opatija, Rijeka, Pula, and Trieste. For example, men in Žejane prepared and sold charcoal and wood for heating, many of those in the Šušnjevica area were miners and sailors, and young women from both locations worked in private homes.

Beginning in the late nineteenth century, there was some movement of people from the villages to the city as well as immigration to foreign lands. In general, however, population growth in the villages in the late nineteenth and across the early twentieth century was large enough to offset the population losses that occurred during the years of the two world wars. Thus, the 1880 Austrian census lists the total population in the Šušnjevica area and Žejane as 2467 (Naselja i stanovništvo RH 1857–2001), and a 1945 Yugoslav regional census gives it as 2255.

In sharp contrast, immediately after World War II, massive depopulation of the villages began, and it has continued virtually to the present day. The most recent census, taken in 2011, shows the population of the villages to be less than one-fifth of what it was in 1945. The depopulation, specifically outmigration abroad, began immediately after the war. It was part of a significant depopulation of the Istrian peninsula more generally. In two large waves, most Italian Istrians, but many others too, left the region in reaction to the political and social changes that followed the war, when Istria was joined to other Croatian-speaking territories and became part of socialist Yugoslavia. Already within the first eight years after the end of World War II, the villages’ population had shrunk by more than a quarter.

Additional factors in post-WWII outmigration involved the socio-economic processes of modernization, industrialization, and urbanization, in ways parallel to what took place in other areas of Central and Eastern Europe (e.g. Gal 1978). The arduous agricultural lifestyle in the rather poor northeastern area of Istria has become increasingly devalued and has gradually been abandoned. Increasingly, younger villagers have sought more lucrative employment, mostly in industrial and service jobs, and have chosen to live outside of the villages. The institution of universal elementary education in Croatian as well as free and more accessible high school education has provided young adults with the resources for greater social mobility. Further, the greater mobility, coupled with regional depopulation, has been accompanied by a more widespread practice of intermarriage, not only among those villagers who left, but also among those who stayed.

Modernization processes have led to the gradual rejection of a very local and largely self-sufficient agricultural lifestyle. The Vlashki/Zheyanski-speaking villages have lost the geographical and cultural isolation that supported the almost exclusive use of the local language by most villagers. In the process, a range of new social domains opened up in which Croatian, not Vlashki/Zheyanski, was useful, even necessary. Previously, the Catholic church had been the main site of exposure to the majority Croatian language. Education, in Croatian or – between the two world wars – in Italian, was not widely available. Now, education became widespread, employment outside the village a norm, and new media – such as newspapers, radio, and television – omnipresent.

---

2 See also Orbanić (1995: 60) and Filipi (2003: 88), whose numbers are slightly lower as they do not include the numbers for one of the villages.
1.4. HISTORY OF THE LANGUAGE AND THE AREA. It has not been established beyond doubt when Vlashki/Zheyanski separated from Proto-Romanian and indeed whether it is a sister or a daughter of the Daco-Romanian branch of Proto-Romanian (Kovačec 1998: 242–244, Frâțilă & Sârbu 1998: 13–19, but cf. Niculescu 1990: 59–70). However, most evidence points to its speakers’ migration away from the rest of Proto-Romanian speakers at some point in the second half of the first millennium. This was prior to the beginning of Hungarian linguistic influence on Daco-Romanian (Frâțilă & Sârbu 1998: 13–17, Niculescu 1990: 67, Mallinson 1990: 303). While Daco-Romanian has many lexical borrowings from Hungarian, Vlashki/Zheyanski has none.

There is greater consensus that at the beginning of the sixteenth century speakers of today’s Vlashki/Zheyanski settled in Istria, specifically in the places where their descendants live today (Kovačec 1998: 242–244). These settlers migrated there from northern Dalmatia, in the same historical period when large numbers of Croatian-speaking people were also moving from Dalmatia to Istria.

The original settlers migrated into an area in central and northeastern Istria, which was then part of the Holy Roman Empire and ruled by the Habsburgs. The larger portion of the Istrian peninsula was under the control of the Venetian Republic at the time. After the fall of Venice in 1797, the entire territory of Istria became part of the Austrian Empire and then Austria-Hungary from 1867 until 1918. After World War I, Italy annexed Istria. The region then remained part of Italy until 1943. After World War II, Istria was placed within the borders of socialist Yugoslavia, specifically within the Republic of Croatia. This was the first time in Istria’s history that the region and its majority Croatian speakers were in the same state with other Croatian-speaking territories. Since 1991, Istria is part of the independent Republic of Croatia.

As this description makes clear, Istria has been a political borderland throughout its history. Only around 60km at its widest point and 120km at its longest, Istria has been crisscrossed by political borders and has been fought over by a succession of rulers and states. Today Vlashki/Zheyanski speakers, like other people in Istria, speak of four generations of their families born in four different states – Austria-Hungary, Italy, Yugoslavia, and Croatia – without ever having left home. Thus, different generations of the same family can easily claim different, as well as multiple, national allegiances. Istria has also been a cultural meeting ground, one characterized by a multilingual and multiethnic population throughout most of its history. Croatian, Italian, and Slovenian speakers have been its main ethnonlinguistic components for centuries. Today, Croatian speakers are the large majority in Istria but there is also a significant population of Italian speakers. Overall, however, as previously noted, the population in Istria today is smaller than it was 100 years ago (Cukrov 2001: 30).

Throughout their five centuries in Istria, Vlashki-speaking and Zheyanski-speaking settlers have shared their history yet have had little contact with one another. Over the course of these centuries, their villages have been part of the same larger political entities, even when they were in different smaller administrative units within them. Today, Vlashki/Zheyanski speakers live in two different administrative regions of Croatia: Vlashki speakers live in the Šušnjevica area, which is in the Istrin Region. Zheyanski speakers live in the village of Žejane, which, while it is geographically on the peninsula of Istria, has been included in the neighboring Primorsko-Goranska Region since the 1990’s. The Istrin regional government promotes political regionalism and officially supports multiculturalism and multilingualism. Two of the languages spoken in the area – Croatian and Italian – are co-official. In the Primorsko-Goranski Region, Croatian is the official language.
State schooling dates to the 1880’s in Žejane, the northern village, and to the beginning of the twentieth century in Šušnjevica (Beltram & Jakovljević 2005, Legac 1983). Prior to that, some villages did have small parochial schools. Formal education before World War I was not extensive, and it was in Croatian. Beginning in 1923, at a time when Istria was a part of Italy, the Fascist government brought schooling to all the Vlashki-/Zheyanski-speaking villages and made Italian the language of instruction for all children. During the Italian period, most children attended school for five years. In 1945, with Istria now part of socialist Yugoslavia, Croatian became the language of instruction once again. In 1951, the government extended universal education to eight years (Beltram & Jakovljević 2005). In sum, education went from being sporadic to regular and universal, and the number of years that everyone went to school jumped from a few years to five to eight. The medium of instruction in schools went from Croatian to Italian and then back to Croatian.

Schools have provided no institutional support for Vlashki/Zheyanski, and neither has the Catholic Church, with all church activities – apart from the Latin Mass prior to the 1960’s – being carried out in Croatian. In addition, Vlashki/Zheyanski speakers have always used their language in spoken situations only, as any written document was in Croatian or, before World War II, in Italian.

2. The sociolinguistic questionnaire and the sample. The goal of our study was to examine domains and frequency of use of Vlashki/Zheyanski as well as speaker attitudes about the language and its relation to their identity. Our questionnaire consisted of 105 questions divided into four sections: personal details, domains and frequency of language use, language competence in Vlashki/Zheyanski, Croatian, and Italian, and language attitudes and identity. In addition to questions for which the respondent chose from a set of answers, we had several open-ended questions that enabled speakers to elaborate on their views. We prepared the questions in Croatian, and fluent speakers of Vlashki/Zheyanski translated the questionnaire into Vlashki/Zheyanski and subsequently administered it early in 2013.

We divided the study participants into groups based on residence, age, and gender. We had 16 participants from each of three locations - Žejane, the Šušnjevica area, and New York City – for a total of 48. For a given location, we took four speakers from each of the following four age groups: 71 and over, 51–70, 31–50, and under 30, with an equal number of female and male participants in each age group. The study participants were native speakers of Vlashki/Zheyanski and completed the questionnaire in their native language; the only exceptions arose because the community members administering the questionnaire were unable to find the requisite number of fluent speakers in the youngest age group. In these cases, the questionnaire was administered in Croatian if in Croatia or in English if in New York.

In this paper, we look at a subset of the total number of questionnaires. Specifically, we focus on 16 participants, namely the two older groups, all native Vlashki/Zheyanski speakers, in the two locations in Croatia.³

³ In examining the responses that people gave us, we have sometimes felt stymied by the small number of respondents for a given age group in a particular place. We had to remind ourselves that in fact our sample represents a considerable portion of the Vlashki- and Zheyanski-speaking populations of the native villages.
2.1. **Respondents’ Attributes.** Table 1 presents the 16 participants in our study.

<table>
<thead>
<tr>
<th>Variety</th>
<th>Sex</th>
<th>Age</th>
<th>Years of Schooling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vlashki</td>
<td>F</td>
<td>84</td>
<td>3</td>
</tr>
<tr>
<td>Vlashki</td>
<td>M</td>
<td>78</td>
<td>6</td>
</tr>
<tr>
<td>Vlashki</td>
<td>F</td>
<td>76</td>
<td>5</td>
</tr>
<tr>
<td>Vlashki</td>
<td>M</td>
<td>75</td>
<td>6</td>
</tr>
<tr>
<td>Vlashki</td>
<td>M</td>
<td>61</td>
<td>12</td>
</tr>
<tr>
<td>Vlashki</td>
<td>F</td>
<td>61</td>
<td>12</td>
</tr>
<tr>
<td>Vlashki</td>
<td>M</td>
<td>60</td>
<td>8</td>
</tr>
<tr>
<td>Vlashki</td>
<td>F</td>
<td>51</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variety</th>
<th>Sex</th>
<th>Age</th>
<th>Years of Schooling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zheyanski</td>
<td>M</td>
<td>83</td>
<td>5</td>
</tr>
<tr>
<td>Zheyanski</td>
<td>F</td>
<td>77</td>
<td>5</td>
</tr>
<tr>
<td>Zheyanski</td>
<td>M</td>
<td>76</td>
<td>5</td>
</tr>
<tr>
<td>Zheyanski</td>
<td>F</td>
<td>73</td>
<td>6</td>
</tr>
<tr>
<td>Zheyanski</td>
<td>F</td>
<td>69</td>
<td>11</td>
</tr>
<tr>
<td>Zheyanski</td>
<td>F</td>
<td>65</td>
<td>8</td>
</tr>
<tr>
<td>Zheyanski</td>
<td>M</td>
<td>64</td>
<td>12</td>
</tr>
<tr>
<td>Zheyanski</td>
<td>M</td>
<td>54</td>
<td>12</td>
</tr>
</tbody>
</table>

**Table 1: Respondents**

Subsequently, we refer to respondents on the basis of the variety they spoke, their sex, and age. Thus, ZF77 refers to the 77-year-old Zheyanski-speaking female respondent.

Whether from Žejane or the Šušnjevica area, the sixteen participants have similar backgrounds. With regard to employment, all but one of the speakers held jobs outside the home, women as well as men. Two women from the younger group in the Šušnjevica area worked in offices; the other 13 jobholders all had blue-collar positions.

When there are differences, they are most often based on age. The older age group consists of people 71 and over, i.e. born between 1929 and 1940. Each person in that group has between three and six years of elementary education. It is only with the “younger” age group, those 51 to 70, that more extensive education took place. Thus, six of the eight respondents in this group have a high school education. In each location, the oldest person attended school in Italian only, while the second oldest started school in Italian but then continued in Croatian after World War II. The rest of the people in the older age group and everyone in the younger age group attended school in Croatian.

All 16 respondents were child bilinguals and fluent in Vlashki/Zheyanski and Croatian. Four speakers report having been monolingual in Vlashki/Zheyanski before they started school.

We asked respondents to grade their skills on a scale of 1 to 5, and we indicated that 5 was “best”. This use of a 1-5 scale taps into local educational practice.

All 16 speakers rated their skills in speaking and understanding Vlashki-Zheyanski as a 4 or a 5, usually a 5, and all but one evaluated their Vlashki-Zheyanski proficiency as equal to or greater than their Croatian proficiency.

In addition, 15 of the 16 respondents say that they understand some Italian, and 12 of them say that they speak some Italian; however, most respondents – all but three – rate their Italian much lower than their Vlashki/Zheyanski and Croatian.

2.2. **Patterns of Use of Vlashki/Zheyanski.**

2.2.1. **Frequency of Use of Vlashki/Zheyanski and Croatian.** In order to get a sense of the patterns of use of the Vlashki/Zheyanski among the speakers of the two oldest age groups and examine the level of language shift to Croatian in the two locations, we looked both at the amount of the Vlashki/Zheyanski use relative to Croatian and the choice
of language made by speakers in interaction with different interlocutors and in different social domains.

We asked the respondents to report on how often they use Vlashki/Zheyanski and Croatian at home and in the village. As noted above, the respondents in both locations are native speakers of Vlashki/Zheyanski. Moreover, with the exception of one speaker in each location, all of the respondents are active users of the language. They report using it every day both at home and in the village.

While the frequency of use of Vlashki/Zheyanski in home and in the village is essentially comparable in the Šušnjevica area and in Žejane, there are important differences in the status of Croatian, specifically in the home. Seven of the eight respondents from the Šušnjevica area report using Croatian too on a daily basis at home (and the eighth uses it “sometimes”). In contrast, only three of the respondents from Žejane report using it daily; two others use it sometimes, and the remaining three use it rarely. In the village domain, the two locales are comparable, with half of the respondents in each place using it daily and the others using it much less.

We also asked speakers to report on a change in the frequency of their use of Vlashki/Zheyanski and Croatian in the course of their lifetime (“now” versus “before/in the past”). They report using Vlashki/Zheyanski as often today as they did in the past. (The lone exception is the Zheyanski speaker mentioned in footnote 3 whose wife is a Croatian speaker. He reports that in the past, presumably prior to his marriage, he routinely used Zheyanski in the house, whereas now he rarely uses it in that domain.) If the frequency of occurrence of Vlashki/Zheyanski has not changed, it stands in sharp contrast to Croatian. Specifically, Croatian is now present on an everyday basis within the communities. Its role is far greater in both domains than was the case in the past: In the Šušnjevica area, for example, of the eight respondents, only one person reported using Croatian on a daily basis in the village in the past. (She is discussed in footnote 3.) All the others report having used it rarely or never. Now, however, five of the eight respondents there say that they use Croatian every day in the village.

2.2.2. LANGUAGE CHOICE WITH DIFFERENT INTERLOCUTORS. So that we could establish how the language choice is made between Vlashki/Zheyanski and Croatian according to the participants, we asked speakers to indicate their language choice with regard to several types of interlocutors – different close family members including spouses as well as relatives, friends, and neighbors. They also indicated their language preference with respect to other specific social domains: work, school, everyday life (such as shopping), at village social events, and at church-related events in the village. Here we limit ourselves to respondents’ reports of their interaction with close family members.

In answering the question “What language do you speak to ___?”, a respondent had five choices to which we later assigned numerical values:

(5) Only Vlashki/Zheyanski
(4) Mainly Vlashki/Zheyanski with some Croatian

The two exceptions are noteworthy. The youngest respondent from the Šušnjevica area reports that, when she was growing up, she spoke both Croatian and Vlashki to her parents, but more Croatian. She reports further that she spoke only Croatian to her older sibling(s). In Žejane, one of the respondents in the 51–70 group reports that he is married to a Croatian-speaking woman and rarely speaks Zheyanski at home. In his case, however, his language behavior in the village is the same as for others of his generation, i.e. he reports using Zheyanski every day.
(3) Vlashki/Zheyanski and Croatian about equally

(2) Mainly Croatian with some Vlashki/Zheyanski

(1) Only Croatian

We established above that the home domain is not reserved exclusively for Vlashki/Zheyanski in either location, but also that in the Šušnjevica area the majority of speakers report an equally frequent use of Vlashki/Zheyanski and Croatian, while in Žejane Croatian is used less than Vlashki/Zheyanski at home. A look at respondents’ language choices with different family members provides corroborating evidence for this geographic difference.

We asked respondents about language use with a range of relatives, from previous generations (parents and grandparents), their own generation (siblings), and subsequent generations (children and grandchildren). In both the Šušnjevica area and Žejane, regardless of the respondent’s relation to the relative: If the relative was in the same or an older generation, the usual answer was “Vlashki/Zheyanski only”.

The exceptions merit discussion. While everyone reported speaking only Vlashki/Zheyanski to their paternal grandparents, three respondents stated that they spoke only Croatian to their maternal grandparents. Presumably these are all instances where the respondent was the child of intermarriage. In one case, the respondent (VM60) nevertheless reported speaking only Vlashki to his other relatives, including his mother, and in a second, the respondent (ZF69) reported speaking mostly Croatian to her mother but only Vlashki to her father. In the third case, the respondent (VF51) spoke mostly Croatian to both her parents and only Croatian to her older brother(s). This last respondent is the youngest person in the groups under study in this sample; as such, her age may also be a factor in her shift to Croatian.3

Overall, then, when the respondents were growing up, Vlashki/Zheyanski dominated in their homes and villages. For the respondents to address elders or even siblings in a language other than Vlashki/Zheyanski was exceptional. However, the Vlashki-speaking respondents show a difference with regard to addressing their spouses and, especially, their children. Four of them report speaking to their spouse only or mostly in Vlashki, but three others report Croatian instead, two “only” and one “mostly”. Further, no matter which language respondents used in speaking with their husbands and wives, their primary language in addressing their children was Croatian, either primarily so (four instances) or exclusively so (three instances).

<table>
<thead>
<tr>
<th></th>
<th>VF 84</th>
<th>VM 78</th>
<th>VF 76</th>
<th>VM 75</th>
<th>VF 61</th>
<th>VM 61</th>
<th>VF 60</th>
<th>VF 51</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spouse</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Children</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: The Šušnjevica area: Respondents’ language choice with spouse and children

---

3 If respondents did not have contact with particular relatives or not did not have relatives of a certain type (e.g. younger sister), they did not answer the question. Apart from the speakers noted above, the only respondents who report speaking to their mother or father in any way other than “only Vlashki/Zheyanski” are two people who did not provide an answer in the box for language spoken to maternal grandparents, thereby leaving open the question as to whether or not the two respondents’ mothers (and their mothers’ parents) were Croatian-speaking. Of the two, ZF65 spoke mostly Zheyanski to her parents, and ZM54 spoke Zheyanski and Croatian equally to his.
In Vlashki, then, the respondents form the pivot generation, speaking Vlashki overwhelmingly to their parents but Croatian overwhelmingly to their children.

In Žejane, language use with spouses displays parallels to the findings in the Šušnjevica area. All four older speakers still report speaking only Zheyanski to their husbands and wives. With the 51 to 70 group, however, there is a shift in progress to Croatian. Crucially, though, respondents’ reported language choice when speaking to their children is much more conservative, with Croatian obtaining only for the two youngest respondents. (One of the respondents, ZF69, reports speaking only Croatian to her husband and only Zheyanski to her child(ren).)

An examination of parent-child interaction in the Šušnjevica area shows that language choice is leading to language shift. The parents may have spoken only Vlashki to their parents, but now they speak Croatian to their children. In Žejane, on the other hand, most parents still speak to their children in Zheyanski (at least most of those in our sample do). The difference between the two sites is not whether or not shift will occur, but rather a difference as to when it will occur – or has occurred. For in terms of what language respondents report speaking to their grandchildren, Croatian now prevails everywhere.

Even with the shift, however, the two speaker groups are not quite identical in terms of their language choice with grandchildren. In the Šušnjevica area, four of the five respondents with grandchildren report speaking only Croatian to them; there is a lone respondent there who reports using mostly Croatian with his grandchild(ren) rather than only Croatian. In contrast, in Žejane, one respondent reports using Zheyanski and Croatian equally, and all six of the others with grandchildren report using mainly Croatian – but no respondent in Žejane reports using Croatian exclusively.
2.2.3. Identity.

National identity, regional identity, and “Istrianness” among Vlashki/Zheyanski speakers. Vlashki/Zheyanski speakers are a part of a linguistic enclave that does not hold official minority status in Croatia. Almost all of Croatia’s 22 officially recognized minorities are national minorities related to populations elsewhere, usually in neighboring national states (Tatalović 2005).

One such minority group is Italians, who are the largest group in Istria apart from Croats. The question arises as to how Vlashki/Zheyanski speakers frame their identity. Because of the origin of their language, Vlashki/Zheyanski speakers have often been associated with a Romanian identity. However, for reasons of history and geography, the speakers themselves have not embraced Romanian national affiliation (Vrzić 2010), as we also show below.

In addition to individuals choosing national identities such as Croatian, Italian, and Slovenian on censuses, in the Yugoslav Regional Census of 1945 a small proportion of the population declared their identity in terms of regional affiliation, instead of the national one. In the Šušnjevica area, the proportion of the population who chose regional, i.e., Istrian, affiliation was small before 1991. In 1991, however, the number of people declaring regional affiliation soared to nearly half the population in the Šušnjevica area. As can be seen in Table 6, residents of Žejane did not choose this option. Most respondents there, like most of those in the Šušnjevica area who did not opt for regional affiliation, declared Croatian identity.

It should be noted that the number of people who opted for a regional affiliation on the 1991 census in the Croatian-speaking villages and towns neighboring to Šušnjevica was also very high, but not so high as in the Šušnjevica area itself.

<table>
<thead>
<tr>
<th></th>
<th>Total population</th>
<th>Regional affiliation</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Šušnjevica area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1945</td>
<td>1596</td>
<td>64</td>
<td>2.9</td>
</tr>
<tr>
<td>1981</td>
<td>463</td>
<td>26</td>
<td>3.6</td>
</tr>
<tr>
<td>1991</td>
<td>408</td>
<td>184</td>
<td>45.1</td>
</tr>
<tr>
<td>Žejane</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1945</td>
<td>651</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1981</td>
<td>250</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>189</td>
<td>8</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Table 6: Census respondents who opted for “regional affiliation” rather than a national identity (such as “Croatian” or “Italian”) in the 1945, 1981, and 1991 censuses

The 1991 census was an important one in Croatian history. It coincided with the establishment of the independent state of Croatia, following secession from the Yugoslav federation, and with a period of significant political turmoil and armed conflict in Croatia and other former Yugoslav republics. It was significant in that it showed, in a rather dramatic way, the importance of regional belonging in Istria, which was boosted by a resistance to the homogenizing Croatian nationalist political ideology of the period. Including all of Croatian Istria – both the Istrian administrative region and the Istrian counties and towns of the neighboring Primorsko-Goranska administrative region – over 37,000 people, more

---

6 While we address the entirety of Vlashki and Zheyanski speakers, we note that the groups have separate histories with little interaction in modern times.
than 16 percent of the population, opted to express a regional identity in 1991 (Duda 2005: 758).

While the proportion of people selecting regional affiliation on censuses in Istria as a whole has gone down since the politically turbulent time of the 1991 census, Istrian regional identity is still represented by significant numbers in the area. On the 2011 census, more than 12 percent of the people (over 25,000) reported regional affiliation instead of Croatian or another nationality in the Istrian administrative region, making Istrian identity the second most represented identity group in the Istrian Region after Croatian.

The phenomenon of regional identity in Istria or “Istrianness” has been studied closely by several different social scientists (Ballinger 2004, Banovac 2004, Sujoldić 2008, Cocco 2010 among others) since the 1990’s. Allegiance to Istrian regional identity is found among people of all ethnic backgrounds, but mostly among Croatian speakers (Cocco 2010: 14). While the high number of people who opted for regional identity on the 1991 census was interpreted as resistance to the policies of social and cultural homogenization being propagated by the Croatian nationalist government at that time (Cocco 2010: 14ff.), Istrianness is recognized to have deeper and more complex roots.

Regionalism in the area is a political stance that goes beyond simple opposition to the extremes of Croatian nationalism; to an even greater extent, it is a cultural identity of people who have lived on the political border and in a multicultural environment for generations and who wish to protect the specific “social, cultural and economic features of the border region“ (Cocco 2010: 17). As such, Istrianness needs not be construed as mutually exclusive with a Croatian national identity (Banovac 2004). Rather, Istrianness can be seen as an expression of “the progressive values of the Western democracy […] opposed to the military hostility and the violent ethnic politics of the former Yugoslavia, which are perceived as detrimental to the local context of hybridism” (Cocco 2010: 17–8, reporting on Šantić 2001). In a similar vein, Ballinger (2004) talks about “Istrianness“ as a hybrid identity and about Istria as a historic borderland. A similar view is expressed by Sujoldić (2008), who frames Istrian identity as multicultural and inclusive. Scholars’ attention to Istrianness comes as a response to its expression on the censuses, especially the 1991 Croatian census. However, already in the nineteenth century Austrian ethnographers were discussing the existence of a population of mixed ethnicity and a hybrid cultural and supra-ethnic/national identity in Istria. They used the terms Verschmelzung, or melting, and Hybridismus, or hybridity, to refer to it (Nikočević 2008: 68, 150, 184).

National and Regional Identity of the Vlashki/Zheyanski Speakers in Our Study. We asked our respondents to answer the following three questions with regard to their group identity: “How Istrian do you feel?”, “How Croatian do you feel?”, “How Italian do you feel?”, as these are the identity designations that are in dominant use in the region and on the censuses. We allowed the respondents to grade their sense of allegiance to them on a five-point scale: “Not at all” / “A little” / “Fairly” / “Very” / “Extremely”. In displaying the results, we present two sets of tables. First, in Tables 7, for the Šušnjevica area, and 8, for Žejane, we use a check mark to indicate a positive response to the identity question, without taking into account the strength of the respondent’s allegiance.

As Tables 7 and 8 show, most respondents (13 out of 16) expressed allegiance to more than one identity, with two speakers claiming all three identities. With three possible identities to choose from and respondents free to choose as many as they pleased, there were seven possible patterns. The sixteen respondents chose six of the seven patterns, with “Italian only” the only one not selected.
In all, 14 of the 16 respondents answered that they felt Croatian (the exceptions being the oldest speaker in each location), eleven said that about Istrian, and six about Italian. The most common pattern (seven out of 16) was allegiance to Istrian and Croatian identity but not Italian.

As a comparison of Tables 7 and 8 shows, the two locations show only minimal differences with respect to the identity choices expressed by respondents. However, respondents at the two locations differ dramatically in the strength of allegiance that they express. This can be seen by comparing Tables 9 and 10 on the next page.

The Vlashki-speaking respondents expressed greater affiliation across the board: Of their sixteen positive responses, fully twelve were “very” or “extremely”. In contrast, of the Zheyanski speakers’ fifteen positive responses, only three were “very” or “extremely”.

**LOCAL AND ETHNIC IDENTITY.** After we had elicited from respondents how Istrian, then Croatian, then Italian they felt, we asked whether these terms (specifically, the ones to which they had responded positively) were sufficient to describe their group identity. In each location, five out of the eight respondents said yes. When asked for other terms, both those who said that the terms were not sufficient and those who found them sufficient then offered other descriptive group identity terms.

The terms that were volunteered differed between the Šušnjevica area group and the Žejane group. All Zheyans mentioned the term “Žejanci”, i.e. “Žejane people”, as a necessary identity term in addition to the national and/or regional designations. One person in Žejane also mentioned the term “Romanian”. In contrast, among the Šušnjevica area respondents, there was more variety. Two people mentioned terms derived from the village names, four people mentioned “Vlach” and one person offered the term “Roma-
### Table 9: Vlashki-speaking respondents: The strength of their allegiance

+1 = A little. +2 = Fairly. +3 = Very. +4 = Extremely

<table>
<thead>
<tr>
<th>Respondents</th>
<th>How Istrian do you feel?</th>
<th>How Croatian do you feel?</th>
<th>How Italian do you feel?</th>
</tr>
</thead>
<tbody>
<tr>
<td>VF84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VF76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VF61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VF51</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 10: Zheyanski-speaking respondents: The strength of their allegiance

+1 = A little. +2 = Fairly. +3 = Very. +4 = Extremely

<table>
<thead>
<tr>
<th>Respondents</th>
<th>How Istrian do you feel?</th>
<th>How Croatian do you feel?</th>
<th>How Italian do you feel?</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZM83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZF77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZM76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZF73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZF69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZF65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZM64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZM54</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
nian/Vlashki Istrian”. It should be noted that no respondent mentioned in this context the

group term “Čiribirci,” which is widely used to refer to Vlashki speakers by the surrounding
Croatian-speaking populations.

Further open-ended questions asked respondents to provide their opinions about the
terms “Vlach”, “Istro-Romanian” and “Romanian”, which are also often used by outsiders.
The name “Vlach” has been reported in the literature as speakers’ own collective term in
the Šušnjevica area in earlier times (Kovačec 1998: 235). The term is one used by Slavic
populations of the Balkans to refer to several different groups in the area who speak Eastern
Romance languages related to Romanian; moreover, for centuries it has been the traditional
outsider term in literature for these groups (Skok 1973: 606–608, Mirdita 2007).

As for “Istro-Romanian”, it is an outsider term commonly used in linguistic and other
literature, as well as in the media, to refer to the speakers of Vlashki/Zheyanski. It was
introduced into usage by nineteenth century linguists. The term has since gained wider
currency in the literature and the media. In the last few decades, the term has become
better known to the speakers themselves as it has been used in the Croatian media as well.
The Šušnjevica area and Žejane groups differed in their reactions to these terms. Three
people from the Šušnjevica accept the term “Vlach” and all of them find the term “Istro-
Romanian” acceptable to a degree. Most (seven out of eight) express their acceptance of the
term by opposing it to some other term they like less: Three speakers find “Istro-Romanian”
preferable to the term “Vlach”, which they consider somewhat derogatory as well as non-
unique, as it is used for other populations in the area and beyond it; four respondents find
“Istro-Romanian” preferable to the term “Romanian”, which they associate with a different
language and group.

Respondents from Žejane unanimously reject the term “Vlach” and three explicitly re-
ject both “Istro-Romanian” and “Romanian”. For five respondents, the latter terms are
acceptable to a degree: Two accept “Istro-Romanian”, one accepts “Romanian” and two do
not make the choice between the two terms explicit. A couple of respondents provide the
linguistic relationship between Zheyanski and Romanian as a justification for their accep-
tance. However, even when accepting “Istro-Romanian” or “Romanian”, no respondent in
Žejane expresses any enthusiasm for either term.

In conclusion, there seems to be a lack of a widely shared – and, importantly, liked –
ethnic group name among the Vlashki speakers. The group terms introduced and used by
outsiders, notably “Istro-Romanian”, are never volunteered but are accepted by the people
in the Šušnjevica area with caveats – in opposition to other, less preferable group terms
such as “Romanian” and “Vlach”. In contrast, Zheyanski speakers are unanimous in their
support for the use of the village-derived group name, corresponding to the language name,
which they all volunteered. This demonstrates the importance that the village affiliation has
to them. Perhaps because of this, they show less acceptance for the outsider group terms,
such as “Istro-Romanian”, and find the term “Vlach” completely irrelevant to them.

2.3. SOCIAL ATTITUDES. Our questionnaire also contained twenty statements expressing
opinions related to the language status (economic, social and symbolic), language vi-
tality and desirability of institutional support for the language and identity (Komondouros
& McEntee-Atalianis 2007). Respondents graded the statements related to these themes on
a five-point Likert scale.

Respondents from the Šušnjevica area and Žejane demonstrated very similar views on
a number of statements. With regard to the social status of the language, the majority agree
that Vlashki/Zheyanski has an inferior social status in Istria. In both locations, people are
evenly divided as to whether or not its social status has improved. A similar division exists among them as to whether or not there is a stigma associated with the language, that is, whether they feel comfortable speaking the language in public, but opinions in Žejane are slightly more positive than in the Šušnjevica area. With regard to the economic status of the language, virtually all respondents agree that Croatian is the language young people must learn to succeed, but Zheyans, unlike the Šušnjevica area people, mostly disagree with the statement that young people need to leave the village in order to succeed. In both locations, the large majority of the respondents would like to have more opportunities to both speak and improve the language and feel that the government is not doing enough to support it.

With regard to the subjective sense of the language’s vitality, people from Žejane demonstrated significantly less concern about this than did the people from the Šušnjevica area: The minority in Žejane, only three out of eight people, believe that their language might disappear in the future. Parallel to that, the large majority in Žejane believe that there will always be speakers of Zheyanski in their village and only one person believes that Zheyanski is only for the use of the old people. In the Šušnjevica area, everyone expressed concern about the possible disappearance of Vlashki in the future.

With regard to the symbolic status of the language and the role it plays in people’s identity, Vlashki and Zheyanski speakers largely share beliefs. They agree that it is important to preserve Vlashki/Zheyanski and pass it on to future generations and that their culture is strongly linked to the Vlashki/Zheyanski language. They expressed pride in their village affiliations and agreed that language plays a defining role in their identity. However, there is a difference between the respondents from the two locations with regard to bilingual practices, such as code-switching and the use of Croatian in communication among Vlashki/Zheyanski speakers. In the Šušnjevica area, all eight respondents expressed their opposition to code-switching and six of eight felt that the use of Croatian by speakers able to speak Vlashki weakened the group’s identity. In contrast to the greater unanimity in the Šušnjevica area, respondents in Žejane were divided with regard to language protection: Three out of eight did not object to mixing Croatian with Zheyanski, and four out of eight did not see the use of Croatian among Žejane speakers as weakening group identity.

In summary, both the Vlashki and Zheyanski speakers express strong ethnic pride and the conviction that their language plays an important role in their identity. They also like the idea of passing the language on to future generations and using the language more in their lives. The Vlashki speakers, however, seem to be more pessimistic with regard to the future of their language, that is, the possibility of its survival in the community.

3. Discussion and Conclusions. Our respondents’ judgments on language use confirm that language shift from Vlashki/Zheyanski to Croatian is under way in both the Šušnjevica area and Žejane. Speaker judgments regarding the domains and frequency of use of Vlashki/Zheyanski also indicate that language shift to Croatian is more advanced in the Šušnjevica area than in Žejane. The Vlashki speakers discussed in this paper – those speakers who are 51 years of age or older – have shifted to the predominant or exclusive use of Croatian in communication with their children. This generation is the pivot generation in the Vlashki-speaking area, the one that has brought about language shift. The same is not true of the Zheyanski speakers in this paper’s sample, most of whom still communicate primarily in Zheyanski with their children. However, the language use of their children, whose usage patterns we do not discuss directly here, seems to have changed in relation to
their own children, our respondents’ grandchildren, as reflected in our respondents’ report that they speak mostly Croatian to them.

The Vlashki and Zheyanski speakers also differ with respect to their awareness of the endangered nature of the local language in their communities and their attitudes toward their language, in particular, toward code-switching between Vlashki/Zheyanski and Croatian and the use of Croatian among Vlashki/Zheyanski speakers. These differences seem to correlate with the difference in degree of advancement of language shift in the two locations: The Vlashki speakers are reasonably pessimistic about the chances for their language’s survival into the future and, at the same time, express stronger attitudes of language protectionism in the face of this doomed prospect. As a group, Zheyanski speakers demonstrate less awareness of the language shift in progress and do not all oppose bilingual language practices.

It was not our goal in this paper to explain the reasons for language shift in these communities. We believe that a complex range of socio-economic and historical factors would need to be looked into and appealed to in order to do that (Himmelmann 2010). Speakers engage in everyday interactions and make language choice decisions under the pressure of a range of socio-economic factors, such as changing local economies, migration patterns, marriage customs, and others. In the communities we studied, the dramatic depopulation of villages since 1945, loss of their geographic isolation due to modernization, and the replacement of the traditional agricultural lifestyle with industrial and service employment are undoubtedly the main socio-economic factors of language loss. In conjunction with these, however, it seems probable that people’s conception of their identity has played a role in this process as well.

The difference in the advancement of language shift between the Šušnjevica area and Žejane might partially be the difference in the conception of identity between the two locations, which in turn can be related to the villages’ locations and the effect of depopulation on the communities. The Šušnjevica area is composed of several villages and adjoining hamlets, many of which were dramatically and quickly depopulated in the period following World War II. In communities where identity is defined in terms of family lineage and language and where other aspects of their culture are indistinguishable from those of their neighbors, depopulation has direct detrimental effects on such communities and their language use, especially when the villages and hamlets are dispersed, as they are in the Šušnjevica area. The Vlashki-speaking villages are surrounded by a number of nearby Croatian-speaking villages. Increasingly frequent contact and the routinization of intermarriage under circumstances of intense recurrent depopulation have heightened the difficulty of (re-)establishing and maintaining a viable cultural boundary with their neighbors. This is expressed and, in turn, possibly compounded by the fact that no ethnic name is both widely accepted and liked by everyone, in particular no ethnic name that is cognate with the name of the language.

Similar socio-economic changes affected Zheyanski speakers, but, as inhabitants of a single village in an isolated location, they seem to have had an easier time identifying their community in unique terms and maintaining a sense of sufficiently distinctive village belonging – and authenticity. Another circumstance might have contributed to a temporary advantage for the Zheyanski speakers over the Vlashki speakers in terms of language maintenance: Žejane lies close to the Slovenian border, outside the Istriя Administrative Region that was established in 1991. While the village is part of Istria in a geographic and historic sense, it is not part of its “core” area. Instead, it is located on the northern border of the Istrian borderland region. Studies exist on other communities on the northern Istrian...
Identity and language shift among Vlashki/Zheyanski speakers in Croatia

border that have embraced a “border identity”, marked by “persistence of ambivalent and blurred local group identities” (Nikočević 2005: 250), especially in terms of membership in any particular nation. While Zheyanski speakers are like the Vlashki speakers with respect to embracing multiple identities, they are different from them and more like their immediate neighbors in northern Istria in that they expressed reserve toward larger group membership, whether national or regional. This disposition, compounded by other factors discussed earlier, has reinforced for Zheyanski speakers the value of their distinctive village and language affiliation and postponed, but not dismissed, the need for language shift. As argued by Vržić elsewhere (2013), in this kind of social context, where identity is transient and pressure for national homogeneity weaker, there is likely less immediacy to assert strongly one’s special identity as well as less immediate pressure and need to assimilate.

A final point we would like to address here is ethnic pride and language loyalty. It may seem surprising that both Vlashki and Zheyanski speakers express strong ethnic pride and place great importance on the role of their language in their identity when, at the same time, language shift to Croatian is well under way in Žejane and largely completed in the Sušanjeva area. However, other studies of endangered language communities show that continued reference to an in-group language as one of the important symbols of a group’s identity is disassociated from everyday language practices, even though the latter is a domain where the shift away from the use of the in-group language may be quite apparent. Among numerous examples, we mention Scottish Gaelic-speaking fisherfolk (Dorian 1980), a Taiap-speaking village in Papua New Guinea (Kulick 1992), and Greek-speaking Istanbulites/Constantinopolitans (Komondourou & McEntee-Atalianis 2007, McEntee-Atalianis 2011). While speakers are aware that the local language and the language of wider use have very different instrumental values, the local language may continue to fulfill its symbolic function for speakers even after it is largely abandoned in everyday use due to socio-economic pressures. As long as there are bilingual speakers or even passive bilinguals, it may continue to provide them with a sense of local authenticity and anchor a key aspect of their multiple and layered identities.

REFERENCES


Zvjezdana Vrzić
zv2@nyu.edu

John Victor Singler
john.singler@nyu.edu
The sociolinguistic evaluation and recording of the dying
Kursenieku language

Dalia Kiseliūnaitė
Klaipėda University

Since the times of the Teutonic order until 1923, the Curonian Peninsula was a part of Prussia, and later – a part of Germany. Baltic tribes’ migration processes of different intensity occurred here. In the 16th century the newcomers from Latvian speaking Courland started to dominate, moving to the spit in several waves up to the 18th century; at the same time, people from the continental part (the majority of them were Germanized Prussians), colonizers from other German lands, and Lithuanians from the Klaipeda area settled in the region. The Kursenieku language, also known as New Curonian (German Nehrungskurisch) can be categorized as a mixture of Latvian Curonian dialects with Lithuanian, German, and elements of the now extinct Old Prussian. Since it had no written form, Kursenieku was roofed by Lithuanian and later by German, which had functioned as languages of religion and education for a long time. The community disintegrated at the end of World War II. After the Kursenieki community left their homeland and settled in different towns and villages of Germany, there was no practical use for the maintenance of Kursenieku. The chronological reconstruction of the Kursenieku is possible and useful for the Baltic studies; however, there is no motive for revitalization: nowadays, there is no community willing to use this language. This article briefly presents the development of the Kursenieku language in its ethnocultural context. Moreover, it raises the discussion around its status (variety or language), provides its sociolinguistic characteristics, describes the work that has been done with the language, and presents urgent goals and research perspectives.

1. BRIEF INSIGHT INTO THE HISTORY OF THE REGION AND ITS COMMUNITY.
Since the times of the Teutonic order until 1923, the Curonian Spit was a part of Prussia, and later – a part of Germany. Baltic tribes’ migration processes of different intensity occurred here. It provided the grounds for the formation of a multi-ethnic community, which lived primarily from fishing.

---

1 I would like to thank my colleague Dr. Christiane Schiller for sharing ideas to improve this paper
2 Term used by Schmid (1989).
3 Curonian Spit (German Kurische Nehrung, Lithuanian Kuršių nerija, Latvian Kuršu kāpas, Russian Куршская коса) – is a 98 km long, thin, curved sand-dune spit that separates the Curonian Lagoon from the Baltic Seacoast. Its southern part lies within Kaliningrad Oblast, Russia, and its northern part within southwestern Lithuania. It is a UNESCO World Heritage Site shared by the two countries.
Besides the people from Baltic origin, the region continued to be inhabited by the colonists from further Prussian and other distant localities. As well as in other places, the German colonists had since the Middle Ages been living on the Curonian Spit, especially on its southern part, which had better routes of communication with the continent. In the 16\textsuperscript{th} century, the newcomers from Latvian speaking Courland started to dominate, moving to the spit in several waves\textsuperscript{4} until the 18\textsuperscript{th} century; at the same time, people from the continental part (the majority of them were Germanized Prussians), colonizers from other German lands, and Lithuanians from the Klaipeda area settled in the region.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Maximum of the settlement of Kursenieku language speakers in Prussia (Einhorn 1649) (http://de.wikipedia.org/wiki/Kuren, 10 December, 2015)}
\end{figure}

Such migration trends on the Curonian Spit have been proved by numerous facts: archaeological and historical sources, as well as toponyms and personal names (Einhorn 1649,

\textsuperscript{4} The first newcomers are recorded already at the end of the 14\textsuperscript{th} century, but their number was not big (Forstreuter 1981: 286). Since the beginning of the 15\textsuperscript{th} century, a few waves of newcomers from Courland had reached the spit, and this process continued during the 16\textsuperscript{th} century (Bezzenberger 1888: 271).
Rimantienė 1999, Bezenberger 1887, Forstreuter 1931, Kiseliūnaitė & Simutytė 2005, Kiseliūnaitė 2005). Thus, when considering the evolution of ethnic culture, not only the principles related to the island are to be applied on the spit, but also the ones of the ethnic margins and multicultural spaces are to be considered. The community of the Curonian Spit was united by trade predetermined by specific natural settings. For the most part, those were fishermen villages with their peculiar lifestyle, which can be distinguished from the continental part by the following features:

- There was no cultivable land on the Curonian Spit, except for a small lowland in Rasite (Germ. Rositten/ Lith. Rasytė). Thus, the inhabitants of the spit did not come to know the life of the farmers, or gradually distanced from it (if we bear in mind that a part came from the continent or were Courland countrymen) (Seraphim 1892).

- People did not live in farmsteads, but rather in street-type settlements, where the homesteads were located on the very coast of the lagoon, but not on the seashore; the current fact is important for the development of the ethnic culture and language, because we dealing with a community that was densely inhabiting a small area.

- The spit was ravaged by plague, and the woods were being cut during the seven-year war, which caused an ecological disaster that dwarfed the development of the region. Until the middle of the 19th century, it had been a poor land, in which public servants felt like the outcasts (Sembritzki 1918: 114).

- Although the spit is a peninsula, people lived there in the same conditions as on an island. The villages were frequently whelmed by sand and snow, becoming difficult to reach and, thus, forming a restricted lifestyle in the middle part of the spit, where the novelties would arrive late; in the 19th century, the travelers and researchers described the villagers as highly conservative (Zweck 1898, Negelein 1902).

- Due to restricted movement and reticence of the villagers, there was always a danger of incest. To prevent it, mixed families with the shore inhabitants had been created (Lithuanian women brought there not only the language, but also the customs); nevertheless, in most villages the majority consisted of a few large families (Strakauskaitė 2001: 91).

- Though the movement was restricted on the land, the fishermen of the spit always maintained the connections with the countrymen of the mainland. They sold fish, bought agricultural goods and household stuff. This exchange was the cause of a big influence of Lithuanian in the Kursenieku language. A lot of Lithuanian words are present in agricultural lexis, since the Kursenieki had to acquire hay and agricultural goods from the other shore. Thus, an amount of trade based linguistic changes occurred, for instance in the composition of the numerals.
Figure 2: Places with Kursenieku speaking population (Bezzenberger 1888)
(The place names are presented in their official forms)
2. KURSENIEKU: LANGUAGE OR DIALECT? From the genealogical perspective, Kursenieku is the Courland variety of the Latvian language, attached to the middle dialect of the Curonian sub-dialect. This fact can be observed in following features:

1. Kursenieku is based on the western dialect of the Latvian language, which was formed in the 15th–16th century, containing elements of the old Curonian language substrate; however, Kursenieku is not considered an uninterrupted continuation of the Curonian language.5

1.1. In the old Curonian language, the nasal diphthongs an, en, in, un were retained from the proto-language (relicts in Lithuanian and Latvian onomastics: Alsunga, Palanga, Tenžė, Kretinga); however, they are replaced in the Kursenieku variety by the Latvian uo, ie, ī, ā (luogs “window”, piec “five”, tiks “net”, sūtina (infinitive) / sūtina (present) “send”).

1.2. The major part of the lexis is Latvian; as for the grammar, there are slight differences. Thus, the language is understood by almost every Latvian.

1.3. Several of its features, as supposed, could belong to the Curonian language substrate. The same features were found as well in the western Courland varieties not long ago, some of them were undertaken by the Latvian standard language.

1.3.1. Phonetics

- uv (zuve “fish”), standard Latvian iv (zīvs)
- an, en, and in in curonisms: krants “steep shore” (standard Latvian kranta), bangā “wave” (standard Latvian vilnis), lenķe “valley”, dzīnars “amber”
- wide ē in verb infinitives (gulēti “to lie”, sēdēti “to sit”), slender ē in standard Latvian (sēdēt, gulēt)

1.3.2. Lexis

- dzievuoti “live”, “work”, standard Latvian dzīvuot “live”, but strādāt “work”
- plekste “plaice”, standard Latvian bute
- dižs, –a “big”, standard Latvian liels, –a

1.3.3. Morphology

- masculine instrumental plural –is (ar veci tīkli “with old nets”), standard Latvian –iem (ar veciem tīkliem)
- prefix āz– (āzmirst “forget”, āzmaksāt “pay”), standard Latvian aiz–
- adjectival degrees: juo tāli “further”, standard Latvian tālāk
- ē– stem form of the primary verb’s preterit: vede “take, lead”, nese “bring”, standard Latvian veda, nesa

2. The Kursenieku language has inherited an early layer of Germanisms and Slavisms, formed in Latvian at the beginning of Christianity: bāznīca “church” < Russian божница, божица; cilēks, standard Latvian cilvēks < Russian

---

5 The problem arises due to the terminology, because users of the Kursenieku language call their language kurse valoda, their Lithuanians neighbors call it kurišų kalba; this term, in ethnographic rather than linguistic sense, is often used in the media, where the language of the inhabitants of Curonian Spit is often confused with the dead Baltic language (Eng. Curonian or Old Curonian, Lith. kurišų kalba), which is a way of making this linguistic reality more sensational, exotic, etc.
3. Language development and language contact. The Kursenieku language, being in other country, was isolated from the rest of Latvian-speaking community. Therefore, it did not undergo the development processes that were peculiar to the dialect of Latvian language in Courland. Due to the isolation, it remained archaic and retained many peculiarities of the 15–17\textsuperscript{th}-century Western Latvian dialects, which disappeared in the course of time. On the other hand, there was a need for innovations in everyday usage and they occurred in two ways: by using old word formation strategies or by borrowing new words and syntactic constructions from the contact languages – Lithuanian and German. The least changes in Kursenieku occurred in the fishing and family domain.

Lithuanian (the western dialect of East Prussia Lithuanians) was used in mixed families, especially if the mother was Lithuanian. Written Lithuanian of Prussian Lithuanians reached the Kursenieiki through schools and church. In the 17\textsuperscript{th} century, there was the

\begin{itemize}
  \item After the separation from its parent language (Curonian Latvian), Kursenieku did not take further part in the process of Latvian language development, as a regional variant, i.e. it did not participate in the development of standard Latvian and in the formation of Latvian national identity. The Kursenieki of the Curonian Spit did not consider themselves part of the Latvian nation. This means that Kursenieku can only conditionally receive the status of Latvian dialect (only in a genealogical perspective).
  \item It became a common language as other non-German languages in Prussia.
  \item Its further development was affected by socially stronger languages – Lithuanian and German.\footnote{Specific forms of numerals are an example thereof. Numbers from 1 to 9 are identical with the numbers in Latvian literary language, except for the Courland peculiar dui (standard Latvian divi, divas “two”); however, numbers from 10 to 19 are clear Lithuanianisms (desint “ten”: Lithuanian dešimt, different from Latvian desmit; vi(e)nālik: Lithuanian vienuolika “eleven”, different from Latvian vienpadsmit, etc.).}
  \item It served an expression of the group’s ethnic identity and contributed to its preservation.
\end{itemize}

Presently, with only a few speakers remaining, it can be included in the group of dying languages.

3. Kursenieku has inherited a Finnish layer, peculiar to Courland: launags “south, south wind, dinner”, karša (Latvian variety karaša) “bun”, valgums “wharf”, etc.

Some linguists consider Kursenieku a dialect of Latvian. However, from a sociolinguistic point of view, it is to be considered an independent Baltic language (in this sense, the author follows the opinion of Vanags 1999). The following arguments support the later categorization:

\begin{itemize}
  \item It became a common language as other non-German languages in Prussia.
  \item Its further development was affected by socially stronger languages – Lithuanian and German.
  \item It served an expression of the group’s ethnic identity and contributed to its preservation.
\end{itemize}

Presently, with only a few speakers remaining, it can be included in the group of dying languages.

\footnote{Specific forms of numerals are an example thereof. Numbers from 1 to 9 are identical with the numbers in Latvian literary language, except for the Courland peculiar dui (standard Latvian divi, divas “two”); however, numbers from 10 to 19 are clear Lithuanianisms (desint “ten”: Lithuanian dešimt, different from Latvian desmit; vi(e)nālik: Lithuanian vienuolika “eleven”, different from Latvian vienpadsmit, etc.).}
aim to evangelize Prussian inhabitants. Thus, priests were trained to serve in non-German parishes, and religious literature was being published. Kursenieku did not obtain written language status, which probably can be explained through the small number of speakers and the possibly high costs for publishing and priests’ support. The church language was Lithuanian, as it was the most popular language among non-German languages in East Prussia. Due to adverse natural conditions, churches and parish schools were located far away from each other – for example, Nida had no church and no school until the 17th century.

At the same time, services and education were conducted in German as well, and the inhabitants were gradually becoming trilingual. Lithuanian had a strong influence on Kursenieku. Being very religious, the Kursenieki adopted all the Lithuanian vocabulary related to religion and moral.

In the northern part of the spit, trilingual inhabitants were not rare. Probably, the use of Lithuanian retreated.

- The life on the spit began to change radically in the times of Bismarck, when German became the only official language in the Prussian part of Germany. This area was Germanized rapidly, because the children attended German schools; the mentality of Kaiser Monarchy was becoming stronger; after World War I only older people were attending Lithuanian services.

- Between the end of the 19th century and the beginning of the 20th century, besides fishing business, resort settlements and infrastructure were developed (roads, piers, hotels; sand slides were prevented by afforesting of the dunes). Modern vacationers from big cities found the spit fishermen a kind of exotic tribe that spoke an unknown language and followed old traditions and superstitions.

- After Klaipėda area joined Lithuania in 1923, the spit was divided into two parts according to the district boundaries at that time; thus, the part north of Nida was under Lithuanian administration and in an intense Lithuanization process, while the southern territory remained German. The most fishery places of the lagoon were left to Germany, which caused dissatisfaction among Lithuanian fishermen. In general, there were not many people who agreed with Lithuanian administration, in contrast to the mainland, and at the same time, the Curonian legacy began to fade, while Germanization reaped its benefits (for more details, see Endzelins 1931).

The community disintegrated at the end of World War II, when the residents of the Curonian Spit, as German citizens, were evacuated or scattered. Several repatriate families emigrated between 1958-1960. After the Kursenieki left their homeland and settled in different towns and villages of Germany, the maintenance of the Kursenieku language became threatened.

4. Current situation of the language. At the end of World War II, when the Soviet army was approaching, almost all Curonian Spit inhabitants were evacuated as German citizens. After a few years, several families came back to the homeland (mostly women

---

7 According to the facsimile of the lessons’ schedule in Nida’s school chronicle (Chronik der Schule zu Nidden edited by Nausėda & Gerulaitienė, page 129), religion lessons were taught in Lithuanian, and reading in German; in other part of the chronicle (Nausėda & Gerulaitienė 2013: 77), it is mentioned that in 1894, only 20 children from 112 had been speaking German at home, others spoke Kursenieku.
and children); however, they did not want to live in the Soviet-occupied territory and, on the basis of the agreement between Germany and the Soviet Union, the families moved west in 1958–1960. Only a few autochthons remained on the Curonian Spit, namely those who married Lithuanians or Russians and those who were not allowed to leave by the Soviet regime. New inhabitants came to the spit, and the autochthons remained an ethnic minority and were called Germans by the new inhabitants. Thus, there were no conditions for the Kursenieku language to be maintained. It was episodically used only in family circles on the initiative of repatriated elders. Today, there are barely a few speakers of Kursenieku left. According to their language knowledge, the speakers can be divided into three groups: 1. Passive speakers, i.e. speakers who, when being talked to or when having a conversation partner, fragmentally insert a few Kursenieku phrases in their discourse, but do not dare to speak more; 2. Speakers capable of keeping a conversation in Kursenieku (there are only few speakers remaining that belong to this group, a couple in Germany and one person in Sweden); 3. Active speakers, i.e., those speakers who are constantly communicating among themselves in Kursenieku and use the language almost every day (only two brothers, Sakuth, who live in Sweden, belong to this group).

4.1. Research on the Kursenieku Language. The research on Kursenieku has started at the end of the 18th century. The first written source was a short dictionary of 278 words in Peter Simon Pallas’ dictionary *Linguarum totius orbis vocabularia comparative*. Since the 19th century several linguists have been working on the language (A. Bezzenberger, J. Plakis, W.P. Schmid, F. Hinze, Chr. El Mogharbel). Except for a small study by J. Plakis, later researches on the Kursenieku language are fragmental, covering only separate aspects; moreover, they are based on a small amount of data. Diachronic changes in the dialect, from the beginning of the 20th century to the present, have hardly been explored.

At the moment, Dalia Kiseliūnaitė conducts most of the research: Kiseliūnaitė collected several hours of audio recordings, on the basis of which research is being carried out and publications are being prepared.

4.2. What Have Been Accomplished? Documents written in the language were discovered. Although the language did not become an official written language, several speakers have attempted to write down their native language, using Lithuanian, Latvian, and German graphemes.

At the moment, there are more than 70 hours of audio recordings but not all them are transcribed yet. However, some of the recordings are in private collections and are still not accessible for linguistic research. Several video recordings were also made during fieldwork stays. Other linguistically valuable documents of the region are in various archives with limited access; especially important are the lists of the inhabitants and place names in historical documents (church books, lists of tax-payers, etc.). Cartographic and iconographic data are being gathered and constantly complemented. In order to complement our corpus and to it interdisciplinary, we are also collecting data about Kursenieku from other

---

8 A small fragment of Kursenieku is registered in the Lithuanian grammar *Anfangsgründe einer Litauischen Grammatik* by Paul F. Ruhig (1747).
9 Richard Pietsch, who comes from Nida, and Paul Kwauka, a teacher from Klaipėda, have published a Kursenieku-German dictionary (*Kurisches Wörterbuch*); later, Pietsch, with the support of the linguist F. Scholz, published a German–Kursenieku dictionary (*Deutsch-kurisches Wörterbuch*, as well as valuable ethnographic data in a book called *Fischerleben auf der Kurischen Nehrung* (1982). We could also found several fragments of small manuscripts by other authors.
disciplines, such as history, regional studies, and ethnography. Considering the need and urgency of documentation (soon there will be no informants left), more fieldwork stays are being prepared for the near future.

4.3. LANGUAGE DOCUMENTATION AND FURTHER RESEARCH. Why is the research on Kursenieku important from an interdisciplinary point of view?

- As a dying language, Kursenieku deserves as much attention as other endangered languages in the world.
- Due to their archaic linguistic structure, Baltic languages are often object of Indo-European researches; however, Kursenieku is normally left out because its data are still almost unknown to the linguistic community.
- There is a substrate of an archaic vanished Curonian language in Kursenieku, which has not been investigated yet. It is reasonable to suppose that Kursenieku, due to its early isolation from Latvian, has retained more relics from the extinct Old Curonian language than contemporary Latvian and Lithuanian dialects.
- Due to its isolation, Kursenieku, originally a linguistic variant of Latvian, retained archaic features of this language and can serve, therefore, as a powerful source of the whole Latvian language history.
- Due to its development and status, Kursenieku can be a valuable source for researches on language and cultural contact between different ethnic groups.
- Onomastic research of the Curonian Spit is of great interest for historians interested in migratory movements, because it shows evidences of many migration processes that are still not clarified.
- In general, Kursenieku documents have also a prominent informational value for researchers in the domains of history, ethnography, and natural sciences.
- Being almost only a fishermen language, Kursenieku represents an interesting research subject from the sociolinguistic and ethnolinguistic perspectives.

4.4. PROBLEMS. The main problem we face in the research on Kursenieku is the lack of more qualified researchers and financial resources. In order to work on Kursenieku, substantial skills in Baltistics, good knowledge of Lithuanian, Latvian, and German, as well as knowledge of Prussian history are required. On the other hand, decreasing attention to the humanities puts Baltic studies aside. As a consequence, only a small number of students enroll Baltic studies’ programs. Thus, the amount of qualified researchers in the area is very limited. At present, the research on Kursenieku is basically in the hands of one person, and what is worst, the collected data also. Due to financial restrictions, there are no possibilities to involve and motivate other colleagues for the work on this language. In order to change the current situation, much more substantial financial resources are needed and more international projects have to be prepared and funded, especially by the interested states. Unfortunately, neither Lithuania, Latvia nor Germany have expressed such interest at the state’s level so far. The work remains in the hands of separate enthusiasts. In Russia, on the other hand, there is no interest in this subject at all, not only at a state’s level, but even in the academic sector.
5. CONCLUSIONS. Kursenieku is considered a separate Baltic language, due to its social status and functions. Its documentation and research is important not only for historic Baltistics and for the study of the region’s history and culture. Their results are urgent for researchers in various fields and even for the economic development of the region, in that research results can be used to support and develop the tourism industry. The territory in which Kursenieku was once actively spoken is protected by UNESCO as a cultural landscape object. Language heritage and ethnic culture are two of the most important priorities of the national park. The research on Kursenieku is urgent for the scientists of at least three states: Latvia, Lithuania, and Germany. Unfortunately, there is no substantial interest on the subject on the part of Russian academia, although half of the territory of the Curonian Spit gets into the Kaliningrad district of the Russian Federation.

Currently, research on Kursenieku language and culture is mostly concentrated at the Klaipėda University. However, it represents only the beginning of a big enterprise. An important task that must be accomplished is the creation of a database, accessible for the scientific community. While evaluating both old and new sources of language data without separating them from historical-social context, it is possible to make a scientifically objective evaluation of the peculiarities of this Baltic language and to develop a consistent base for its reconstruction.

The chronological reconstruction of the Kursenieku language is possible and useful for Baltic studies; however, there is no reason for revitalization efforts, since nowadays there is no community willing to use this language.

REFERENCES


Dalia Kiseliūnaitė
dalia.kiseliunaite@gmail.com
Language Revitalization: The case of Judeo-Spanish varieties in Macedonia
Esther Zarghooni-Hoffmann

University of Bamberg

Judeo-Spanish is a secondary dialect of the Spanish language having evolved from the ancient standard Spanish in the course of its expansion southwards. Although the language enjoys a heritage and presence in the Balkans of over five centuries, it is now facing language death – its acuteness depending on the region. In Macedonia, the two varieties of Bitola and Skopje last documented by Kolonomos (1962) need to be labelled “moribund” or “nearly extinct”. This paper aims to point out some of the aspects relevant to the author’s doctoral research study, in which a documentation of the current language status of Judeo-Spanish in Macedonia is envisaged. The deliberations look at the reasons for language endangerment and at the same time evaluate possibilities and opportunities for language revitalization – what priorities are to be set, what role do linguists and especially the community play, what is the approach, what are skills, methods, and steps to be taken into consideration to ensure not only a documentation of the language, but also and foremost its conservation and revitalization.

1. INTRODUCTORY NOTES. The year of 1492 marks a milestone in Iberian and world history not only against the background of Christopher Columbus’ discoveries and the publication of Nebrija’s Gramática de la Lengua Castellana, but also it brought about the re-establishment of political and religious unity in Spain. Along with it came the edict of expulsion enunciated by the Reyes Católicos on 31st March 1492, towards the Jewish community, whereby, if they did not convert to Christianity, they were expelled from the country. This contributed decisively to the emergence and development of a variety that had conserved traits of a language which, before the expulsion, had differed little from Spanish despite its distinctive Jewish features; traits that now generated the core of a new language: Judeo-Spanish (Harris 1994: 65).

After their expulsion from Spain, Jews settled in the neighboring country of Portugal, where the same edict reached them in 1497 as a consequence of the intermarriage of the royal families. This again prompted them to leave the country to Amsterdam in particular, but also France, Italy, and later Great Britain (Born 2012: 126). Jews from South Spain

1 The denomination “Macedonia” does not imply the notion of the geographical territory of Macedonia divided between Bulgaria, Greece and Macedonia, it henceforth refers to the current territory of the Republic of Macedonia, admitted to the United Nations under the provisional reference of the Former Yugoslavian Republic of Macedonia (FYROM).

2 This edict was revoked in Spain no earlier than 1968 (Bollée 2003: 118).
mainly emigrated to North Africa, where the Judeo-Spanish variety *Hakitia* emerged. The majority followed the invitation issued by Sultan Bayazid II to settle in the regions of the Ottoman Empire, where they were granted religious freedom. A considerable proportion emigrated there from Castile and Northern Spain, a phenomenon that can be observed in the denominations of the respective national synagogues, as well as in the lexis and phonology expressed with Catalan, Aragonese (both spoken in Northern Spain), and Portuguese components (Wagner 1990: 38f.).

The most significant Sephardic communities emerged in the Balkans, the main centers being Sarajevo, Saloniki, Monastir (today’s Bitola), Constantinople and Smyrna (Born 2012: 126, Wagner 1990: 38.). In some of those communities, the local Balkan Jews adapted to their new neighbors, forsaking their own language (Greek (Born 2012: 126) or German (Wagner 1990: 38)) and adopting the Judeo-Spanish vernacular.

Imported into its new homelands, this language however underwent an assimilation directed towards the respective regional variety (Greek, Turkish, etc.), mostly manifested in lexical terms. Phonetic-phonological changes experienced by the Spanish language on the Iberian Peninsula during the 16th and 17th century did not exercise any influence on Judeo-Spanish (Born 2012: 126). Communities based in Western Europe stood in constant contact with the Spanish mother country, and as a result these communities followed the development of the Castilian Spanish, whereas the Balkan communities conserved much more ancient stages of the language (Bunis 1991: 7f.).

Although the Judeo-Spanish varieties enjoy a heritage of five centuries, they are now facing language death – its acuteness depends on the region and reasons which will be elaborated below. A look into the research to date suggests that the Judeo-Spanish varieties in the Balkans are the least prevalent in linguistic research. The largest present-day survey for languages of the world – *Ethnologue* – locates the variety under the denomination “Ladino” as prevalent in Israel, but also existent in Greece, Turkey, Puerto Rico, and the United States. A Balkan variety is referred to as a dialect. When it comes to Macedonia, out of the formerly smaller communities such as Kumanovo, Veles, Strumica, Gevgelija, and the bigger, thriving Jewish communities such as Bitola, Skopje, Štip (Kolonomos 1962: 22), only Bitola and Skopje are still noteworthy (Ibd.: 22) and subject of the author’s doctoral research study.

Given the fact that the last thorough research was undertaken by Kolonomos in 1962 and the circumstances whereby the field research in the Judeo-Spanish varieties in Macedonia is still rather in its fledgling stages (field access, documentation, ageing of community members), the following survey will be based on scholarly findings rather than field results at this given point. Nevertheless, it aims to provide insights into reasons for the endangerment of Judeo-Spanish as well as tendencies, perspectives, and practical steps of its documentation and conservation based on the needs expressed by the local community.

2. LANGUAGE DEATH. “There is agreement among linguists who have considered the situation that over half of the world’s languages are moribund, i.e. not effectively being passed on to the next generation.” (Foundation for Endangered Languages 1996: 3). This is certainly the case when it comes to Judeo-Spanish in Macedonia.

Taking *Ethnologue* into consideration and basing field observation upon the suggestions of the Expanded Graded Intergenerational Disruption Scale (EGIDS) (Lewis & Simons

---

Language Revitalization: The case of Judeo-Spanish varieties in Macedonia

2010), which provides concise labelling and description of the 13 levels of language status, Judeo-Spanish varieties in Macedonia find themselves labelled “moribund” (“The only remaining active users of the language are members of the grandparent generation and older.”) or “nearly extinct” (“The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language.”). Considering the fact that Ethnologue allocates to the language, globally known as “Ladino”, an educational language status, the language being “in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education,”9 there is striking divergence between the global notion of the minority language and its status in Macedonia, given that there are not more than a handful of native speakers in the country. A widespread system of institutionally supported education is not existent and due to the language status not envisaged at this point. The Directorate for Promotion and Development of Languages in the Education for the Ethnic Minorities lists the languages of the Turkish, Bosniaks, Serbians, Vlachs, and Roma as the languages to be developed, Judeo-Spanish or “Ladino” is not included in this list. It is hoped that the advice of the OSCE to the government of Macedonia, whereby the preservation and promotion of a multi-cultural society is encouraged not only in words but in deeds,10 will set a precedent for institutionally supported education in the case of Judeo-Spanish. However, such a development seems fairly remote at this point.

Both the overall development of Judeo-Spanish in Macedonia along with the endangerment of the language and its official recognition must lead to the categorization elaborated above.

How, then, does such a categorization come about? What is language death? A language is declared dead when no one speaks it anymore, and it is “effectively dead when there is only one speaker left, with no member of the younger generation interested in learning it.” (Crystal 2000: 11). However, if there are two or more speakers, it is the context in which the population is included that must be evaluated. For example, isolation and density of a population play a vital role in the discussion about language endangerment (Ibd.: 12ff.). Crystal (2000) states that any discussion of language death must be seen in the following perspective:

---

8 The scientific denomination is “Judeo-Spanish” – despite some controversy: A denomination “Judeo-German” does not exist anymore, the term “Yiddish” has been established for a long time. Moreover, the term “Judeo-Spanish” does not take into consideration the importance of the Portuguese language, since with the reference to language fusion and affiliation both Spanish and Portuguese play a significant role (Faingold 1989: 26). “Djudezmo” concerns the traditional, everyday vernacular (Bollée 2003: 117f.) within the Jewish communities settled in the region of the ancient Ottoman empire (Bunis 1991: 7ff.). From a philological point of view, the term is correct, yet it has not managed to establish itself since no adjective can be derived from it. Speaking of “Sephardic”, originally religious ordinances are referred to and it does not imply that all Sephardim are Hispanophone (Faingold 1989: 26). “Ladino” designates the literal literary language of religious and liturgical translations, thus meeting solely sacral needs (Bollée 2003: 117f., Born 2012: 126). For some linguists, “Ladino” is methodologically opposed to the written sub-variety of “Judeo-Spanish” inasmuch as it stands in strong lexical as well as mainly morphological and syntactic accordance to Hebrew (Bollée 2003: 177f.). Speakers themselves and numerous researchers use the terms analogically as a synonym for “Judeo-Spanish” (Born 2012: 126). Besides that, the term is very common in Anglophone scholarly literature. It is also the term with which many Jewish communities are familiar.
[I]t is evident that a very small number of languages account for a vast proportion of the world’s population (…). The 8 languages over 100 million (Mandarin, Spanish, English, Bengali, Hindi, Portuguese, Russian, Japanese) have nearly 2.4 billion speakers between them (…). If we continued the analysis downwards, we would eventually find that just 4

Every part of the world has its own underlying, varying conditions, a population number alone can thus give but generalized, insufficient insight and information about the degree of language death.

2.1. Reasons for Language Endangerment and Death in the Case of Judeo-Spanish Varieties in Macedonia. The range of factors leading to the dying of a language is vast and mostly there is a combination of reasons. A global overview over language death cannot be given (Crystal 2000: 70), but in case studies these factors can be easily identified. The following provides a summary list of causes of why languages disappear:

- Physical endangerment of a community of speakers: natural causes such as earthquakes, floods, etc.; famine, drought; diseases; exploitation of resources; conflict and war (Crystal 2000: 70–76).
- Endangerment of culture: assimilation; domination of one culture over the other; loss and lack of prestige of a language (Crystal 2000: 76–88).

There are many more factors that are conducive to language endangerment, as in the case of the Judeo-Spanish varieties of Macedonia. In the course of the 19th century, numerous lesser Sephardic communities in the Balkans found themselves under the influences of secularization, assimilation and emigration (Born 2012: 126). With the beginning of the 20th century, modernization found its way into the Balkan region and eliminated extensively Jewish life-forms, which is reflected in the replacement of the language of the Sephardim with the respective national language (Born 2012: 126). Besides these impacts of an incipient nationalism in the Balkan states, Harris (1994: 197f.) describes further factors that proved conducive to the decline of Judeo-Spanish:

- Decrease of prestige forced by the national governments and at last carried by the Sephardim themselves through negligence or renunciation;
- Israelization;
- Americanization;
- Jewish life now following an Ashkenazi standard;
- Establishment of the Alliance Israélite Universelle, in the course of which the French influence became indispensable;
- Lack of language cultivation and standardization on the part of language academies or respective central institutions;
- Indifference on the part of the Spanish mother country and eventually the final breach with Spain;
• Application of the Hebrew alphabet and the Rashi script in lieu of the Latin alphabet;
• Absence of prestigious Judeo-Spanish literature as well as ultimately the reduction of ranges of speech from a formerly powerful language of economy towards fewer linguistic domains, focused on rather private life.

The reduction to the communicative spaces to private life clearly escalated alongside the growing threat deriving from anti-Jewish policies introduced by German occupiers and collaborators in situ. Speaking the language was avoided so as not to be identified as member of the Jewish community. The decline of Jewish way of life, language, and culture culminated drastically in the persecutions and the deportation of Jews to Treblinka.

In March 1943, German SS and Bulgarian authorities deported 7,144 Jews to Treblinka from Bulgarian-occupied Macedonia and 4,000 Jews from occupied Thrace. (…) Some 900,000 Jews and approximately 2,000 Roma (…) were murdered at Treblinka. There were fewer than 100 known survivors. (Mais 2011: 126)

98% of the Jewish population in Macedonia was thus almost entirely annihilated (Quintana Rodriguez 1997: 47f.). In consequence of the genocide, no nameable Judeo-Spanish influence can be documented in ancient thriving centers such as Bitola. Small groups of speakers are observed in Istanbul, Israel, Latin America, and the USA (Bollée 2003: 117f., Born 2012: 126), which are mostly from the older generations. Among younger generations the language falls prey to assimilation in the particular regional language. No monolingual Judeo-Spanish speakers are listed, and it is in a minor share of communities where it performs the function of a primary language. In 1979 already, Malinowski reported upon his field research in Israel: “I was unable […] in spite of repeated inquiries to find any informant under twenty years of age.”( Berschin 2005: 21, after Malinowski 1979: 165). Today, after over 30 years, we must make the assumption that a Judeo-Spanish speaker is likely to be at least 70+. Judeo-Spanish, thus, looks back on its vast heritage, and is at the same time confronted with its extinction. The UNESCO Atlas of the World’s Languages in Danger locates Judeo-Spanish under the denomination “Judezmo” among the languages severely endangered in their vitality. The total population amounts to 112,130 speakers worldwide.

3. ASPECTS AND STAGES OF LANGUAGE REVITALIZATION IN THE CASE OF JUDEO-SPANISH IN MACEDONIA.

To address the on-going linguistic loss in times of standardization and unification, measures need to be taken to prevent a fatal decrease in the vitality

---

11 There has been discussion on the validity of a Bulgarian involvement in the deportations. It is true that this question does not have a simple answer. Almost 50,000 Jews from Ancient Bulgaria escaped the deportation to the concentration camps and survived World War II – thanks to numerous, courageous interventions on the part of their fellow countrymen. The synod of the Bulgarian orthodox church, several deputies of the parliament as well as the former deputy President of Parliament Peschew protested against the extradition. Nevertheless, 11,343 Jews from the territories that were assigned to Bulgaria by Hitler (Thrace in Greece and Yugoslavian Macedonia) were not able to escape and were eventually deported – an incident which was confirmed by the German police attaché in Sofia at that time and is nowadays assessed correct by the Jewish Community of Bulgaria. http://www.faz.net/aktuell/politik/ausland/holocaust-romantische-luege-oder-heldentat-1228323-p2.html (10 December, 2015).
13 http://www.ethnologue.com/language/lad (10 December, 2015). This number is from 2013. Bollée (2003) reports about 130,000 speakers and upon 160,000 speakers ten years before (Bollée 2003: 117f.). The decline is thus remarkable.
of communities and further loss of cultural knowledge. It is indeed to be expected that Judeo-Spanish will soon not be spoken any more in Macedonia. Yet there is interest in linguists’ involvement within the local community – as well as a need for such involvement being expressed at a global level. Crystal (2000) asks the question “Why should we care?” if there is a far-ranging belief that sharing one single language would foster peace, unity, and global understanding. He answers this question with the need for diversity and the expression of identity, pointing out that languages are repositories of history and contribute to the sum of human knowledge, quite apart from the fact that they are interesting in themselves (Crystal 2000: 32-65).

How can a revitalization of the Judeo-Spanish language in Macedonia come about? What are the steps and stages? What is the role of the linguist and most of all the community itself?

Given the worldwide complexity of problems, the case of Judeo-Spanish can rely on numerous case studies carried out so far. These have shown that the loss or conservation of a language depends greatly on the involvement of indigenous communities, local support groups, and outside bodies. Linguists have shown increasing interest in language revitalization particularly since the 1990s, and yet it is not only about “doing something”, but utmost importance and deliberation must be dedicated to what is to be done – or initially, what needs to be taken into consideration.

3.1. PRIORITIZATION. Prioritization is indispensable in the face of constraints and restraints of time, personnel, and resources. According to Ostler (1997, after Crystal 2000), “The first step in language rescue must be an informative assessment of a language’s current situation.” (Crystal 2000: 92) The predictability of language vitality depends solely on elaborate data (Crystal 2000: 92). In the vastness of information it needs to be clear what kind of information is envisaged: Besides the number of speakers, the context of speaker fluency, accuracy, and age levels needs to be taken into consideration to assure adequate evaluation of linguistic vitality, the possibility of revitalization, and its consolidation towards continuity (Crystal 2000: 92).

Consequently, the question needs to be raised as to which information is valid information and which is irrelevant information. A unified framework cannot exist given the extensive parameters exerting their endangering influence on a given language. The framework alongside which language revival can take place will greatly depend on the attitude of a community and its aspiring intentions towards its language (Crystal 2000: 92). Despite the acute need and exigency of a language-conserving framework, an attitude of haphazard collection of empirical data must be avoided, as well as postponing the elaboration of a framework until all the facts are gathered – the former proving short-term and inefficient, the latter inhibiting (Crystal 2000: 93–94). This typological framework then leads to the classification of a language in the light of demography, sociology, psychology, history, politics, geography, education, religion, economy, and technology – broadened by external factors such as local, regional, national, and extra-national influence, all of which are based on case studies (Crystal 2000: 94), all leading to a foundation based on facts and a theoretical frame of reference.

3.2. CONSCIOUSNESS. Another aspect to be taken into consideration is the lack of consciousness of languages in general and endangered languages in particular, in consequence of their abstractness and complexity on different levels, which is shown in the following account.
I am Rapa Nui, and, of course, I cannot speak for all Rapa Nui. But I can speak about what I see happening on my island (...). First, the people of Rapa Nui are not free to do what we want: we are controlled by another government, the government of Chile. This government has not always been kind to us or allowed us our self-determination. Because of this, our language is dying. I want to do something about it, but there is resistance from many different quarters. The Chilean government and sometimes our own elected officials do not really support our efforts to revitalize our language. Also, linguists and anthropologists have come to study us. While we appreciate their efforts, we want to control our own diversity. In addition, some efforts have created more division in our community. (...) Language revitalization should start in the home and be a community effort. It should be everybody’s concern. The people of Rapa Nui need to come together and decide for themselves what the future should be. (Hotus Tuki 2009: 199f.)

Hotus Tuki (2009) thus refers to several variables essential to the process of language revitalization: Conducive political conditions (when it comes to the preservation of ethnic and cultural rights, willingness of financial support, among others), experts with disposition towards language documentation and dissemination, and the community itself. Consciousness is thus indispensable on every level. What that implies in terms of a community and its contribution to language conservation as well as linguists’ assistance is to be looked into more thoroughly below.

3.3. INTERDEPENDENCE OF LINGUIST AND COMMUNITY. It seems that linguistic research has been carried out without consideration for community’s needs for a long time. Tsunoda (2006) confirms a change in culture, whereby communities now benefit more directly from research and see their personal and intellectual rights being respected. Despite the academic exigency to document endangered languages (and the awareness of such in the first place), meeting the needs of a community must be ensured and go hand in hand (Tsunoda 2006: 228).

Field work thus is no longer simply a professional task a researcher is responsible for; it rather implies a role that follows given ethics. Putting academic success over the integrity of a local community is one of the traits of exploitation. In the synopsis given by Tsunoda (2006), basing the insights on documentation aims of different scholars, the following aims are prevalent: (i) to satisfy academic curiosity, (ii) to foster academic success, (iii) to secure linguistic data, (iv) to ensure cultural heritage of humankind, and (v) to contribute to cultural heritage of the communities for their benefit, e.g. with the aim of language revitalization. So far, the linguistic concerns seem to rest largely on deliberations which are remote from the needs of the community. An attitude of having the right rather than the privilege to carry out field research in a respective community is not conducive and rather limiting to any variable that plays a role in language documentation and conservation (Tsunoda 2006: 218).

Nevertheless, communities have started to express their displeasure towards merely external influence and have begun to formulate a catalogue of imperatives:

(a) The community should decide who is allowed to conduct research there (...). (...) A request for permission needs to provide information on (i) the language to be investigated, (ii) the aim and significance of the planned research, (iii) the topics of the research, (iv) method of the research, (v) method
of publication of the research results, and (vi) whether the research will be returned to the community, and how? That is, “academic theft” is no longer allowed. (b) The community, and not the researchers, should choose research projects. (…) The community may assign a research topic which is different from the one initially envisaged by the researcher. (c) Researchers should inform the community of what is happening (…). (d) Researchers should collaborate with community members (…), and should not marginalize community members in research. However, some community members “want to do the work themselves without help from foreign experts” (…) (f) It is not sufficient for researchers to publish for their own benefits (…). They should not monopolize knowledge (…). They should benefit the community in question, by giving back the research results to the latter (…). Thus, (i) research results should help the education of the children of the community (…). (ii) Linguists should train community members in linguistics (…) – so that the latter can conduct research. (iii) Linguists should aid language revitalization activities (…). (iv) Research results should be fed into the general education system (…). (Tsunoda 2006: 219-220)

With reference to the community in Macedonia, certainly the allowance to conduct research on Judeo-Spanish has been given. So far, this permission focuses on Skopje in default of access to the community in Bitola. The language to be investigated is Judeo-Spanish. The dissertation research upon which this paper is based does not aim at a language-geographic description of the Judeo-Spanish in the Balkan countries, in particular consideration of Macedonia, since this was undertaken by Quintana Rodriguez (1997, 2006), neither does it envisage further research upon the phonological characteristics of the Judeo-Spanish in a continuum. In fact, a description of the current status of Judeo-Spanish in Macedonia is to be attempted, the more so as Kolonomos thoroughly and exclusively described the two varieties of Bitola and Skopje in her dissertation in 1962. The selective description of the language enclaves of Bitola and Skopje does not derive as much from an attitude of dependence on previous research, but is based on the reality whereby the formerly dynamic center of Bitola has the most thriving cultural heritage and the community of Skopje the largest and most active Jewish Community in Macedonia. The central questions of this investigation thus relate to the current language status of the Judeo-Spanish and its varieties in Macedonia. The most significant group of speakers is nowadays to be situated in Skopje, out of which some are to be attributed genealogically to the community of Bitola. Is it thus possible that assimilation has taken place in Macedonia whereby both varieties have assimilated towards a “Skopje variety”?

If phonetics and lexis have already been thoroughly studied and documented, what insights can be formulated about syntax? What influence does language contact exert on the syntactic structure of the respective languages when a Romance language develops in a non-Romance environment? What is structurally or typologically of Romance origin, where do Balkanisms can be observed, where is language innovation to be located? It is in this context that the interrelation of Romance and Balkan linguistics will reveal its scope, since so far Judeo-Spanish has been surveyed mostly under the lens of Romance Studies (Kramer (2007).

Hand in hand with these linguistic questions, light needs to be shed on the benefits for the local community to such a degree as to frame tendencies for its language, as to whether there is a possibility of language conservation and revitalization in a community
as small as the one in Macedonia and how respective measures could look like. In personal contact with the Jewish Community of Sofia and Skopje as well as in conversations with the Director of Centropa, Dr. Serotta, a consistently homogeneous opinion was confirmed, namely that there is the desire to actively maintain the language and to make it subject of scientific research after a long period of denial, which reflects the need to hide the Jewish identity and protect the members of the community in times of growing nationalism.

A description of the language should therefore be realized in close connection with the local community, taking its needs into account. In consultation with the current president of the Jewish Community of Skopje, Berta Romano-Nikolic, it was found that it would be ideal to offer the study of Judeo-Spanish as part of the community curriculum. Such a study should not be overly based on the aspect of language itself, rather on conveying Sephardic culture to foster an affirmative attitude within the community towards the reassuming of an identity that has been lost and towards the understanding that language is part of culture (Crystal 2000: 119). The preparation of language material should proceed in close collaboration between the linguist and the president as well as the secretary of the Jewish Community, who is well versed in linguistics and thus able to perform the connection between external and internal variables in relation to the community.

There are certainly innumerable other factors to be observed, especially when it comes to public awareness, not to speak of the political scale. However, one major step has already been taken in terms of cinematography, when in 2012 the film “The Third Half” was introduced to a worldwide public. It depicts the true story around the football team FC Macedonia and its Jewish coach in the Bulgarian occupation zone of Macedonia during World War II. One of the languages spoken in the film besides the local Slavic languages, English, and German is Judeo-Spanish, which is vividly present. In Macedonia, the film stayed in the theatres over a period of months and proved exceedingly successful, which certainly lead to increased awareness among the local population as well as to win praise from Jewish groups.

3.4. METHODS OF RESEARCH AND PRACTICAL STEPS.

3.4.1. APPROACH. Bowern (2011) draws attention to a special approach when it comes to highly endangered languages. Documentation cannot really be built up over a period of time, yet certain temptations should be avoided, such as recording raw data as quickly as possible without being alert to its nature. It is beneficial to identify the number of speakers, their fluency as well as their language skills at the earliest stage. A suggested scenario in which the project could focus on the recording of elderly fluent speakers, which then could be trained in translation, transcription, etc. (Bowern 2011: 469-470) will not be performable in Macedonia anymore. It is highly challenging to record the speakers, who are but a handful and belong to the generation of at least 80+. The link between this generation and the community who expressed the wish to resume the study of the language of their ancestors will have to be a linguist or a member of another Judeo-Spanish speaking community, ideally in a dialect that approximates to the varieties of Macedonia.

---

16 Geographically speaking, both varieties of Judeo-Spanish belong to different isogloss bundles, based on their phoneme inventory: (1) the North-Western group (center: Sarajevo) including Bitola, which was not influenced by any language innovation (conservation of the consonant group –rd–, no reduction of /rr/); (2) the South-Eastern group (centers: Istanbul, Saloniki) including Skopje, situated in a continuum that discloses but
Moreover, the principle of simplifying complex material later on rather than the other way around should prevail (Bowern 2011: 470).

3.4.2. SKILLS. Language documentation and conservation requires skills that differ from general linguistic training. Those are based on descriptive linguistics and directed towards typological and structural matters. Jukes (2011) suggests that beyond basic skilled phonological and morphosyntactic analysis, an amply diversified knowledge of grammatical and other linguistic analysis is needed (Jukes 2011: 424).

Skills relevant to practical fieldwork can be divided into five areas: linguistic theory, fieldwork methods, methods of language conservation, area studies, and field skills (e.g. ethics, see above) (Jukes 2011: 425). Documentation skills that are necessary include proficiency in information and communication technology, and applied linguistics such as orthography development, lexicography, translation, pedagogy, language politics, and advocacy, among others (Jukes 2011: 426).

3.4.3. METHODOLOGY. On the basis of Austin (2006), Endruschat & Ferreira (2011: 15) formulate the different phases of documentation as follows:

- Project design
- Negotiation with the community
- Acquisition of financial support
- Collection (field research) and processing of data
- Formulation of outputs
- Evaluation and final report.

On the technical side, primary data stand in the center of any kind of language documentation and represent the communicational reality of a language community. They can consist of audio and video data, interviews, written sources of such a community as well as metalinguistic affirmations on the part of the speakers about their own language. A corpus of linguistic data will always seek to represent the community in its typology, relating primary data to the context in which they are recorded, analyzed, and archived. These metadata are added data (referring for instance to persons, situations, technical information, etc.) and allow insights into the project realization as well as into the context of the realization of a linguistic occurrence. To complete the corpus, primary data need to be accompanied by annotations whereby the data are transcribed, translated, and embedded into a specific linguistic context (phonetics, prosody, syntax, etc) (Endruschat & Ferreira 2011: 15–20).

Primary data for the documentation of the Judeo-Spanish of Macedonia will, on the one hand, be elicited from handwritten and/or printed texts found in archives, libraries as well as in private collections of literature, and comprise religious and secular texts, poetry, drama, novels and tales, folklore, journals, statutes, etc., and on the other hand, from transcribed interviews. It is to be evaluated to what extent there is divergence between the written and the spoken language.

few differences, which are manifest in the frequency of the realization of the initial /f/ (Quintana Rodriguez 1997: 47f).
As the case may be, interviews can also result in case studies. It lies in the subject nature of the dissertation envisaged that presumably there will be very few interview partners at the disposal of the researcher. The advantage of this is that a case can offer valuable insights both into its own entirety and complexity as well as being able to be set into a wider context, and thus lead to more detailed and profound findings (Mayring 2002: 42).

For access into the field, several interview methods have proved feasible. The interview as a tool for data collection is conducive to gathering answers to the enquiries about the number of speakers, syntactic structures in the discourse as well as tendencies of viability and perspectives of language conservation fairly close to the context. The following interview methods are envisaged:

Problem oriented, semi-structured interview
In this open form of interview, the interviewee is asked to elaborate upon given topics, problem statements, etc. Central questions particularly refer to tendencies and could include the following:

- How does the interviewee assess the language status?
- What are, following personal assessment, the reasons for the decline of the language? What is the interviewee’s position towards this matter?
- Is there interest in language conservation?
- How could such a language conservation be brought about?

Narrative interview
Narrative interviews, which allow for subjective questions asked by the interviewee him/herself, have proved to be less inhibited. Interviewees do not feel confined to the interest of the research with which they are approached. Besides, narratives mostly follow a universally applied framework, whereby the interview is structured in an organic way, when required with the interviewer’s minor assistance. Such natural structuring is conducive to the syntactic analysis.

Group discussion
When, in a group, several subjective questions are asked, this interaction leads to interesting and always dynamic contexts such as faith, community life or identity. In addition, conversations become more open in terms of discourse and register.

The above mentioned interview scenarios have already taken place within the Jewish Community of Sofia, Bulgaria, and are to be applied to the community in Skopje. Should the linguist’s/interviewer’s proficiency of Judeo-Spanish not be sufficient there is the possibility of formulating interview questions and affirmations in Spanish and the interviewees on their part answer in “Ladino”. So far there have not been any challenges in comprehension. The interviewees are not proficient in Spanish on their own account, hence a linguistic assimilation to the interviewer’s language is very unlikely – which would falsify the data collection along with its evaluation drastically. In any case there are persons available who could, in the interviewer’s stead, conduct the interview in Judeo-Spanish.

The respective data acquisition methods follow Mayring (2002).
4. CONCLUSION. The research of the Judeo-Spanish varieties and its relics in Macedonia brings with it obvious benefits for the field of linguistics. The questions about (i) current language status, (ii) the influence which languages in contact exert on one another specifically in the domain of syntax, and especially when a Romance language evolves in a non-Romance environment, (iii) typological and structural identification, (iv) language innovation, and (v) language assimilation within the varieties of Macedonia, all of which to be studied following a corpus-based perspective, combine language documentation with contact-linguistic evaluation.

How can such research, which is valuable in itself, benefit the Jewish Community of Skopje and beyond that ideally contribute to the heritage of Judeo-Spanish as a whole? Simply answering the question about the tendencies for Judeo-Spanish in Macedonia will not suffice, and neither will a linguistic survey presented to a linguistically rather untrained audience. The language documentation needs to disclose the possibilities of pro-active language conservation and revitalization in a community as small as Macedonia and to identify steps that can be taken in ownership of the process.

The community has already taken one decisive step, expressing the wish for language classes in Judeo-Spanish focusing on the aspect of culture. What has dwindled along the last decades can now return to the common awareness of a community. Despite the broad public attention attributed to the Jewish Community of Macedonia with the release of “The Third Half”, it seems more sustainable to raise consciousness at grassroots level within the community itself and then proceed with efforts within the larger public.

It is evident that, at this stage, there is a necessity for external assistance since the number of native speakers is vanishingly small and their age around 80+. Two generations later there will be no local speakers who could carry out the endeavor in situ. In agreement with the Jewish community of Skopje, this task was bestowed upon the author, since her linguistic training includes Romance linguistics with special focus on ancient Portuguese and Spanish (inclusive of studies of Judeo-Spanish), Oriental Studies and now Slavic Linguistics. It is hoped that this knowledge will contribute greatly and fruitfully to the enhancement of Judeo-Spanish within and outside Macedonia.

In an atmosphere of consultation and mutual assistance, where eventually the linguist will play a consulting and assisting role, while the community has ownership and sovereignty over its own endeavors, the response to sustainable language conservation seems to be encouraging.

REFERENCES


18 I would like to express my gratitude for and acknowledge the constant encouragement, accompaniment, and assistance on part of Mrs. Berta Romano-Nikolic, and Dr. Samuel Sadikario-Kolonomos as well as the Jewish Community of Skopje.


Esther Zarghooni-Hoffmann
esther.s.hoffmann@googlemail.com
El árabe ceutí, una lengua minorizada. Propuestas para su enseñanza en la escuela

Francisco Moscoso García
Universidad Autónoma de Madrid

The Arabic of Ceuta is the native language of 40% of the Spanish population of Ceuta, which also speaks Spanish. The remainder 60% is mostly monolingual and their native language is Spanish. There is also 1% of bilingual citizens whose native tongue is Sindhi. The Arabic of Ceuta is Moroccan Arabic, the native language of 60% of the population of the neighboring country and, specifically, it shares common features with the northern dialect area (Yebala region and the Atlantic coast down to the city of Larache). But its use in Spanish territory since the second half of 19th century gave rise to two phenomena: Spanish borrowings and code-switching in the case of bilingual speakers. The Arabic of Ceuta is an oral language, like Moroccan Arabic, which has never been standardized from the political sphere, in contrast with literal Arabic (also called cultivated, standard, modern or classic), which is not the native language of any Arab in the world and has emerged as the only means of educational, political, and cultural expression due to political and religious power. Despite this, there is a whole literary tradition, oral and written, in Moroccan Arabic, especially from the 20th century. Currently, there is a group of Moroccan professors and intellectuals working on its coding in order to generalize a writing system in Arabic script. Ceuta is the Spanish region with the highest school dropout rate in Spain, and this is particularly acute in schools where the majority of students are bilingual. Many experts recommend teachers and professors to teach in the native language of their pupils, at least at the beginning of their education. In this paper we will put forward some proposals for the recognition of Ceuta Arabic as coded by the movement of Moroccan intellectuals who are already working on the development of a dictionary, a grammar, text collections, and translations of works from the European literature to Moroccan Arabic. The ultimate goal should be its inclusion in the educational and administrative services of the city as well as to achieve an official status in the future, rightly recognized by the Spanish Constitution.

1. INTRODUCCIÓN. El árabe ceutí es la lengua materna o nativa de casi el 40% de los ciudadanos españoles de la ciudad de Ceuta. La única lengua oficial que se reconoce en la ciudad es el español, contraviniendo lo dispuesto en el artículo 3 de la Constitución Española, que prevé la cooficialidad de aquellas lenguas del territorio nacional junto a la oficial de todo el Estado, es decir, el español. Ni siquiera existe un reconocimiento formal
El árabe ceutí, una lengua minorizada. Propuestas para su enseñanza en la escuela

en el Estatuto de Autonomía de la ciudad. En Moscoso (2013a) defendíamos la cooficialidad y argumentábamos el derecho a la misma. Un derecho que se le niega a otras lenguas minorizadas en el estado español: el amazige en Melilla, el gallego en Castilla y León, el asturleonés, el aragonés, el catalán en Aragón, el asturiano, el gallego-asturiano, el aranés en Cataluña o el portugués en pueblos y ciudades como Olivenza. Algunos argumentan que el árabe ceutí o el amazige no son lenguas históricas de España. Creo que habría que definir qué es Historia en primer lugar, ya que en ningún lugar he encontrado dónde acaba esta. En el caso del árabe ceutí, su presencia en la ciudad de Ceuta data de la segunda mitad del siglo XIX. La mayoría de sus hablantes accedieron a la nacionalidad española en 1986 (Planet 1998: 25–30 y 85–106), un año después de la entrada en vigor de la Ley de Extranjería. Reivindicamos, por consiguiente, el derecho al reconocimiento del árabe ceutí como lengua histórica de España.

¿Qué es el árabe ceutí? Se trata de una variante del árabe marroquí que se caracteriza por pertenecer a la franja dialectal del norte de Marruecos y por sus peculiaridades, es decir, los préstamos del español y la alternancia de códigos, árabe y español, entre sus hablantes. Sus especificidades han sido descritas detalladamente por Vicente (2005). También hemos publicado algunos cuentos en este registro y un artículo basado en cuestionarios (Moscoso 2004 y 2007). Y sobre la interferencia del español en el árabe ceutí y de esta lengua en aquella, hay un estudio interesante escrito por Rivera (2013: 88–89). Otra cuestión que abordamos detalladamente en dos trabajos (Moscoso 2013a y 2015) es que no se trata de un dialecto del árabe clásico, tal como erróneamente se suele decir de él desde ciertos alminibares no lingüísticos, sino que comparte con este registro unas mismas raíces que remontan a un protoárabe del que poco sabemos. El reconocimiento del árabe ceutí como lengua no pasa únicamente porque así lo sea desde la Administración española, sino también por sus hablantes, quienes afirman que “hablan un mal árabe”, aunque consideran que la conservación de su lengua es una cuestión de “identidad cultural y religiosa” (Vicente 2005: 61–69). Pero esta idea no es más que la visión equivocada que la educación religiosa ha ido inculcando entre los arabófonos, a saber, que la lengua por excelencia es la coránica y que el resto no es más que una vulgarización. Y así, el registro clásico, que no es lengua materna de ningún árabe, ha sido, a lo largo de la Historia, el único vehículo administrativo y, en gran parte, literario y cultural. Al desprestigio de la lengua materna o nativa se unen también el hecho de que esta ha sido una lengua de carácter oral que no ha pasado nunca a una codificación y normalización oficiales, a excepción del maltés, la única lengua árabe contemporánea que se ha normalizado y convertido en lengua oficial de Malta en 1934, y también de la Unión Europea. Su caso, como el de tantas lenguas, puede servirnos de ejemplo y camino a seguir para la codificación del árabe ceutí en el territorio español (Moscoso 2015 y Moscoso 2013b).

¿Por qué entonces se trata de una lengua minorizada tanto por las autoridades españolas como por las religiosas musulmanas de Ceuta? Porque tanto el español como el árabe clásico son instrumentos al servicio del poder, ya que – dice Bourdieu – “no se trata únicamente de comunicar, sino de hacer que se reconozca un nuevo discurso de autoridad, con su nuevo vocabulario político, sus términos para dirigirse y referirse, sus metáforas, sus eufemismos y la representación del mundo social que vehicula y que, al estar ligada esta a los nuevos intereses de los nuevos grupos, es indecible en las hablas locales formadas por usos ligados a intereses específicos de grupos de campesinos” (Bourdieu 2001: 74).1

1 Texto en francés: “Bref, il ne s’agit pas seulement de communiquer mais de faire reconnaître un nouveau discours d’autorité, avec son nouveau vocabulaire politique, ses termes d’adresse et de référence, ses métaphores, ses euphémismes et la représentation du monde social qu’il véhicule et qui, parce qu’elle est liée aux intérêts
Todavía se sigue calificando a la población ceutí de “población cristiano-occidental” y “minoría árabo-musulmana”, o sencillamente “cristianos” y “musulmanes”. Estos calificativos siguen estando bastante enquistados en cualquier institución ceutí, ya sea administrativa o educativa, e incluso investigadora. Stallert (1995: 128) nos da la clave del porqué este tipo de denominaciones al afirmar que Ceuta “ilustra cómo los valores tradicionales del casticismo vuelven a surgir y recobran valor de actualidad en una situación de contacto con el Otro histórico”, lo contrario que sucede en la Península en donde “el nacionalismo étnico español ha quedado arrinconado por los nacionalismos periféricos y donde la etnicidad bajo el signo de la religión es sentida como un anacronismo”. Nosotros nos negamos a este tipo de clasificación en una sociedad aconfesional y en la que todos somos españoles. Desde un punto de vista lingüístico, si tenemos que establecer algún tipo de diferencia, preferimos hablar de ciudadanos monolingües y ciudadanos bilingües.

En esta comunicación reflexionaremos en un primer lugar sobre la necesidad de tener en cuenta la lengua materna o nativa de casi el 40% de los españoles bilingües de Ceuta, empezando por su introducción en el colegio como instrumento de acogida y seguimiento del desarrollo académico de la niña o niño. Y en segundo lugar haremos algunas propuestas del registro lingüístico que se podría emplear en la enseñanza a partir de textos escritos que hay publicados y otros que están en ejecución en estos momentos, así como la creación de un equipo de trabajo cuya finalidad sea la elaboración de material pedagógico.

2. El colegio en Ceuta, un espacio que puede llevar al fracaso educativo. El desarrollo de la lengua materna o nativa es perceptible desde el vientre materno en el que el feto empieza a distinguir sonidos a partir de los seis meses de gestación, prefiriendo el recién nacido la voz de su madre a la de otras mujeres (Brooks et al. 2012: 25). En la segunda parte del segundo año, el niño empieza a hacer uso de sus conocimientos gramaticales, experimentando la adquisición de vocabulario un incremento considerable. A partir de esta etapa empiezan a aparecer las partículas auxiliares, los artículos y las flexiones, y cuando se completan los tres años, el niño ha desarrollado más o menos la gramática (Meisel 2011: 28–30). Es en esta etapa de la vida del niño en la que empieza a contar relatos, aunque sea de forma rudimentaria, y al llegar a los cinco, su narración se manifiesta más coherente (Karmiloff et al. 2005: 238–242).

La niña o el niño ceutíes se incorporan a los tres años al colegio sin que este proporcione las herramientas adecuadas para su desarrollo académico en su lengua nativa. La profesora Rivera (2013: 153–155), ahondando en esta situación, afirma que “[...] los alumnos arabófonos no reciben ningún tipo de input escrito en su lengua materna dentro del entorno familiar, puesto que el dariya es un dialecto oral. Así pues, el estudiante llega a la escuela careciendo de una mínima capacidad de desciframiento de lo escrito y nunca se ha planteado las hipótesis primarias que realizan los niños al comenzar la escritura”. Esto es cierto, ya que los padres desconocen o minusvaloran su lengua nativa, a la que Rivera denomina erróneamente “dialecto oral”, por lo que ya hemos expuesto en la introducción. Y, además – sigue reflexionando esta autora –, el material de enseñanza del español no está enfocado didácticamente para que esta lengua sea enseñada como nueva. Esta situación es a la que Rivera denomina “la paradoja ceutí”, ya que el español no es totalmente una lengua extranjera, pero tampoco es una L1 para estos niños. La única lengua de instrucción es el español, exigiendo al discente su uso sin contemplaciones. En el estudio de la profesora Rivera (2013: 162), es interesante destacar la encuesta realizada a los alumnos arabófonos

nouveaux de groupes nouveaux, est indiscible dans les parles locaux façonnés par des usages liés aux intérêts spécifiques des groupes paysans".
El árabe ceutí, una lengua minorizada. Propuestas para su enseñanza en la escuela

en la que pone de manifiesto que casi el 48,97% dicen que aprendieron el español al llegar al colegio, mientras que el 36,73% dice haberlo aprendido en casa y el 14,28% restante no lo recuerda. A estos datos hay que añadir que el porcentaje de los alumnos arabófonos escolarizados en la escuela pública es del 68,93% (Rivera 2013: 162) y que el fracaso escolar es el más alto de toda España (Rivera 2013: 181). Con esta situación y estos datos, no es de extrañar que los docentes en Ceuta digan que “los alumnos arabófonos poseen unos niveles de comprensión lectora más bajo que sus compañeros castellanohablantes” (Rivera 2013: 189).

¿Y a qué se debe este fracaso educativo entre los arabófonos? Respondemos a esta pregunta con dos reflexiones. La primera de ellas está en relación al bilingüismo de los arabófonos, que Rivera (2013: 151 y 104–108) caracteriza como “sustractivo”, es decir, mientras el español progresa, el árabe ceutí va retrocediendo; “adscrito”, ya que el bilingüismo se adquiere temprano; “bicultural”, por la doble pertenencia cultural; “secundario”, ya que se aprende solo el español en las instituciones académicas; y “equilibrado”, porque se pone de manifiesto la fluidz de las dos lenguas en contextos diferentes. Pero – apostillamos – la fluidz de las dos lenguas se da a nivel comunicativo, no académico, ya que el árabe ceutí está fuera de este ámbito. Y aquí llegamos a nuestra segunda reflexión. Cummins (2002: 94 y 199) distingue en el desarrollo de una lengua dos niveles, el comunicacional y el académico, siendo la adquisición de este más lenta. La lectoescritura se sitúa en el nivel académico, que – según este autor – debe de desarrollarse en la lengua materna o nativa, lo cual ayudará más tarde en el aprendizaje de la lectoescritura en español o en árabe clásico. En relación a esto último, Vygotsky (2010: 226) explica que “al aprender una lengua extranjera, usamos significados de palabras que ya están desarrollados en la lengua nativa, y sólo los traducimos”. La conclusión a estas reflexiones es que la niña o el niño arabófono bilinngüe es educado únicamente en español y que este particular incide negativamente en el desarrollo de su lengua materna o nativa a nivel académico, colaborando así a que se acen-túe el fracaso escolar, cuyas causas también hay que buscar en la falta de apoyo familiar, del profesorado y de la administración educativa (Rivera 2013: 200–208).

¿Se deberían enseñar tanto el español como el árabe ceutí en la escuela? La respuesta no puede ser más que afirmativa. Solo así se podría pasar de un bilingüismo sustractivo a otro de tipo aditivo (Rivera 2013: 105) en el que el desarrollo de la lectoescritura en ambas lenguas permita un equilibrio real tanto en el nivel comunicativo como académico. Otra cuestión, que no vamos a resolver aquí, sino que debe de proponerse al grupo de expertos del que hablaremos en el apartado siguiente, es cómo hacer del colegio un lugar de enseñanza bilinngüe. Habría varios caminos para proponer, aunque un principio básico es la inclusión de la lengua materna o nativa desde la etapa infantil del discente. Se podrían emplear tanto el español como el árabe ceutí desde un primer momento, o solo emplear los primeros años la lengua materna o nativa de los niños para que la lectoescritura se consolide en la L1 antes de pasar a la L2. También somos conscientes y realistas, ya que esta última solución se enfrentaría a la falta de material pedagógico exclusivo existente en estos momentos. Por consiguiente, mientras que no haya un desarrollo considerable de este material, que ya se ha iniciado, lo más conveniente sería que las dos lenguas estuvieran presentes en el aula desde la etapa infantil. Por otro lado, sería interesante observar los resultados que se han dado en otras zonas del mundo bilinngües que han apostado por la enseñanza de las lenguas maternas o nativas minoritarias junto a las de mayor número de hablantes.
3. EL ÁRABE CEUTÍ MODERNO COMO LENGUA DE ENSEÑANZA. Nuestra propuesta de inclusión del árabe ceutí en el colegio parte de un marco legal y educativo que ampara el derecho a que la lengua materna o nativa de los alumnos sea tenida en cuenta, tanto a nivel social como educativo. En primer lugar, tenemos la Constitución Española, la cual avala en su artículo tercero la cooficialidad del árabe ceutí con estas palabras: “1. El castellano es la lengua española oficial del Estado. Todos los españoles tienen el deber de conocerla y el derecho a usarla. 2. Las demás lenguas españolas serán también oficiales en las respectivas Comunidades Autónomas de acuerdo con sus Estatutos. 3. La riqueza de las distintas modalidades lingüísticas de España es un patrimonio cultural que será objeto de especial respeto y protección”. Aunque es cierto que los estatutos de la Ciudad Autónoma de Ceuta ni siquiera reconocen al árabe ceutí como lengua de casi un 40% de la ciudadanía española, solo se apunta a la pluralidad cultural de la ciudad en su artículo quinto. Aquí nos encontramos con un “escollo”, que no deja de ser sino una dificultad que se resolvería con su reconocimiento en las esferas políticas e institucionales.

El segundo documento es la Carta Europea de las Lenguas Regionales y Minoritarias, cuyo comité de expertos pide en un informe sobre su aplicación en España al Gobierno español “aconsejarnos, en colaboración con los hablantes, sobre la situación de las siguientes lenguas, y a incluir comentarios detallados, en su próximo informe periódico, sobre la aplicación del artículo 7 de la Carta a las mismas: el gallego en Castilla y León; el portugués en la ciudad de Olivenza; el bereber en la Ciudad Autónoma de Melilla; y el árabe en la Ciudad Autónoma de Ceuta [...]”. El comité de expertos habla de “árabe” en Ceuta, sin establecer diferencia entre el árabe ceutí y el árabe clásico. Como venimos diciendo, somos de la opinión de que se tenga en cuenta la lengua materna o nativa, es decir, el árabe ceutí. En cualquier caso, el árabe clásico debiera de ser enseñado como una L2. Más adelante, abordaremos esta cuestión.

Por otro lado, y en tercer lugar, nos parecen relevantes las anotaciones de dos informes y una declaración. El Marco Común Europeo de Referencia (MCER) para la enseñanza de las lenguas, dice lo siguiente: “Mantener y desarrollar la riqueza y la diversidad de la vida cultural europea mediante un mejor conocimiento mutuo de las lenguas nacionales y regionales, incluidas las menos estudiadas”. Las recomendaciones pedagógicas de la Unesco, cuyo grupo mixto de expertos, reunido el 09 de mayo de 2009, decía lo siguiente: “La educación básica se imparte en la lengua materna, por lo menos en sus primeras etapas, respetando después las exigencias y necesidades que impone el plurilingüismo”. La Unesco lo tiene claro, la educación básica se imparte en la lengua materna o nativa, es decir en la L1 y no en la L2, que en nuestro caso son el árabe ceutí o español y el español o el árabe clásico respectivamente. Y la Declaración Universal de los Derechos Lingüísticos recoge en dos de sus artículos lo siguiente: “Toda comunidad lingüística tiene derecho a codificar, estandarizar, conservar, desarrollar y promover su sistema lingüístico, sin interferencias inducidas o forzadas (art. 9)” y “Toda comunidad lingüística tiene derecho a que su lengua sea utilizada como oficial dentro de su territorio” (art. 15.1).

En Moscoso (2015: 411–414) hemos presentado como ejemplo para el árabe ceutí a tres lenguas, el español, el griego moderno y el maltés, que fueron codificadas y pasaron a

---

5 unesdoc.unesco.org/images/0018/001871/187101s.pdf (pág. 3, 10 de diciembre de 2015).
ser lenguas oficiales. Convendría, antes de abordar esta cuestión, que tomemos conciencia del valor creativo de cualquier lengua, por muy oral que sea, ya que la escritura no es la que confiere el carácter de lengua a un sistema de comunicación, sino el propio sistema; y que la diferenciación entre dialecto y lengua no es más que una cuestión política. En el caso del árabe ceutí, y de cualquiera de las lenguas árabes que se hablan en el mundo, ya hemos dicho que no proceden del árabe clásico, razón de más para no considerarlas dialecto. El calificativo de “dialecto” a las distintas lenguas árabes que se hablan es fruto del desprestigio que se les ha conferido desde los poderes religiosos y políticos del Mundo Árabe a lo largo de la Historia y en el presente. ¿Es el árabe ceutí árabe marroquí? Sí, igual que el valenciano es catalán o el gallego portugués. Política y lingüísticamente puede ser considerado como un dialecto del árabe marroquí; pero también, desde los mismos presupuestos, puede ser tenido por una lengua propia del territorio español, ya que posee – y ya apuntamos a ello en la introducción – sus especificidades. Basándonos en el esquema propuesto por Bibiloni (2007) para la codificación y normalización de una lengua, proponemos el siguiente proceso:

a) Selección: Como alternativa al calificativo de diglosia (Ferguson 1959) con el que ha sido definida la lengua árabe, nosotros propusimos en relación al árabe marroquí, pero bien pudiera valer para la mayoría de árabes del mundo, incluido el de Ceuta, el de “pentaglosia” (Moscoso 2010), que se acerca de una manera más comprensible, al carácter multiglosico (Ferrando 2001: 135) de la lengua árabe. Y así, en Ceuta, podemos encontrar: 1. el árabe antiguo, en el que está escrito el texto coránico que se enseña en las escuelas coránicas de la ciudad o se emplea en el rezo. 2. El árabe moderno o estándar, también llamado culto o estándar, que es el registro empleado en el mundo árabe en la educación, la administración y las relaciones internacionales; este árabe es el mismo que el antiguo, pero más asequible a la modernidad y, al igual que el primero, no es lengua materna o nativa de nadie hoy en día. 3. El árabe ceutí, que es el hablado por la población española bilingüe arabófona en general entre ellos, con sus características propias. 4. Los distintos dialectos del árabe ceutí que se hablan en las distintas zonas de la ciudad; es cierto que, al ser un territorio tan pequeño, las diferencias no son tan grandes, pero sí las hay. 5. El árabe ceutí moderno, que es un árabe con la estructura del árabe ceutí y enriquecido con voces y expresiones del árabe estándar. La incidencia de este registro en el árabe ceutí no es tan elevada como la que encontramos en Marruecos, ya que la población arabófona está sometida a la imposición del español a través de la escuela o los medios de comunicación y no tanto del árabe estándar. Pero sí pensamos que el árabe ceutí moderno debiera ser el registro que empleemos en la educación y en otras esferas sociales, ya que toda lengua necesita préstamos para adaptarse a la modernidad y, sobre todo, al lenguaje académico. Somos de la opinión que estos préstamos deben de ser incorporados a partir del árabe estándar, ya que este registro posee esta terminología. Y, por otro lado, porque en los países árabes en general, y Marruecos en particular, los préstamos del árabe estándar cada vez están más presentes. Sería una forma de establecer vínculos entre el árabe ceutí y el resto de árabes, así como con el árabe estándar. Si tenemos en cuenta también este registro como L2 en la escuela, el paso desde la lengua materna o nativa, el árabe ceutí, al árabe estándar será más rápido, ya que son dos lenguas con las mismas raíces.

b) Codificación: Para la codificación del árabe ceutí moderno proponemos en primer lugar el uso de la grafía árabe, ya que esta serviría de vínculo de unión con el resto de los pueblos árabes. Es cierto que podemos escribir cualquier lengua en cualquier alfabeto – el
El árabe ceutí, una lengua minorizada. Propuestas para su enseñanza en la escuela

El árabe maltés ha empleado la grafía latina –, pero pensamos que el uso de la grafía árabe serviría también para establecer vínculos de unión con el árabe estándar, que debiera ser una de las fuentes principales de préstamos. En cuanto a la grafía árabe, nuestra propuesta es incorporar las pautas de escritura que el profesor Abderrahim Youssi, de la universidad Mohammed V, emplea en las traducciones de obras de la literatura europea al árabe marroquí. Su objetivo es demostrar que el árabe marroquí es una lengua creativa y que estos textos son un instrumento importante para luchar contra el alto índice de analfabetismo que sufre el país, casi el 50%.

El registro que emplea este profesor es el árabe marroquí moderno, que, al igual que el árabe ceutí moderno, incorpora voces y expresiones del registro estándar con el fin de acercar al árabe marroquí a la modernidad. Su propuesta de uso de la grafía, entre otros aspectos, está basada en el acercamiento de las raíces entre las dos lenguas árabes, el árabe marroquí y el árabe estándar, permitiendo de esta forma que el paso de la lengua nativa a la estándar pueda hacerse con menor dificultad. En esta línea, Youssi publicará próximamente una gramática del árabe marroquí moderno en la que se exponen detalladamente, no sólo los contenidos de esta, sino sus propuestas de codificación. Así como una colección de textos con pautas pedagógicas para su uso en el colegio. Y por otro lado, un grupo de profesores está trabajando actualmente en la redacción de un diccionario de definiciones enteramente en árabe marroquí. A todo esto, habría que sumar las diferentes publicaciones que se han ido haciendo en los últimos años: novelas en árabe marroquí, revistas, recopilación de tradiciones populares (historias, cuentos, adivinanzas, proverbios, ...), poesía contemporánea y un largo etcétera de textos que debieran ajustarse a unas normas de codificación unificada.

En esta etapa de codificación, se trataría de ir escribiendo material pedagógico en árabe ceutí, empleando la grafía árabe y la propuesta de escritura del profesor Abderrahim Youssi. Los textos en árabe ceutí moderno se caracterizarían por sus rasgos particulares en relación al árabe marroquí moderno. Por el momento, se podrían emplear los textos ya publicados en Marruecos y adaptarlos a las especificidades del árabe ceutí. Nosotros publicamos (Moscoso 2004) una pequeña colección de cuentos en árabe ceutí que también puede ser empleada. Los encargados de la recopilación y escritura de estos textos sería un grupo de expertos designados por el comité, del cual hablaremos más adelante.

c) Aplicación: Una vez que se dispusiera de un material medianamente aceptable, habría que aplicarlo a nivel social. Su aplicación se haría efectiva, en un primer momento, en la educación, para luego plasmarse en otras instancias: administrativas y sociales en general. Pensamos que, mientras se va elaborando más material pedagógico que incluya distintas disciplinas, el árabe ceutí moderno debiera de emplearse en la etapa infantil con la misma presencia y derechos que el español, aplicándose desde un bilingüismo aditivo. Y que, en la etapa primaria, debiera de emplearse como lengua moderna de enseñanza, tanto para bilingües como monolingües, siendo además un instrumento de apoyo para aquellos alumnos que empiezan su educación en la etapa primaria, sin haber pasado por la infantil.


El grupo de expertos estaría encargado de ir valorando la puesta en práctica de su enseñanza en la escuela y cómo ir mejorándola, desarrollando al mismo tiempo material pedagógico que cubra las necesidades detectadas.

d) Elaboración: Siguiendo con lo planteado en la etapa de aplicación, en esta se persigue la constante actualización y creación de nuevo material pedagógico y todo aquel dirigido a la administración y la sociedad en general. Se hace necesaria la labor de traducción al árabe ceutí moderno y la creación de nuevos textos, no sólo pedagógicos, sino también literarios.

Para la puesta en práctica de este proceso, se hace necesaria, como hemos dicho, la creación de un comité de expertos, que debiera estar formado por profesores de infantil, primaria y secundaria, expertos lingüistas, representantes de las administraciones educativas, representantes de los partidos políticos, representantes de los padres de los niños escolarizados y representantes del resto de estamentos de la sociedad civil (tanto bilingües como monolingües). La hoja de ruta, retomando lo expuesto anteriormente, podría guiarse por las siguientes pautas:

- Selección del registro: árabe ceutí moderno, cuya estructura sería el árabe ceutí estándar, o aquellos rasgos más empleados, con vocabulario y expresiones del árabe estándar.
- Codificación: emplear la grafía árabe y las propuestas del profesor Abderrahim Youssi.
- Aplicación: en un primer momento en el colegio y más adelante en la Administración y las instituciones sociales en general. Para ello, se requiere que el comité de expertos designe a un grupo de especialistas que adapten los textos existentes en árabe marroquí moderno al árabe ceutí moderno y que lleven a cabo traducciones.
- Elaboración: de material pedagógico y otro.
- Incentivación de la creación en árabe ceutí moderno, de tal forma que se escriban textos literarios: poesía, novelas, relatos, etc.
- Creación de un grupo de expertos que supervise la aplicación de la lengua en la escuela y revise el material pedagógico empleado.
- Reconocimiento del árabe ceutí en el Estatuto de Autonomía de la Ciudad.
- Petición de la cooficialización de esta lengua al Gobierno Español.

4. CONCLUSIONES. A modo de conclusión, creemos que se hace necesario el reconocimiento del árabe ceutí, primero como lengua minorizada entre las que están presentes en el Estado Español, pero no el árabe clásico o estándar, que no es lengua nativa de ningún arabófono. Esperamos que nuestros colegas universitarios dedicados al estudio de las lenguas de España, incluyan también al árabe ceutí y al amazige como lenguas históricas y de pleno derecho del Estado Español. En segundo lugar, se hace necesario su reconocimiento en el Estatuto de Autonomía de la Ciudad de Ceuta. Y en tercer lugar su cooficialidad dentro del Estado Español. El índice de fracaso escolar en Ceuta, el más alto
El árabe ceutí, una lengua minorizada. Propuestas para su enseñanza en la escuela

...del Estado Español, más acentuado en aquellos colegios en los que la presencia de bilingües es mayor, hace necesaria una educación bilingüe aditiva. Para ello, es imprescindible que la Administración tome medidas encaminadas a la creación de un Comité de Expertos que sienten las bases de una hoja de ruta que permita la codificación del árabe ceutí moderno y la elaboración de material escrito que convenga a los intereses del desarrollo educativo del niño arabófono que entra en el sistema educativo español.

**References**


Moscoso García, Francisco. 2013b.


Francisco Moscoco García
francisco.moscoso@uam.es
Multilingualism and structural borrowing in Arbanasi Albanian

Jana Willer-Gold, Tena Gnjatović, Daniela Katunar, and Ranko Matasović

University of Zagreb

In this paper we present a brief overview of the history of linguistic contacts of Arbanasi Albanian, a Gheg Albanian dialect spoken in Croatia, with Croatian and Italian. Then we discuss a number of contact-induced changes in that language. We show that Arbanasi Albanian was subject to strong influences from Croatian (and, to a lesser extent, from Italian) on all levels of linguistic structure. Using the data from our own fieldwork, we were able to show that there were also influences on the level of syntax, including the borrowing of certain constructions, such as analytic causative and imperative constructions, as well as the extension of the use of infinitive in subordinate clauses.

1. INTRODUCTION. Arbanasi Albanian (AA) is a Gheg Albanian dialect currently spoken in a suburb of Zadar, Croatia. Its speakers moved to Zadar in the early 18th century. They originated from a region on the border of Albania and Montenegro, which was ruled by the Ottoman Empire at the time. Since their arrival in Zadar, the speakers of Arbanasi Albanian lost all contact with other Albanian speakers, and their idiom was heavily influenced by the surrounding Čakavian dialect of Croatian, as well as by Italian, which was the language of education and government during much of the history of Zadar before 1945 (Barančić 2008).

The Christian Arbanasi emigrated to Zadar (Zara) from their former homeland in three waves in the beginning of the 18th century: The 1st emigration in 1728 brought 21 families with 450 members from three villages around the Lake Scadar – Šestana, Briske (and Livari) (south of Ulcinj). The Arbanasi first arrived to the Venetian port of Kotor, whence they were transported to Zadar; the 2nd migration in 1733 brought 28 families counting 199 people from Šestana and surrounding villages, and the 3rd one in the year 1754 also included a large number of immigrants. Apart from Zadar itself, the speakers of AA also settled in villages around the city, today called Prenđe, Šestani, Ćurkovići, and Paleke, as well as in Zemunik, Dubrovnik, and Kotor. However, during the following three centuries only the Arbanasi in Zadar have resisted integration, and therefore they are of interest for this study.

The Arbanasi fled from the Ottoman Empire mainly to avoid recruitment into the Ottoman army, and in part because of religious oppression. They fled to Zadar, then under Venetian rule, and settled there with the help of the archbishop of Zadar, Vicko Žmajević. Their migration has to be seen in light of Venetian efforts to repopulate the devastated parts of Dalmatia. One of the local magnates in Zadar, Erizzo, provided the first families to arrive with a house and land just outside the town walls. This led to the founding of the
Multilingualism and structural borrowing in Arbanasi Albanian

settlement named Arbeneši by Arbanasi, Borgo Erizzo by Italians, and Varoš Erćina by Croats (after the magnate Erizzo). Later all three communities started calling it Arbanasi. Arbanasi is today part of the town of Zadar.

The town of Zadar and the community of the Arbanasi shared the fate of Dalmatia. After the collapse of the Republic of Venice (1797), they were ruled by Austria, which evolved into Austro-Hungarian monarchy (except for a brief French rule during Napoleon). Under Venetian and Austro-Hungarian rule, most education in Zadar was conducted in Italian and, at lower levels, in Croatian. In 1748, archbishop Zmajević established a Glagolitic school, educating young clergy that would later disseminate nationalist ideals in northern Dalmatia. Some educated Arbanasi clergy wrote letters in Croatian, in Glagolitic script, and during the period of the Croatian national revival, Arbanasi cultural and intellectual elite matched that of Croatians. Yet the community was ethnically (if not linguistically) divided, as some of its members became ethnic Italians. In 1896, a primary school was established where the pupils were taught in Italian, with only two lessons a week in Croatian and AA. There was also a Croatian primary school, where most lessons were given in Croatian. In 1901, learning Albanian became obligatory in the Croatian school in Arbanasi, but only for those pupils whose first language was AA.

In the period of the two world wars, 1920-1943, during the Italian Fascist occupation of Zadar and Arbanasi, the Croatian language was forbidden. The Arbanasi Albanian language was tolerated at first, but during the later years of the Fascist regime it was forbidden to speak Albanian and to teach it at school.

In 1945, Zadar became part of Croatia and Yugoslavia, and most speakers of AA who were ethnically Italian either chose to leave their town, or were expelled by the Communist regime. The Arbanasi Albanian language was not taught in school, and until quite recently there were very few attempts at its cultivation and preservation.

1.1. ETHNOLINGUISTIC IDENTITY. As there has been no contact between Arbanasi Albanians and the Albanians in their homeland until the very end of the 19th century, the cultural values they have kept until today belong to the times when they left Albania. Only recently, since the middle of the 20th century, contacts with Kosovo and Albania have been revived, as new immigrants started arriving from these areas.

When discussing the degree to which the Albanian Gheg dialect has been preserved in Zadar for the last three hundred years, some extra-linguistic factors have to be taken into consideration. First, oral tradition is the primary means of linguistic transmission of the AA culture. Social meetings (attended mostly by women) where folk stories were told were an important vehicle in the transmission of the language. Secondly, a strong sense of community allowed Arbanasi to stay aware of their uniqueness when compared to the Croatian-speaking majority surrounding them. Thirdly, in many cases of mixed marriages with the local population the spouse would move and be integrated into the Arbanasi family. More often than not this allowed the spouse to acquire the Arbanasi language to the degree of first language proficiency.

1.2. THE CURRENT STATUS OF ARBANASI ALBANIAN. Today, the dialect is endangered, with less than 200 fully competent speakers, although there are probably around 500 people that understand the language to some extent (Kovačec 2002). It is generally not written, except in a handful of occasional publications (journal Feja) and in the works of collectors of traditional lore (Stipčević 2011).
The attitude of the speakers towards their language and its use is generally positive, and the same holds for all of the informants that have participated in our research. Since the speakers of Arbanasi Albanian are ethnically Croats and confessionally Catholic, like the large majority of the inhabitants of Zadar and its surroundings, they generally do not face any ethnic or religious discrimination. The use of their language is not stigmatized, but until quite recently it has not been particularly encouraged. Our informants have not reported any problems associated with the use of their mother tongue.

1.3. HISTORY OF RESEARCH. The first scholar who wrote about the Arbanasi dialect of Albanian was the great Slavic scholar Franz Miklosich. In his *Albanische Forschungen* (1870–1871) he provided some information about the Arbanasi dialect and its speakers, but at that time Albanian linguistic studies were still in their infancy. Tullio Erber (1883) was primarily interested in ethnography and history of the Arbanasi community, but he also gave a brief account of their language. The idiom was described briefly by Gustav Weigand (1911), and in 1937 the Italian linguist and Indo-Europeanist Carlo Tagliavini wrote the first grammar of Arbanasi Albanian. Since he spent only a few weeks in Zadar and consulted only a handful of informants, his description is sketchy and unreliable. Moreover, he was interested mostly in historical phonology and etymology, and a large part of his book is devoted to the etymological dictionary of the words from his corpus. His account of the morphology is brief and incomplete, and his book contains very few remarks about the syntax. More extensive is Ismail Ajeti’s Ph.D. thesis, published in Serbian in 1961, in which the Arbanasi dialect was compared with the other Albanian dialects. However, even Ajeti’s focus was primarily diachronic, and he seems to have been more interested in sound laws deriving AA vowels and consonants from Proto-Indo-European than in the exhaustive description of the idiom. Even though his account of the AA phonology is more detailed and reliable than Tagliavini’s, he still failed to give a complete description of the morphology, and the information he gives about the syntax is scarce and impressionistic. Both Tagliavini’s and Ajeti’s books contain folklore texts they recorded among the speakers of AA, and these can be used to compare the syntax of the language around the middle of the 20th century with the present situation.

The only dictionary of AA was published by a Croatian scholar, Kruno Krstić, who was a native speaker of the language, in 1987. His dictionary contains around 4500 entries, and it is especially valuable because the use of words in the dictionary is amply illustrated with examples collected by Krstić himself.

Finally, some up-to-date information on AA can be found in the brief encyclopedia article by August Kovačec (2002), and some sociolinguistic problems of language identity among Arbanasi Albanian speakers are discussed by Maksimilijana Barančić (2008), also a native speaker of the language. Aleksandar Stipčević’s book about the traditional culture of the Arbanasi (2011) is a rich source of texts that the author collected himself or re-published from earlier sources. Furthermore, some occasional publications containing folklore texts have appeared during the last decade (e.g. *Feja* 2005). However, it is still fair to say that Arbanasi Albanian is linguistically under-studied (especially with regard to its syntax), and that a larger collection of texts in the idiom would be a big desideratum.

As we saw, all of the research published about AA was written from the point of view of traditional dialectology, and important issues of syntax were left undescribed, especially with respect to contact-induced changes affecting the Arbanasi speech during its three hundred years of co-existence with Croatian and Italian.
2. LINGUISTIC INFLUENCES ON ARBANASI ALBANIAN. In the following chapters we will provide an overview of lexical, phonological and morphological data on language contact between AA and Croatian and Italian, supported by data from our own research.

2.1. LEXICAL INFLUENCES. Lexical borrowing occurred both from Croatian (Cro.) as well as Italian (Ital.), although Krstić’s dictionary (1987) notes mostly Croatian loanwords, many of them borrowed from the local Čakavian dialect, e.g. AA bodull “islander” < Cro. bodul vs. Standard Cro. otočanin. Some loanwords seem to have been international words also borrowed via Croatian, e.g. AA gazet “gazette, newspaper” < Cro. gazeta. As is usual in such situations of linguistic contact, semantically full words were the most commonly borrowed ones (i.e. nouns, verbs, and adjectives, e.g. AA junc “calf” < Cro. junac, junica, AA brod “beard” < Cro. brada, AA pole “field” < Cro. polje (Čakavian pole)). An interesting case is AA breg “hill”. This word was borrowed in Common Albanian from Slavic at a very early stage, and its original meaning was “shore”, as in Standard Albanian (Std. Alb.) breg (cf. also Russian bereg, which preserves the original Slavic meaning). The meaning “hill” in AA developed under the influence of Cro. brijeg (Čakavian brijeg), which means “hill”. It is interesting to note that in addition to semantically full words, some conjunctions were borrowed as well, e.g. AA ma “but” < Ital. ma, and the subordinative conjunction AA da “that” < Cro. da. Krstić also notes the borrowing of the emphazer Cro. baš > AA bash “very, really”.

2.2. INFLUENCES IN PHONOLOGY AND MORPHOLOGY. Arbanasi Albanian underwent a number of sound changes, mostly described by Tagliavini (1937) and Ajeti (1961). Although the details of these changes need not concern us here, it is important to note that most of them led to phonological convergence between the Arbanasi dialect and the surrounding idioms, both Čakavian Croatian and Italian (including the Venetian dialect):

- Gheg nasal vowels are lost, e.g. AA an “moon”, Gheg. ān, Std. Alb. hënë. Neither Italian (Venetian) nor Čakavian Croatian have nasal vowels.
- The phoneme /b/ was lost: AA und “nose” (Std. Alb. hundë), AA anger “eat” (Std. Alb. hëngrë). Italian and Čakavian also lack /b/ as a phoneme.
- The palatalized velars became affricates, i.e. gj > dž (AA gjum “sleep” is pronounced [dʒum]), q > ě (AA qen “dog” is pronounced [tʃen], cf. also AA pleq “old men” [pletʃ]); both Croatian and Italian do not have palatalized velar stops, but they do have affricates [dʒ] and [tʃ], phonetically very similar to the AA phonemes.
- While Albanian generally distinguishes two different vibrants, rr [R] and r [r], in AA the two sounds merged. The single remaining vibrant is very similar to Croatian and Italian [r], cf. AA ar-a “nut”, Std. Alb. arrë.
- In the idiom of most speakers th [θ] > s [s] (cf. AA san “said” vs. Std. Alb. thane) and dh [ð] > l [l] (AA le “earth” vs. Std. Alb. dhe, AA mal “big” vs. Std. Alb. madh). Tagliavini (1937) noted the change in progress, but recorded that many speakers still differentiated interdental spirants th and dh.

Similarly, some morphological changes occurring in Arbanasi Albanian made it more similar to Croatian (Čakavian) and Italian. Among such changes, we may note the elimination of plurals formed by ablaut, e.g. AA dora “hands” (pl. of dorë), cf. Std. Alb.
It is also interesting to note that AA borrowed the vocative case ending –e from Croatian in some nouns, chiefly those denoting close relatives, e.g. tate “o father”, nane “o mother”. Other varieties of Albanian do not have the vocative case.

3. **Present research.** During our research visit to Arbanasi in March 2012 we interviewed ten informants. All were native speakers of Arbanasi Albanian, i.e. bilingual in Arbanasi Albanian and Croatian, and they belonged to different age groups (from mid-twenties to mid-sixties). Nine were male and one was female, and all of them claimed to be using the language every day. We collected our material by elicitation, asking the informants to translate sentences and short texts from Croatian into Albanian. Through these tasks we tested the fluency of our informants and the richness of their vocabulary, and assessed the degree of syntactic interference between Croatian and Arbanasi Albanian.

For the purposes of our analysis it was necessary to distinguish the phenomena of linguistic interference and code-switching, in order to be able to analyze the instances of the former in our data. Therefore, we will discuss some of the literature on these phenomena in the next chapter.

3.1. **Language interference or code switching?** The terminology in studies dealing with language contact phenomena is often inconclusive when it comes to clearly distinguishing code-switching from language interference. This is clearly illustrated by various terms that can be used to refer to the two, for instance code mixing, language mixing, language transfer, convergence, intersystemic influence, among others (Treffers-Daller 2012).

Although both terms stem from the field of contact linguistics, their distinction seems to depend on various factors, including the goal of a linguistic analysis (for instance bilingual or sociolinguistic studies), whether the speech event is made continuous or discontinuous by the language switch, the linguistic levels appropriated from one language into the other and so forth (Auer 1998, Skiba 1997). Code-switching is a relatively new term, introduced into bilingual studies because of the need to stress the online switch made between two language systems in bilingual language use (see Auer 1998). From this perspective, code-switching is primarily a usage-based phenomenon which presumes parallel existence of two rather clear cut language systems in the mind of a bilingual. The switch occurs due to various communication factors, for example establishing the familiarity of speakers or the societal status and meaning one language carries over the other (Llamas et al. 2007). Moreover, as Treffers-Daller (2012) states, the notion of speaker control is more present during code-switching than language interference situations. We came across such examples in our research, where speakers of AA would use a Croatian word because of their lack of vocabulary in AA, in order to continue with the task at hand.

Language interference, on the other hand, relies more on the notion of a “language mixture”, a blend exhibiting the competing properties of more than one language at a time. This notion of a “mixture” has been addressed in works by Haugen (1956, 1972), who deems it rather vague when it comes to describing properties of the process of borrowing. In an attempt to classify stages of language contact, Haugen (1956) distinguishes “switching” from “interference” as two subsequent stages of an ongoing process of diffusion. According to Haugen, “switching” is the first stage, with alternate use of lexical items from L1

---

1 Our informants used only one general past tense roughly corresponding to Croatian perfect.
“Interference” is the second stage, where two languages overlap, and it is followed by the third and last stage, “integration”, where the linguistic structures from L1 are fully integrated in the system of L2, and there is no overlapping, except in a historical sense (Haugen 1956). Furthermore, loanwords seem to vary according to the level of their morphological, phonological or syntactic integration within a new linguistic system, and thus full integration on all three levels is usually considered to be a classic case of a completed transfer. Also, transfer of certain syntactic constructions can be related to the pragmatic salience induced by communicative needs (Matras 2012).

To differentiate the cases of code-switching and language interference in our own AA data we used two criteria: A lexical item was considered to be a Croatian or Italian loanword if it is a) attested in Krstić’s dictionary (1987), or b) it is found in at least two of our sources (including our informants). Krstić (1987) makes explicit reference of his attempt to put only “proper” AA words in his dictionary, and the words in question demonstrate the phonological, morphological and syntactic integration discussed above. The inclusion of a word in the vocabulary of more than one speaker means that we were able to separate nonce loanwords from conventional loanwords. Furthermore, words are considered to be adapted to Arbanasi Albanian if the speakers can inflect them according to Albanian grammatical patterns.

3.2. STRUCTURE OF THE QUESTIONNAIRE. The questionnaire consisted of five sections:

a) Narrative translation – Cinderella story

The first task for the informants was a translation of a short story about Cinderella from Croatian into Arbanasi Albanian. We chose the Cinderella story because we assumed that all the informants would be familiar with it. We asked the informants to read the story one paragraph at a time and translate it orally. Our objective was primarily to assess their fluency, but also to check how certain syntactic constructions would be translated when encountered within a longer text instead of just a single isolated sentence. For example, we used infinitival as well as finite subordinate clauses in order to see whether the informants would differentiate between these two structures in their translations. We also used constructions with a causative meaning to check whether we can elicit the expected Italian-type causative construction “make + INF”. One part of the story was construed so as to potentially elicit the admirative construction which exists in Albanian, but has not been described in the literature on the Arbanasi Albanian language. The Cinderella story also served to distinguish between the highly fluent speakers and those whose language seems to be influenced by Croatian to a greater extent.

b) Subject / object control sentences

The second part of the interview consisted of a set of semantically unrelated sentences with subject and object control constructions, which the informants were again asked to translate. Some of the sentences contained an infinitival clause, while in others the verbal complement consisted of a finite subordinate clause. The task was divided into three parts throughout the interview in order to avoid too much repetition. For example, a sentence meaning “He needs to wash windows constantly” appeared at two separate times in the questionnaire, construed in two different ways – once with an infinitival complement, and the other time with a finite clause introduced by the subordinator da “that”. The aim was

---

2 Some of these examples can include the transfer of e.g. imperative constructions (see also our data below).
to see whether the informants’ choice of the construction in Arbanasi Albanian would be influenced by the structure used in the Croatian examples. Our findings on this specific construction will be discussed in more detail in the next chapter.

c) Vocabulary check – newspaper texts

In the following section of the questionnaire the informants were asked to translate orally two very short newspaper texts – a weather forecast and an excerpt from an article on the parliamentary elections in Croatia – and a brief passage from a high-school chemistry book. This task was designed to show us how rich the informants’ vocabulary is in three very specific domains in which we expected them to use primarily Croatian. The results proved our expectations correct, the informants admittedly could not translate most of the technical vocabulary related to politics, chemistry, and meteorology into Arbanasi Albanian and used Croatian lexemes instead.

d) Admirative construction

Another set of Croatian sentences that the informants were asked to translate were construed so as to elicit the admirative construction, in order to see if it exists in the language. However, the data we collected from our informants does not suggest that this construction is present in Arbanasi Albanian.

e) Phonological task

In the phonological task we presented the informants with twelve pairs of Croatian words, the Albanian equivalents of which were minimal pairs differing in either $dh - l$ ([ð] – [l]), $rr - r$ ([R] – [r]), $th - s$ ([θ] – [s]), or $ll - l$ ([H] – [l]). Our aim was to check whether the distinction between these phonemes has been lost in the language of our speakers, and the data we collected shows that this is indeed the case: i.e. $dh$ and $l$ merged into $l$, $rr$ and $r$ into $r$, $ll$ and $l$ into $l$, and $th$ and $s$ into $s$.

f) Sociolinguistic questionnaire

In the final part of the questionnaire we asked the informants about their attitudes regarding the Arbanasi Albanian language, attempts towards its preservation, and their personal experiences as speakers of this language within the wider Zadar community. Though they all spoke their language on a day-to-day basis with their friends and/or family, most of them seemed rather skeptical about the preservation of the language and recent attempts made by the local community to revive Arbanasi Albanian by organizing courses, workshops etc. One part of the informants considers it a good idea to introduce Arbanasi Albanian courses in kindergartens and schools, while others believe that there are not enough young people and especially children who are fluent in the language, and, as one of the informants explicitly said, language courses are not an adequate substitute for everyday communication and spontaneous language use. All the informants would like to see their language survive, but in general they seem to think it is too late for that, given that the speakers have stopped teaching their children the language. The informants also explained that being bilingual in Croatian and Arbanasi Albanian brings neither significant advantages nor disadvantages in terms of job opportunities, social status, etc. However, given that there are occasional cases of discrimination towards Albanians in general in Croatia, some mentioned that their origin sometimes, though rarely, does trigger a somewhat negative attitude toward them, most often during arguments or fights.
4. SYNTACTIC INFLUENCES ON ARBANASI ALBANIAN. In this chapter we make explicit the results in our data specific to syntactic influences of Croatian and (to a lesser extent) Italian on Arbanasi Albanian. We have observed that influences from Croatian and Italian are evidenced in various syntactic structures of Arbanasi Albanian, including the calquing of certain constructions (4.1, 4.2, 4.3) borrowing of complementizers (4.4), and the extended use of the infinitive instead of the finite verbs in subordinate clauses (4.5).

4.1. THE ROMANCE-TYPE CAUSATIVE. The Romance-type causative with the verb “to do” (Ital. fare, Cro. Čakavian činit) is calqued:³

1. Ai i ko bo me iq but-in [Ta 53]
   3SG 3PL.ACC AUX.3SG do to raise barrel-ACC.SG.DEF
   “He made them raise the barrel.”

2. E ko bo me sua [T]
   3SG.ACC AUX.3SG do to learn
   “He made him learn.”

3. Ai ko bo soldata m’u qesh [T]
   3SG AUX.3SG do soldiers to-REFL laugh
   “He made soldiers laugh.”

This type of causative construction is otherwise unattested in Albanian (both Gheg and Tosk). It developed under the influence of Italian, whence it also spread to Croatian Čakavian dialects. The Arbanasi Albanian examples are structurally perfectly equivalent to Italian causatives, such as for example La frase li ha fatti ridere “The sentence made them laugh”. One cannot establish with certainty whether Arbanasi Albanian acquired it directly from Italian, or through Croatian (Čakavian) intermediary.

4.2. THE CROATIAN CONSTRUCTION čini se da. The Croatian construction čini se da “it appears that” (lit. “it makes that”) is calqued:

4. Po mu bajet da ko ro shi [S]
   PROG 1SG.DAT do.PASS.3SG that AUX.3SG fall rain
   “It appears to me that it has rained.”

5. Më baet se e ko lan zjarmi [Kr. 132]
   1SG.DAT do.PASS.3SG that 3SG.ACC AUX.3SG leave heat.NOM.SG.DEF
   “It appears that the heat will stop.”

It is relevant to note that, although AA has the verb meaning “to seem/appear”, in this structure they use the verb bo “to make”, a literal translation from Croatian, with the clitic personal pronoun rather than the reflexive pronoun.

4.3. The Croatian particle *nek*. The imperative construction with the Croatian particle *nek* is borrowed. Unlike the standard language and the Gheg dialects closest to it, Arbanasi Albanian uses the particle *nek* in the 3rd person of the imperative mood (both singular and plural), cf. AA *nek vinje* “let him come”, *nek vinjen* “let them come” (Ajeti 1961: 140). The particle *nek* has an equivalent use in the Čakavian dialect of Croatian, as in (6):

(6) \[\text{Nek dođu } \text{IMP:particle come.PRS.3PL} \]

“Let them come.”

In Standard Albanian, the imperative of *la* “let” is used in this construction, together with the subjunctive of the inflected verb, cf. Std. Alb. *le të vinjë* “let him come”.

4.4. The Croatian subordinator *da*. The Croatian subordinator *da* was borrowed. In those instances where AA uses full subordinate clauses, rather than infinitives (to be discussed below), the Croatian subordinator *da* “that” is used alongside the inherited subordinator *se* “that”. We have not been able to discover any systematic difference in the use of these two subordinators, which seem to be quite synonymous. The Croatian loanword *da* is attested already in the texts collected by Tagliavini (1937), but it still has not ousted completely the Albanian subordinator *se*. Example (7) illustrates the use of *da* with a complement clause, and examples (8) and (9) its use with an obligatory control construction.

(7) \[\text{Ató i kan than da nu(k) ko kúrgj [Ta]} \]

3PL 3SG.DAT AUX.3PL say that NEG have.3SG nothing

“They told him that there was nothing.”

(8) \[\text{Po mendoj da kam me vot nesër në qine [S]} \]

PROG intend.PRS.1SG that AUX.1SG to go tomorrow in cinema

“I intend to go to the cinema tomorrow.”

(9) \[\text{Marko ko san da ko me art [T]} \]

Marko AUX.3SG say that AUX.3SG to come

“Marko said he would come.”

4.5. The infinitive. The use of the infinitive (formed with *me* + participle) is extended to nearly all types of subordinate clauses. Example (10) exemplifies the inchoative construction, example (11) the obligatory object control construction, and example (12) the obligatory subject control construction. It is worth noting that in Croatian, the equivalent of example (11) would involve the use of subordinator *da* and the finite verb, rather than the infinitival construction, while in Italian the infinitive would be used. It appears that AA has generalized the use of infinitives even in those cases where Croatian prefers the use of finite subordinate clauses.

(10) \[\text{Ali gjël-i ko zan me kantát [Ta]} \]

but rooster-NOM.SG.DEF AUX.3SG start to sing

“But the rooster started to sing.”
(11) Gjuqi i ko lishua me fol [T]
judge 3SG.ACC AUX.3SG let to speak
“The judge let him speak.”

(12) Ko aruá me vot në tempo [V]
AUX.3SG forget to go in time
“He forgot to leave on time.”

Standard Albanian (and Tosk dialects generally) uses subordinate clauses with finite verb forms in all of the constructions listed above; the Gheg dialects use the infinitive in some of these constructions, but not as frequently as Arbanasi Albanian, where the finite verb is used in subordinate clauses only with verba dicendi, e.g. san “say”:

(13) Po folin da ko Gjani në brombi kap shum peshki [S]
PROG say.PRS.3PL that AUX.3SG Gjani in tonight catch much fish
“They say that Gianni caught a lot of fish tonight.”

(14) Ajo ko san se mu ko me dimua me interpretat [T]
3SG AUX.3SG say that 1SG.DAT AUX.3SG to help to interpret
“She said she would help me translate.”

However, when the verb of saying (san) is used to mean “to advise” or “to order”, the infinitive is used in the subordinate clause rather than the finite verb:

(15) Ivica iu ko san me vot në muzej
Ivica 3PL.DAT AUX.3SG say to go in museum
“Ivica advised them to go to the museum.”

4.6. WORD ORDER. Interestingly, it appears that Croatian and/or Italian word order have not significantly influenced AA. Although these languages shared some word-order features (e.g. they are all SVO, in all of them the relative clause precedes the nominal head, and the numerals and demonstratives precede the head nouns), where they do disagree AA preserved the original Albanian word order, e.g. with respect to the order of nouns and adjectives: While in Croatian the adjective generally precedes the noun (e.g. lijepa djevojka “beautiful girl”), in AA it follows the noun (Alb. vajza e bukur lit. “girl-the-beautiful”); note, however, that in this respect Albanian agrees with Italian word order, which may have helped its preservation in AA. AA also preserved the postposed Albanian articles (clitics) rather than developing preposed articles (as in Italian), or losing the articles altogether (Croatian does not have any articles).

5. CONCLUSION. In this paper we presented a brief history of language contacts between Arbanasi Albanian and Croatian and, to a lesser extent, Italian. We subsequently discussed a number of instances of contact-induced changes in Arbanasi Albanian, and we showed that the language was influenced by its neighboring adstrates (Croatian and Italian).

The exact conditions on the use of infinitives in Gheg dialects is difficult to ascertain, and the syntax of the Gheg dialect area from which the speakers of AA emigrated is virtually undescribed. It is worth noting that the use of the infinitive in Old Albanian (especially in Old Gheg texts in the 16th century) was more widespread than in the modern language (Demiraj 1989, Demiraj 1993).
and superstrates on all levels of linguistic structure. In our fieldwork, we focused on syntactic influences of Croatian and Italian on Arbanasi Albanian. They include the calquing of certain constructions, the development of particular analytic causative and imperative constructions, as well as the extension of the use of infinitive in subordinate clauses. This type of syntactic influence may be due to the pragmatic salience of constructions induced by communicative needs, a fact which is well attested in the literature on language contact (e.g. Thomason 2001, Grinevald Craig 1988, Matras 2012), and we believe that Arbanasi Albanian data contribute to such conclusions.

REFERENCES


Erber Tullio. 1883. La colonia albanese di borgo Erizzo presso Zara. Dubrovnik.


Jana Willer-Gold
jwgold@ffzg.hr

Tena Gnjarovics
tgnjavovics@ffzg.hr

Daniela Katunar
danielakatunar@gmail.com

Ranko Matasovic
rmatasovic@ffzg.hr
Reflections of an observant linguist regarding the orthography of *A Fala de Us Tres Lugaris*

Miroslav Valeš

*Technical University of Liberec*

A Fala has never had a standardized orthography as it is a language of oral tradition and almost all written documents have always been produced only in Spanish. The few documents which exist in A Fala use orthographies that vary considerably, especially when indicating the phonemes which are absent in standard Spanish. However, in the past decades there have been signs of an increasing interest regarding the language and cultural identity in the three villages and there have also been attempts to establish organizations to promote the language, such as *A Fala y Cultura*, *U Lagartu Verdi*, and *A Nosa Fala*. This increase in language awareness leads inevitably to situations, when the speakers want to express their linguistic identity in written form and the lack of written standard makes this task rather difficult. The objective of this paper is to analyze the public inscriptions, direction signs and street names written in A Fala. The appearance of these signs expresses the willingness of the speakers of A Fala to claim their linguistic identity. At the same time, their inconsistent orthography reveals the problems that arise in the course of writing their language. There are two main causes of these difficulties: The influence of Spanish, as all the speakers are bilingual in Spanish, and variation within the language itself. Regarding the first cause, the main issues include the uncertainty how to write the phonemes that do not exist in standard Spanish, and also whether the phonemes that do exist in Spanish should be written in the same way or not. In respect of the second cause, the signposts and street names reflect the three main varieties: Valverdeñu, Lagarteiru and Mañegu. They also partially reflect the ideas of those who created them and testify to a certain evolution in time. In general, the linguistic data in the form of street names and direction signs provide relevant information about the options for writing those phonemes which do not have an equivalent in Spanish, as well as geographical (diatopic) variation, and the changes of ideas regarding the orthography. This paper will use this valuable linguistic material to reflect on the issues that are involved in the establishment of an orthographical standard.

1. **INTRODUCTION.** *A Fala de Us Tres Lugaris* is a language spoken in three villages in Extremadura: Valverdi du Fresnu (Valverde del Fresno), As Ellas (Eljas), and Sa Martín de Trevellu (San Martín de Trevejo), in Sierra de Gata on the border between Spain and Portugal. According to Ethnologue there are about 5,500 speakers (Lewis et al. 2014) of the
language. The legal status changed in 2001 when A Fala was declared “bien de interés cultural”, yet, in the latest *Status of Autonomy of Extremadura*, A Fala was still not mentioned. Regarding its classification and origin, A Fala forms part of the Ibero-Romance subgroup of Romance languages. There are various theories affiliating A Fala with Portuguese, with Astur-Leonese and Galician. According to the most recent investigations carried out by Costas González (2011), Galician seems to be the closest “relative”. However, A Fala is clearly an independent language, not a dialect of some of those previously mentioned languages, and its origin will not be the topic of this paper.

The objective of this paper is to analyze the public inscriptions, road signs, and street names written in A Fala and, in general, to discuss the various ways of writing the language. Before I start with the specific examples and problems that we encounter when thinking about the orthography, I would like to mention the sociolinguistic situation of the three villages. First of all, each of the three places has its own variety and their differences are quite considerable. At the same time, however, the speakers are aware that they share the same language, as they sense a common identity. The largest of the three villages, Valverdi du Fresnu, has always served as a kind of local “metropolis”; it has had more contact with the outside world and its variety, Valverdeñu, has received more influence from Spanish. It also seems that in Valverdi the proportion of people who use A Fala in their daily routine is lower than that in the other two places. The variety of Sa Martín du Trevellu is called Mañegu. Sa Martín is a beautiful, picturesque village and its variety has some specific features not shared with the other two villages. The variety of As Ellas is called Lagarteiru. As Ellas seems to be a largely traditional place and most people, even children, use the language on a daily basis. The linguistic situation of the three places can be defined as diglossia, according to the definition of Ferguson (1959), where A Fala represents the low (L) variety while Spanish the high (H) variety. For the purpose of my observation it is relevant to stress that all or almost all the speakers of A Fala are bilingual. They have received their school education in Spanish, and this considerably influences their ideas on how to write the language.

A Fala is a language of oral tradition. It has never had any standardized orthography, and practically all written documents have always been produced only in Spanish. It is some kind of general belief that A Fala is for speaking while Spanish is for writing. In the past, speaking A Fala used to be a source of prejudice and its speakers were a target of ridicule. However, in the last decades the situation has been changing. Most people appear to be proud of their own specific language and identity. There have also been indications of a growing interest regarding the language and attempts to establish organizations to promote it, such as *A Fala y Cultura*, *U Lagartu Verdi* and, most recently, *A Nosa Fala*. This increasing awareness of the language inevitably leads to a desire for the speakers to express their linguistic identity in written form, yet the lack of a written standard makes this task quite difficult.

The first time I visited the three villages I was searching for some written documents in A Fala and was surprised that I could find hardly anything. I had been expecting to come across some posters advertising local celebrations, announcements of the local authorities but I was surprised that there was nothing, or almost nothing of this nature. The few documents I could find in A Fala were primarily linguistic descriptions and these used orthographies that varied considerably, especially when indicating the phonemes that do not exist in standard Spanish. These descriptions were also influenced by the theoretical point of view of their authors regarding the origin of the language. For this reason some tried to highlight the relationship with Portuguese, others, the relationship with Galician
Reflections of an observant linguist regarding the orthography of A Fala

2. Orthography Issues. When searching for A Fala in its written form I encountered it most frequently in the street names and direction signs. I believe that the appearance of these signs expresses the willingness of the speakers to claim their linguistic identity and to show the visitors and the rest of the world that A Fala exists. At the same time, however, the inconsistent orthography reveals the problems that arise in the course of writing the language. As I will try to demonstrate the difficulties have two basic causes: First, the transcription of phonemes that do not exist in standard Spanish, and second, the language variation itself. It should also be noted that there are both genuine difficulties, by which I mean problems for which there is no easy solution, as well as “artificially created” difficulties. I will describe both of them.

I will start with the most serious problem, which is one of the genuine ones. A Fala has four sibilant phonemes that do not exist in Spanish: The voiced alveolar fricative /z/, the voiceless postalveolar fricative /ʃ/ and the voiced postalveolar fricative /ʒ/ and also the voiced postalveolar affricate /dʒ/ sound. There are various options when it comes to representing these four sounds in written form.

The first mentioned, the fricative voiced alveolar /z/ is the voiced equivalent of the voiceless alveolar /s/. It is an “occasional phoneme” as there seems to be only one distinctive pair, but the sound is rather specific, and when people write in A Fala, they usually try to mark it in some way, to highlight that the sound is different from Spanish. It used to exist in Spanish as well, but here it lost its distinctive function sometime at the end of the 16th century. As it used to exist, it also had its written form, the most frequent of which was double <ss> for the voiceless /s/ and one <s> for the voiced /z/ in intervocalic position: passo /paso/ (“step”) – casa /kaza/ (“house”). On the street name signs we can find various possible options concerning how to reflect this sound. One of them is the “historical” spelling casa /kaza/ (“house”) with one <s>, the other is a combination of letters <sh> casha and the last one is an underlined <s> casa. Each of the three solutions has its advantages and disadvantages. The disadvantage of the first one, the historical solution, is that it would be necessary to write all intervocalic /s/ sounds as double <ss>, which is quite easy to consider but much more difficult to carry out. This is the solution that Frades Gaspar (2000) uses in his book Vamus a Falal, which is a respected manual of A Fala. However, there are many errors in the book itself which only goes to show that it is not easy to comply with this orthography. Another disadvantage is that the voiced alveolar /z/ sound written with one <s> “invites” to read it as the voiceless /s/ to all the people who are familiar with Spanish. The specific nature of the sound somehow disappears. Yet another problem is the fact that in Valverdeñu, for example, the word casa (“house”) is pronounced with voiceless /s/ and for this reason it would be rather confusing to write it with double <ss> while the other varieties would use one <s>. The <sh> solution ignores the fact that <sh> is frequently used to mark the sound /ʃ/ in other languages, however, it seems to have support of many community members. The positive aspect is that it expresses that there is a “different” sound. The solution with an underlined <s> also expresses well that this sound is different from standard Spanish. However, when it comes to writing it is rather bothersome to use letters that do not exist on a standard keyboard and you have to find them elsewhere. It would probably discourage the users and make them look for some
other solution. The most frequent solutions are the first two, the first one being promoted by Domingo Frades in his early publications and the second one by U Lagartu Verdi in their magazine Anduriña, a magazine written in A Fala.

The other two problematic sibilants – the voiceless postalveolar fricative /ʃ/ and the voiced postalveolar fricative /ʒ/ are usually written in the same way. Even though they are “occasional phonemes” (i.e. there is only one pair which is distinctive in one of the three varieties), it is sensible to write them in the same way, as there is considerable variation between the two. There is also the voiced postalveolar affricate /dz/ sound that seems to be a positional variant of /ʒ/ in Mañegu but a phoneme with only one distinctive pair in Lagarteiru. This sound is also usually written in the same way as the previous two. On the street names and direction signs it is possible to find three potential solutions concerning how to mark these sounds. One is <sh>, as in aishuntamentu (“city council”), the other is <s>, as in baixu (“small”), yet another is an underlined <x>, as in calexa (“alley”). The <sh> solution is sensible as the phoneme /ʃ/ is marked in this way in other languages. However, then it is not possible to use <sh> for the /ʒ/ phoneme and it is necessary to find other reasonable solution to mark this sound. The <s> solution is not a bad option either, but the same symbol marks the sound /ks/ as in examen (“exam”), /s/ as in xtraniuru (“foreigner”), and also /x/ as in México (“Mexico”). The one grapheme <s> thus comes to represent a number of different phonemes. Nevertheless, this solution enables us to use <sh> for /ʒ/. The underlined <x> is similarly problematic as the underlined <s>, it is not entirely “user friendly”. Out of the three, the most frequent solution is probably <s>.

On the other hand, there is not much consistency in the writing and authors often become confused when writing the sibilants due to the influence of Spanish, for example when they write: coisha (“thing”) with <sh> and ixenti (“people”) with <s>, even though they contain the same phoneme /ʒ/. The logical question regarding the sibilants is: What is the best solution? As mentioned before, sibilants are one of the real problems, which implies that every viable solution will nevertheless have some disadvantages.

Turning now to an example of a problem that is not a “real one”, we can have a look at the palatal lateral approximant /ʎ/. The same phoneme exists in Spanish, even though it sometimes appears in different words. It also exists in Portuguese. It is not certain that the Portuguese sound is exactly the same, but the Spanish and A Fala sounds coincide. Spanish uses double <ll> to mark this sound while Portuguese uses <lh>. On the signs, the Spanish solution was much more frequent, but I could also find the <lh> orthography. The latter tries to highlight the relation with Portuguese and at the same time it makes the written form different from Spanish. However, since all speakers of A Fala are bilingual and they were schooled in Spanish it seems rather illogical to use <lh> for the sound the people have always been accustomed to write with double <ll>. In my opinion the orthography should be as easy as possible to serve the users well. The <lh> orthography only reflects the opinions regarding the origin of A Fala, but it makes the writing more complicated for the users. A similar situation can sometimes be found with the nasal palatal phoneme /ŋ/. The Spanish orthography is <ñ> while the Portuguese is <nh>. Likewise, this leads to inconsistencies in representing the sound in written form. However, I could not find any <nh> in the street names and public inscriptions, only in older issues of Anduriña, as well as on some web pages.

The second cause of difficulties that I mentioned was the language variation. The variation can be of different kinds but first I would like to mention the diatopic or geographical variation. The three main varieties, Valverdeñu, Lagarteiru and Mañegu, were also reflected in the signs that were the subject of my study. For example, the definite article in Mañegu is
Reflections of an observant linguist regarding the orthography of A Fala

119

while it is /u/ in Lagarteiru, and Valverdeñu. The street names reflect this variation: Calli o Corchu in Mañegu, U Petril in Valverdeñu, Calexa do Portu in Mañegu and Calli du Portu in Lagarteiru. Similarly, there is considerable variation in the sibilants that were mentioned previously. We can also find variation on a lexical level, Centru médicu ("medical center") in Lagarteiru and Centru mécu in Mañegu or Calli Currieira in Lagarteiru and Calli Correeira in Mañegu. The diatopic variation that is reflected in the signs gives us an idea about the range of differences between the three varieties.

Observing the differences and considering that the three varieties are of the same status, we easily draw to a very important conclusion: It is not possible to have one standardized orthography for the three varieties. In general, it is possible to standardize the match “sound-symbol” so that the same sounds are written in the same way, but it is not possible to write the words in the same way in all the three varieties. The three varieties need to be written in different ways and it is a positive feature that the inscriptions mostly follow the local way of pronunciation.

Moreover, the variation in the writing and pronunciation of /l/ and /r/ is an example of diastatic/social variation. It seems to result more from the individual preferences and situation of speech rather than from belonging to one of the three varieties. We can find people pronouncing [pláða] and [práða] (“square”) in all three varieties. The <r> pronunciation or the one that is different from Spanish is sometimes considered more appropriate to A Fala and so it has certain “covert prestige”, while the <l> pronunciation or the Spanish-like is considered to be the influence of Spanish and it has the “overt prestige”. The variation works the other way round as well: árbol – álbuli (“tree”) where the Spanish-like form is always considered the influence of Spanish, while the other is considered more traditional. Nevertheless, most speakers use both pronunciations and the signs reflect both forms as well. We can find Plaza and Praza da Constitución. Other street names sometimes reflect the traditional and others the Spanish-like pronunciation: Calli Castelu Artu but Calli a Plaza. The street signs also reflect the changing opinions on the <l/r> writing. An example is Calli du Par/lqui, one sign with <r>, the other corrected for <l>, to distinguish it from Spanish.

Another topic regarding the orthography that can be observed from the street names is the use of <y> and <i>. These two symbols can represent the semi-consonant /j/. It is evident that at the moment of writing there are no linguistic reasons for using any particular one of the two. The only criterion is the similarity to Spanish or differentiation from Spanish. In the inscriptions we can find both solutions, the Spanish-like but also the attempts to write in a different way. For example, Calli 1er de Maio prefers the differentiation (it is to be observed that in Valverdeñu, it should in fact be 1er de Maiu, with the final <u>). However, we can find both solutions with ayuntamentu (“city council”). In general, the use of <i> instead of <y> is more frequent. It seems that the people who write in A Fala like the idea of making it look different to express their particular identity.

The apostrophe is another feature of A Fala orthography. In general, the apostrophe sometimes tries to indicate that the author of the sign considers something to be “missing”. However, the use of apostrophes is rather chaotic and thus we can find the sign Casha d’a cultura (“house of culture”) right next to the building where we can read the same word without an apostrophe. In another sign the author was so confused by the apostrophes of Bar d’us Jubilaus (< de us) and Parqui d’u Castelu (< de u) that he or she wrote Centru d’e Día (“day care center”) where we would expect simple de as nothing really disappears. In another sign there occurs Camiñu d’u Portu, to be compared with Calli du Portu without an apostrophe. However, most signs and street names avoid apostrophes altogether. By
contrast we can find apostrophes frequently in the few written documents (older issues of Anduriña and some web pages). In general, the apostrophes make the writing rather complicated and it is very difficult to be consistent and always use them in the same way.

3. CONCLUSION. To summarize, the linguistic data in the form of street names and direction signs reflect most of the general problems of A Fala orthography. These problems are caused primarily by two factors: The use of phonemes that do not exist in Spanish and for that reason they do not have any habitual written form and, second, the variation inside the language, whether diatopic or diastratic. At the same time, the signs also bring relevant information about possible solutions concerning how to write the phonemes that do not have a parallel in Spanish and they also give us information about the variation of the language. A Fala has survived over centuries in oral form, without being written, but it is not certain that it can survive further centuries in the same way. The written form would definitely help to promote the language and it would make it stronger with more possibilities to survive. It has not been the objective of this paper to give the guidelines how to establish a standardized orthography. Nevertheless, when considering the orthography of A Fala, it is important to concentrate on the genuine problems. It is also necessary to realize that it is not possible to unify the three varieties. For this reason each of them will need its own written form, slightly different from the other two. The most important consideration is that the final orthographical standard should make the writing as easy as possible because only “user friendly” orthography will motivate the speakers to write in their own language.

REFERENCES


Miroslav Valeš
miroslav.vales@tul.cz
Language Landscape: Supporting community-led language documentation

Sandy Ritchie, Samantha Goodchild, and Ebany Dohle

Language Landscape/SOAS

Different groups have differing motivations for participating in language documentation projects. Linguists want to increase our knowledge of languages and linguistic theory, but constraints on their work may lead to issues with their documentation projects, including their representations of the languages they study. Native speakers participate to maintain and develop their language, and may choose to represent it in a way which showcases their culture and attitudes. In order to encourage more native speakers to take part in documentation projects, a simple integrated system is required which will enable them to record, annotate and publish recordings. Language Landscape, our web-based application, enables native speakers to publish their recordings, and Aikuma, a mobile application for documentation, enables them to record and orally translate recordings, in both cases with minimal cost and training required. Language Landscape benefits communities by allowing them to document their language as they see fit, as demonstrated by our outreach program, through which some London school children created their own projects to document their own languages and those spoken around them.

1. INTRODUCTION. Linguists and native speakers participate in language documentation projects for different reasons. “. . . [T]he linguists’ research aim is to contribute to our scientific knowledge of the world’s languages or to linguistic theory, while the local language workers’ aim is to do something for the maintenance and development of their language and culture” (Mosel 2006: 67). These differing motivations may lead to differences in the way that the two groups choose to represent languages. As Dobrin et al. (2007) opine, “. . . linguists’ efforts to preserve [endangered] languages seem to lead inexorably to their reduction . . . to indices, objects, and technical encodings . . . ” (2007: 60). Native speakers may prefer to represent their language in a way which showcases the culture it encodes and expresses the community’s collective view of themselves and others in their society. So far, documentation projects initiated and led by native speakers outside of the academic sphere have remained largely hypothetical. Two reasons for this are lack of resources and lack of technical knowledge. Documentation projects often rely on costly and complex technology and are carried out by students and professionals who have undergone extensive training. These requirements are barriers to entry for native

---

1 We would like to thank the students and staff of Bow School, east London for working with us through our outreach program, and Teresa Poeta for her comments on an earlier draft of this paper. All remaining errors are our own.
speakers who might otherwise be motivated to participate in and even initiate documentation projects. Removing these barriers is desirable as native speakers can then exercise greater control over the way their language and culture is represented in the public sphere. In order to achieve this, a system is required which enables people to create, annotate, and publish recordings of their language with minimal training and with the technology that is available to them. Given the availability of cheap internet-enabled mobile phones and high speed mobile telecommunication networks (GLA 2014) in an increasing number of countries, it seems that a mobile application which incorporates all of the capabilities listed above would be the best format for such a system.

This paper sets out to show how Language Landscape (http://languagelandscape.org/), a web-based application developed by our team, may meet some of the requirements of such a system. The website enables users to publish their recordings on a global map of languages. The paper also discusses a mobile application for documentation currently in development: Aikuma (http://www.aikuma.org/) is an Android application developed by researchers at the University of Melbourne which enables users to create recordings and provide time-aligned oral translations of them. Both applications may be viewed as attempts to engender greater participation in documentation projects by employing simple interfaces which require minimal training and cost to be able to use. Language Landscape offers simple forms which make publishing a recording or project online quick and easy, and collaborative features which enable people to work together on documentation projects. Aikuma employs an innovative “phone call” interaction style whereby the speaker listens to a recording as if they were on the phone and can interrupt it at any time to add a segment of oral translation simply by speaking into the phone (Hanke & Bird 2013).

Each of these applications goes some way to lowering the barriers to entry in language documentation projects. However, neither currently provides a complete integrated system for people equipped only with internet-connected mobile phones to create and publish a documentation project. This paper speculates on how the development of an accompanying mobile application for Language Landscape would bring our system closer to this ideal.

The paper also briefly discusses issues with current documentation practices and contrasts them with how documentation may be done using software like Language Landscape and Aikuma. It goes on to highlight the benefits to language communities and individuals offered by Language Landscape. An evaluation of a pilot outreach programme carried out in a London school is also included as a case study of the benefits of community-led documentation.

2. Issues in Language Documentation Practices. Current language documentation practices have led to the creation of a previously unparalleled range of materials relating to minority and endangered languages. The scope and detail of these materials is due in large part to the efforts of a small community of highly specialized researchers. These linguist-documenters have spent years collecting, annotating, and archiving bodies of material which often come close to Himmelmann’s “comprehensive record of the linguistic practices characteristic of a given speech community” (1998: 166). However, linguist-led documentation projects are not without their problems. There is often a large cultural gap between the linguist and the community, leading to a lack of contextual understanding and a potential for misrepresentation.

---


3 Language Landscape team. Directors: Ebany Dohle, Samantha Goodchild, Karolina Grzech, Charlotte Hemnings, Teresa Poeta, Sandy Ritchie (all SOAS); Main Developer: Graham Ritchie (The University of Edinburgh); Designer: Nick Ritchie (BBC Radio).
between linguist and community, leading to problems with recording certain cultural and linguistic practices. Linguists are often working on tight budgets and timescales which limit the potential of their documentation project. Many linguists are also engaged in teaching and research activities which take them away from their fieldwork.

One solution to these problems is to carry out documentation work with the community. Grinevald advises “fieldwork with teaching, training, and mentoring native speakers for sustainable documentation projects” (2003: 60). Typically, the linguist-documenter trains community members in the techniques of language documentation both to collect data which would normally be inaccessible to the researcher and in the hope that they will continue the work of documenting the language beyond the scope of the funded research project. This is an admirable endeavor and there have been some success stories, for example Grinevald’s own work with the Rana community in Nicaragua and McGill’s work with the Cicipu community in Nigeria (McGill 2009).

However, for many projects, training of community members in the use of recording equipment and software packages is impracticable due to lack of time and resources. The requirement of basic computer literacy to complete tasks such as transcription and translation work is also sometimes an issue as some communities have had little or no experience with computers until the arrival of the researcher.

3. THE ROLE OF TECHNOLOGY IN COMMUNITY-LED DOCUMENTATION. In the face of these issues, truly community-led language documentation projects might seem like a distant prospect. However, with the increasing availability of cheap mobile technology, it is possible to envisage a future in which many more people own or have access to tools which could readily be harnessed for language documentation purposes. In a community with several internet-connected mobile phones installed with a dedicated language documentation application, making, annotating, and publishing recordings could be carried out by a much wider range of people. In such a situation, the possibilities for involvement in or even initiation of a documentation project by the community seem much more likely.

3.1. CROWDSOURCING TRANSLATION: AIKUMA. A mobile application which includes some of these features is already available: Aikuma uses an innovative user interface which enables users to create recordings and provide time-aligned oral translations of them.

The developers’ motivation for creating the app is to remove the bottleneck in current typical language documentation workflows in creating transcriptions and translations of recordings (Hanke & Bird 2013). Trials of the app have been successful and the developers plan a series of improvements which will make the app even more practical for use in language documentation projects.5

Aikuma is an excellent example of how to harness the capabilities of low-end mobile phones6 to broaden participation in a language documentation project. By removing the requirement of an established writing system and/or literacy, a much wider range of people become potential collaborators in a project. Even in communities which do have high rates of literacy, using Aikuma negates the need to provide extensive training in recording techniques and the use of computer software packages. This frees the linguist-documenter of a considerable commitment and allows them to involve a wider range of people in the documentation project.

4 See also http://cicipu.org/ (10 December, 2015).
6 The researchers used inexpensive Android mobile devices in their trials.
Aikuma demonstrates that technology can be an inclusive force in language documentation projects. However, the typical documentation workflow does not end at transcribing and translating recordings. Beyond that, recordings are tagged with metadata, annotated using dedicated software such as Toolbox/FLEx and ELAN and eventually deposited in an archive. Respecting access restrictions, the archive will then publish the collection online, making it accessible to anyone with an internet connection.

3.2. CROWDSOURCING PUBLICATION: LANGUAGE LANDSCAPE. Where Aikuma is designed for crowdsourcing of the early stages of the documentation workflow – that of making and translating recordings, the Language Landscape website is designed in part for crowdsourcing of the final stage – publishing recordings online. We will return to this point but before that we will introduce the project and describe some features of the website.

Language Landscape is a not-for-profit organization based in London, UK\(^8\) which aims to raise awareness of language diversity. We run a website and outreach program to bring together language communities online, to help people to better understand the languages spoken around them, and to help to raise the profile of minority and endangered languages. Our website enables anyone with an internet connection to add language recordings to a world map. The recordings are mapped where they were made in order to reflect the geographical spread of languages. Other users can access the recordings by browsing or searching the map. In order to encourage more people to contribute to the project, we have initiated outreach work with several communities and completed a pilot educational program with a school in east London. The program provided students with practical training in recording techniques and helped them to learn about issues such as multilingualism and language endangerment, using the website as a starting point for discussion and activities (see Section 5).

The website is based around a series of maps which display where recordings added to the site were made.\(^9\) Each recording is represented by a marker which users can click on to see information about the recording and play it in page. The homepage displays all the recordings currently available on the site.

---

\(^7\) Image source: http://www.aikuma.org/getting-started.html (10 December, 2015)

\(^8\) Language Landscape is a company limited by guarantee registered in England and Wales (8694275).

\(^9\) We use the Google Maps API and an in page JavaScript media player to play the recordings. The website runs on a Python web application server with recordings and associated metadata is tracked in a relational database.
Figure 2: The Language Landscape homepage with an in-page recording being played

As well as this primary map displaying all of the recordings currently available, the website also features potentially limitless “submaps” which show subsets of the recordings. For example, each language represented on the site has its own submap which displays only the recordings in which someone is speaking that language (see Figure 3). These pages also offer other information about the language, including a description and its position in the genealogical tree.\(^\text{10}\)

If a user is looking for something specific, they can also define their own submap by using the Advanced Search functions (http://languagelandscape.org/search/recordings/). These allow users to conduct complex queries on the database with the results displayed as a custom submap (see Figure 4).

For reasons of space we have only outlined a few of the website’s features here. For more information and help on using the features of the site, see our Help pages (http://languagelandscape.org/help/).

Now that the website has been explained in more detail, we can return to the point of how it may help to encourage greater participation in the publication of language documentation materials. The central idea driving the design and implementation of the website was

\(^{10}\) Genealogical trees are primarily taken from Glottolog – http://glottolog.org/glottolog (10 December 2015) – as well as other sources.
Figure 3: A Language Landscape submap: recordings of Polish

Figure 4: A custom submap: Recordings of either French or Polish made after 24 January 2012. The filters on the left show which recordings match which search term.
to make this process quick and easy for anyone with an internet connection and a Language Landscape account. To add a recording, contributors upload it to the website and fill out a simple form. Contributors can also map existing content from YouTube on Language Landscape by pasting the URL in the relevant field instead of uploading a file. Once the recording has been added to the database, it is flagged for moderation and published after being checked by a moderator.

![Language Landscape Add Recording form](image)

**FIGURE 5: Language Landscape Add Recording form**

If the contributor agrees to it, the recording’s metadata can also be edited by other users. This allows people to improve on existing recordings by adding transcriptions, translations, and other information which the original contributor may not have provided.

As well as adding individual recordings to the map, contributors can also create their own “projects” on the website. A Language Landscape project is a user-defined submap featuring a collection of recordings grouped together around a particular theme. Projects are designed to be collaborative: several users can work together on a project by contributing recordings and metadata and personalizing their project page with our customization tools. The projects each have their own page with options to add a description and video.

---

11 You can sign up for an account by registering your details here: [http://languagelandscape.org/accounts/login/](http://languagelandscape.org/accounts/login/).

12 The full list of metadata categories in this form is available here: [http://languagelandscape.org/help/add-recording/](http://languagelandscape.org/help/add-recording/).
Users can also customize the style of their project map using the Google Styled Maps Wizard\(^\text{13}\) and add a custom marker by uploading their own marker image. The form for adding a project is also designed to be quick and easy to fill in.

![Figure 6: A customized Language Landscape project page (http://languagelandscape.org/project/WOLP)](http://languagelandscape.org/project/WOLP)

We believe that by offering these intuitive content creation features for free online, we will help to encourage more people to get involved in documentation projects. The system would also be greatly improved by the addition of an accompanying mobile application à la Aikuma. Such an app would integrate the ability to make a recording using a phone with the tagging and publication tools we have already developed. This would make the entire process of making, tagging, and publishing recordings online even simpler and more streamlined. Given the global trend towards smart phones as the primary communication tool, development of such an application is an important next step for the project.\(^\text{14}\)

This section described how software like Aikuma and Language Landscape can encourage greater participation in documentation projects by simplifying the processes in-


\(^{14}\) If you are interested in supporting the development of Language Landscape, please get in touch with us: admin@languagelandscape.org
volved. The next section will discuss why using Language Landscape in particular may benefit communities and individuals.

4. Benefits to communities and individuals. The primary advantage of Language Landscape is that it enables speakers to represent their languages and cultures online as they see fit. Representing one’s own languages offers several possible benefits to communities and individuals, in particular to those who speak lesser-known and endangered language varieties.

4.1. Naming the language. Firstly, it allows people to decide on their own name for their language variety. A linguist who is not a native speaker of the language he/she studies may not be aware of subtle dialectal differences and may represent many varieties as one and the same language. Native speakers make these distinctions intuitively, and the freedom to name their variety as they wish allows them to represent these differences. Other speakers who also speak the same variety can then contribute to the new submap, giving the variety an identity and a “homepage” online. This also extends to smaller and perhaps less prestigious varieties of larger languages. Some varieties of British English, to take an example close to the authors’ home, are popularly thought of as aberrations of the “standard” variety or indications of low educational attainment (e.g. Watts & Trudgill 2002). Enabling speakers to showcase their unique variety by putting it on the map may lead to greater respect and understanding of the diversity present within large languages as well.

Allowing user-defined language names is potentially problematic as in some cases these names will not match formal terminology used in linguistics and related disciplines. We intend to work around this by retaining the user-defined name for a variety but also adding alternative names in the metadata, so for example colloquial names for varieties such as Jafaican will also be tagged with more formal designators like Multicultural London English. This will allow for the use of standard terminology while retaining the ability for native speakers to name their variety as they see fit.

4.2. Genres and registers. As with language varieties, native speakers are also best placed to document the diversity of registers and genres present within a speech community. Young people in particular are often at the cutting edge of new forms of expression as well as being au fait with the speech of their parents’ and grandparents’ generations. Encouraging young people to record these genres and collaborate on innovative projects to document and map them will help them to appreciate the expressive power of their own language in comparison with dominant global cultural forms.

4.3. Code-switching, code-mixing and multilingualism. For many communities, using several languages in the course of their daily lives is the norm (see e.g. Lüpke & Storch 2013 for an overview of the situation in Africa). A single language map will not be sufficient to represent these communities’ linguistic practices and repertoires. Offering multilingual communities the ability to tag their recordings with multiple languages and represent their practices as project submaps rather than language submaps will enable them to document the complex and dynamic nature of their linguistic situations.
4.4. **COLLABORATION.** The ability to collaborate is key to encouraging more people to get involved in a project. Collaboration allows communities to represent their language more democratically: the work of making and translating recordings is distributed amongst many people, and everyone can have a say in which recordings get selected for publication. A particular advantage of using an online platform is that it also enables communities who are separated geographically but share linguistic and cultural ties to collaborate remotely. This will also help to maintain links or establish new connections between diaspora communities and the community back home.

We will now turn to the benefits of our outreach work, focusing on a case study of an outreach project with a group of mainly Sylheti-speaking school children living in east London.

5. **LANGUAGE LANDSCAPE OUTREACH.** In 2013 Language Landscape carried out a pilot outreach project with Bow School, in East London, with a view to hosting regular outreach activities with various communities in the city. In this section we will look at our aims in our outreach programs, including increasing participation, before focusing on the Bow School case study.

5.1. **INCREASING PARTICIPATION.** Language Landscape can be used to document any variety of language anywhere in the world. In order to encourage greater use of the website, we feel it important to focus our efforts on individuals and communities who are most likely to become active users. We carried out our pilot study with a school in London (see below section 5.3) and will be working in the future with other minority and endangered language communities. We hope to encourage further uptake of our outreach projects which aim to increase community- and individual-led documentation efforts through providing basic training on using readily available technologies and a background to linguistic issues. By promoting our website and outreach program, we hope to engage with a greater number of speakers and communities, fostering greater interactivity among users of the website and creating an enthusiasm for undertaking community-led documentation projects.

5.2. **GENERAL AIMS.** The outreach projects carried out by Language Landscape are flexible in nature and can be adapted to suit the individual needs of different communities. There are, however, core themes which form the basis of our outreach projects. As all current Language Landscape directors and volunteers have been trained in documentary linguistics, we are able to offer classes and workshops in the basics of sound recording and editing, both using professional equipment, and also how to achieve good results using the technology available to most people. In addition to providing a solid grounding in technology training, the outreach projects also give a broad overview of linguistic issues to present a better understanding of the multilingual world which we live in. Some of the areas which we cover in the outreach project are: an introduction to the discipline of linguistics, why people study language(s), minority and endangered languages, and the benefits of multilingualism. The outreach projects provide speakers and communities with the skills and knowledge required to create their own projects to document their language varieties in whichever way they wish, as demonstrated in our pilot outreach program with Bow School below.

---

15 If you are interested in running an outreach project with Language Landscape, please contact our outreach team at outreach@languagelandscape.org
5.3. CASE STUDY: BOW SCHOOL. Language Landscape carried out a pilot outreach project with Bow School in east London in spring 2013. The pupils created their own projects which are available to view on the website. The outreach project consisted of an eight-week series of workshops which culminated in a visit to SOAS, University of London, where the pupils took part in a debate on endangered languages and presented their projects to students, staff, and the general public.

Bow School is set in an area of East London, the Borough of Tower Hamlets, where over 70% of schoolchildren have English as a second language. In this particular school, the outreach team worked with a class of pupils between the ages of 12–13, where the majority of children also speak Bengali and/or Sylheti. Throughout the eight weeks the team devised a number of interactive activities to introduce the children to topics such as minority languages, multilingualism, and endangered languages. They were also taught the basics of sound recording and were able to try it out on semi-professional equipment provided by the outreach team. The main focus, however, was on applying the recording techniques for use on their own mobile/smart phones to carry out their own projects. The pupils were encouraged to come up with their own topics for their projects and many of the topics chosen were inspired by issues covered in the workshops and by the children’s own experiences. For example, one group surveyed teachers in their school and recorded them speaking various languages, whilst another group chose to focus their project on languages and scripts used for writing, interviewing teachers who can write in the Bengali script. As is evident, when free to choose their own topics, the pupils created projects which were interesting to themselves, projects that reflect their community.

The recordings and projects created by the pupils in the case study could certainly be of interest to linguists and researchers investigating multilingual repertoires, certain types of speech genres and adolescent speech. Above all, though, the recordings are of definite interest to the speakers themselves, and by extension, their communities. We believe that this project shows some of the hallmarks of a community-led documentation project, which forms the basis of Language Landscape’s model of documentation, which may provide an alternative or accompaniment to linguist-led documentation projects.

6. CONCLUSION. As demonstrated by the case study, simple, intuitive tools such as Language Landscape do enable communities and individuals to collaborate on documentation projects which represent their language and culture in the way they see fit. The publication features offered will also be greatly improved in future by the addition of a mobile application which enables users to add recordings directly from their mobile phone. The website offers many benefits to native speakers of smaller languages, not least the opportunity to take part in a global project which represents smaller language varieties on an even footing with larger languages. The website may also be a driving force for the establishment of new online communities, “bring[ing] together speakers and learners who are scattered over great distances” (Holton 2011: 371).

In addition to our online platform, the outreach projects provide an opportunity for some communities to receive face-to-face training in linguistics and technical skills. Although these are not required to create an interesting project on the website, our outreach work allows us to engage with communities directly, and we benefit from this by incorporating their suggestions and ideas into the website and outreach program. Gener-
ally, our aim is to raise awareness of language diversity and the many issues surrounding it by making language documentation projects open to a wider and more diverse range of people, thereby supplementing current language documentation practices with community- and individual-led documentation projects.

REFERENCES


Sandy Ritchie
sandyfritchie@gmail.com

Samantha Goodchild
299225@soas.ac.uk

Ebany Dohle
ebdohle@googlemail.com
BaTelÒc: A Text Base for the Occitan Language

Myriam Bras and Marianne Vergez-Couret

CLLE-ERSS, UMR 5263 Université Toulouse Jean Jaurès & CNRS

Language Documentation, as defined by Himmelmann (2006), aims at compiling and preserving linguistic data for studies in linguistics, literature, history, ethnology, sociology. This initiative is vital for endangered languages such as Occitan, a romance language spoken in southern France and in several valleys of Spain and Italy. The documentation of a language concerns all its modalities, covering spoken and written language, various registers and so on. Nowadays, Occitan documentation mostly consists of data from linguistic atlases, virtual libraries from the modern to the contemporary period, and text bases for the Middle Ages. BaTelÒc is a text base for modern and contemporary periods. With the aim of creating a wide coverage of text collections, BaTelÒc gathers not only written literary texts (prose, drama and poetry) but also other genres such as technical texts and newspapers. Enough material is already available to foresee a text base of hundreds of millions of words. BaTelÒc not only aims at documenting Occitan, it is also designed to provide tools to explore texts (different criteria for corpus selection, concordance tools and more complex enquiries with regular expressions). As for linguistic analysis, the second step is to enrich the corpora with annotations. Natural Language Processing of endangered languages such as Occitan is very challenging. It is not possible to transpose existing models for resource-rich languages directly, partly because of the spelling, dialectal variations, and lack of standardization. With BaTelÒc we aim at providing corpora and lexicons for the development of basic natural language processing tools, namely OCR and a Part-of-Speech tagger based on tools initially designed for machine translation and which take variation into account.

1. INTRODUCTION. Occitan is a Romance language, spoken in southern France and in several valleys of Spain and Italy. The number of speakers is hard to estimate: According to several studies, it can be evaluated between 600,000 to 2,000,000 (Martel 2007, Sibille 2010). Occitan is not a unitary language, it has several varieties. The most accepted classification of Occitan dialects was suggested by Bec (1995) and includes Auvernhàs,Gascon, Lengadocian, Lemosin, Provençau, and Vivaro-aupenc.

Occitan is not standardized as a whole. Nevertheless, it is written since the Middle Ages and has a very important literary tradition. Its literature has been translated to other languages (Mistral, Boudou, Rouquette, Manciet, etc.). Although much less socialized than it was before the Second World War, Occitan is now present in newspapers, on the internet, on

This project was carried out with the help of the Région Midi-Pyrénées, which has funded, together with the University of Toulouse-Le Mirail, a two-year post-doctoral fellowship devoted to BaTelÒc.
the radio and television, and in some public schools and universities. Non-governmental organizations maintain and spread Occitan: the Felibrige, the Institut d’Estudis Occitans, the associative network of immersive schools Calandreta, the linguistic training institute for adults – Centre de Formacion Professionala Occitan. However, Occitan has no official status in France.

In this paper, we present a text base for the Occitan language, called BaTelÒc. Nowadays, Occitan documentation mostly consists in dialectological data (several regional linguistic atlases are gathered in the THESOC database searchable online\(^2\)), digitized lexicographic data (few bilingual dictionaries are searchable online\(^3\)), virtual libraries (books in PDF format) from the modern (Bibliotheca Tholosana Occitana,\(^4\) 16\(^{th}\)–18\(^{th}\)) to the contemporary period (CIEL d’Òc,\(^5\) 19\(^{th}\)–21\(^{st}\)), and machine-readable texts of the Middle Ages (Concordance of Medieval Occitan (Ricketts et al. 2001) and Linguistic Corpus of Old Gascon (Field 2013)). In addition, the CIRDOC (International Center for Occitan Documentation) is developing a multimedia library, Occitanica\(^6\) which offers access to a multiplicity of sources: written texts, images, virtual exhibitions, documentary films, sound records, etc. The BaTelÒc project aims at complementing those resources with machine-readable texts for modern and contemporary periods (see Bras 2006, Bras & Thomas 2011 for a description of the text base experimental version). It aims at developing a wide range of text collections by gathering written literary texts (prose, drama and poetry) and others genres such as technical texts and newspapers, and also by embracing dialectal and spelling variations.

This resource will be relevant for the linguistic description of Occitan in lexicology, morphology, syntax, semantics, and discourse studies. It could also provide data for lexicographic projects (Bras & Thomas 2007) and for studies in literature, anthropology, ethnology, history, etc. And last but not least, the text base is also meant for teachers and new speakers by giving real language uses in learning competencies.

In this paper, we first discuss the aims of BaTelÒc (Section 2): As Occitan is an endangered language, building a data base of Occitan texts can be regarded as a language documentation process; at the same time, as there is a big amount of written texts, the text base can be built in a corpus linguistic perspective. We then present the first version of BaTelÒc (Section 3) and end with some perspectives (Section 4).

2. Text Bases within the fields of Language Documentation and Corpus Linguistics.

2.1. Language Documentation and Corpus Linguistics. Language Documentation, as defined by Himmelmann (2006), aims at compiling and preserving linguistic data for studies in linguistics, literature, history, ethnology, sociology. It should include all language modalities, covering spoken and written language, various registers, for the study of the language as social practice and cognitive faculty. The special situation of endangered languages, namely the unsustainable small and declining speaker base, influences greatly the objectives: to collect on an opportunistic basis all the possible varieties, first and foremost spoken data (because if writings remain, spoken words fly away!). Language

\(^3\) http://www.locongres.org (10 December, 2015).
\(^6\) http://occitanica.eu (10 December, 2015).
Documentation offers well-documented material which might serve for linguistic analysis. From this material, linguists can extract a coherent corpus for their own specific studies (Cox 2011).

Corpus Linguistics has been mostly developed in response to the problems caused by introspective methods. It consists in studying languages based on attested examples. It mostly concerns well-studied languages and is based on specific goals of linguistic studies. This generally requires the development of big corpora (millions of words), “sampled texts, written or oral, in machine readable form” (McEnery et al. 2006: 4, cited by Mosel 2013), to be representative of the language variety under study.

2.2. TEXT BASES. Text bases can be seen as the result (an online production) of both language documentation and corpus linguistics projects. Data in text bases must then consist of marked-up, organized, and well-documented machine-readable authentic texts. Those data can therefore be gathered as corpus. Text bases are available online with a range of tools that allow the user to select his/her own corpus and interrogate it (concordancer, frequencies, ...). The efforts to create text bases range from well-studied languages such as English, French, Portuguese, etc., to vulnerable languages\(^7\) such as Basque or even endangered languages such as Picard.

The British National Corpus (BNC) is probably the most famous general corpus. It has been created by the industrial/academic BNC Consortium (led by Oxford University Press and composed of the following publishers: Addison-Wesley Longman and Larousse Kingfisher Chambers, and academic research centers: Oxford University Computing Services (OUCS), the University Centre for Computer Corpus Research on Language (UCREL) at Lancaster University, and the British Library’s Research and Innovation Centre). The corpus is designed to represent a wide range of modern British English, from written (newspapers, letters and memoranda, school and university essays) to spoken data (everyday conversations, formal business conversations, radio material).\(^8\) Other corpora have been designed as the BNC to allow comparisons across genre and provide a reliable basis for contrastive language studies, namely the American National Corpus, the Korean National Corpus, and the Polish National Corpus. There are several ways to access BNC online, for instance through the access provided by the Brigham Young University (the BYC-BNC).\(^9\) The Brigham Young University also provides other corpora with the same exploring tools (see Table 1 below), such as for example the Corpus of Contemporary American English.\(^10\)

Another well-known text base is The Bank of English (BoE), extracted from the 2.5-billion-word Collins Corpus. It is a balanced corpus of current English regarding genres and diatopic variation. It includes a wide range of written (websites, newspapers, magazines and books) and spoken data (everyday conversations, radio and TV material). The text base also considers diatopic variation including material mainly from UK, US, Australia, and Canada but also from India, New Zealand, South Africa, and Ireland. Moreover, BoE is constantly updated with new material. This base was designed to support the development of COBUILD dictionary by giving real-life examples, collocations, and frequencies.

As for French, Frantext\(^11\) was designed to support the creation of Trésor de la langue française. Frantext has been developed by ATILF laboratory (Analyse et Traitement

\(^7\) The classification chosen here is the one established by UNESCO.
\(^8\) [http://www.natcorp.ox.ac.uk](http://www.natcorp.ox.ac.uk) (10 December, 2015).
Informatique de la Langue Française). It is a reference text base in French linguistics and literature. It includes a collection of nearly 4,800 texts comprising 286 million words (literary, scientific, and technical) from the 12th to the 21st century. Unlike the two corpora described above, Frantext is not a balanced corpus. Almost all the text bases offer the possibility to select texts to work on. But in the case of Frantext, this step is achieved separately. Exploring the text base is then done in two steps: first building a corpus and then searching for concordances. The main advantage is that once the corpus is defined, it may be used for the entire work session.

Considering Catalan, an example of a more recent official language, the Corpus Textual Informatitzat de la Llengua Catalana\textsuperscript{12} was created mostly to support the development of a lexicographic project (Diccionari del Català Contemporani) by the Institut d’Estudis Catalans. It only includes written texts (for instance literary, newspaper, technical, and scientific materials).

However, not all text bases are designed for lexicographic purposes and many other rich-resourced languages have their text bases online as for instance the Reference Corpus of Contemporary Portuguese (CRPC)\textsuperscript{13} developed by the Centro de Linguística da Universidade de Lisboa. It contains written (literary, newspaper, technical, scientific, didactic, etc.) and spoken material (formal and informal conversation) and considers diatopic variation as well (Europe, Brazil, Angola, Cape Verde, Guinea-Bissau, Mozambique, São Tomé and Príncipe, Goa, Macau, and East-Timor). Only the written part of the corpus is available online.

Text bases are being also created for vulnerable languages such as Basque and endangered languages such as Picard. The XX. Mendeko Euskararen Corpus Estatistikoa has been developed by the Real Academia de la Lengua Vasca. It includes 6,351 texts organized by historical period, dialects (Bizkaiera, Gipuzkera, Zuberner, Lapuertera-Nafarrera, Unified Basque) and genre (administrative, literary, scientific, newspaper writing). Picartext\textsuperscript{14} has been developed by CERCLL-LESCLaP (Laboratoire d’Études Sociolinguistiques sur les Contacts de Langues et la Politique Linguistique, EA 4283). It includes 300 written texts organized by genre (literature, letters, songs, comics, and dictionary) and diatopic features (by reference to the place of the authors – different French and Belgian districts).

As far as technical aspects of text bases are concerned, the Text Encoding Initiative seems to be chosen by all bases as a means to provide and classify texts. The different bases also provide tools for analyzing languages and/or help lexicographic work, for instance collocation, frequency, and concordance checks to observe different word usages. Most of them are enhanced with linguistic annotations, Part-Of-Speech tagging (POS), and lemmatization when relevant. Text bases of under-resourced languages (as for example Basque and Picard) normally do not provide such kind of annotation. Texts and tools are made available online to a wide range of users: academic researchers, general public, learners, new speakers, etc. Table 1 offers a synthesis for all the text bases presented above.

2.3. BUILDING A TEXT BASE FOR THE OCCITAN LANGUAGE. We are developing a text base for Occitan, called BaTelÒc. As Occitan is an endangered language, building a data base of Occitan texts can be regarded as a language documentation process: Our

\textsuperscript{12}http://ctilc.iec.cat (10 December, 2015).


\textsuperscript{14}http://www.u-picardie.fr/LESCLaP/PICARTEXT/Public/index.php (10 December, 2015).
<table>
<thead>
<tr>
<th>Text Bases</th>
<th>BYC-BNC</th>
<th>BoE</th>
<th>Frantext</th>
<th>CTILC</th>
<th>CRPC</th>
<th>XX MECE</th>
<th>Picartext</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>English</td>
<td>English</td>
<td>French</td>
<td>Catalan</td>
<td>Portuguese</td>
<td>Basque</td>
<td>Picard</td>
</tr>
<tr>
<td>Language Status</td>
<td>Official language</td>
<td>Official language</td>
<td>Official language</td>
<td>Recent official language</td>
<td>Official language</td>
<td>No official status at country level</td>
<td>No official status</td>
</tr>
<tr>
<td>UNESCO Atlas of the World's Languages in Danger</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Vulnerable language</td>
<td>Endangered language</td>
<td></td>
</tr>
<tr>
<td>Corpus Type</td>
<td>Balanced corpus</td>
<td>Balanced corpus</td>
<td>Collection of texts</td>
<td>Balanced corpus</td>
<td>General corpus</td>
<td>Diversified corpus</td>
<td>Diversified corpus</td>
</tr>
<tr>
<td>Content</td>
<td>Written and oral British English</td>
<td>Written and oral Diatopic variations</td>
<td>Written</td>
<td>Written</td>
<td>Written and Oral Diatopic variations</td>
<td>Written</td>
<td>Written</td>
</tr>
<tr>
<td>No. words</td>
<td>100 million</td>
<td>650 million</td>
<td>270 million</td>
<td>50 million</td>
<td>300 million</td>
<td>4.6 million</td>
<td>8 million</td>
</tr>
<tr>
<td>Define your own corpus</td>
<td>by genre</td>
<td>by genre, domain, country, period</td>
<td>by author, title, genre and period</td>
<td>One corpus</td>
<td>by genre and country</td>
<td>by genre, period and dialect</td>
<td>by genre, period and region of the author</td>
</tr>
<tr>
<td>Annotations provided</td>
<td>POS</td>
<td>Lemma POS</td>
<td>Lemma POS</td>
<td>Lemma POS</td>
<td>Lemma POS</td>
<td>Lemma</td>
<td></td>
</tr>
<tr>
<td>Tools</td>
<td>Concordancer Frequencies Collocations</td>
<td>Concordancer Frequencies Collocations</td>
<td>Concordancer (Regular expressions) Frequencies Collocations</td>
<td>Concordancer</td>
<td>Concordancer Frequencies</td>
<td>Concordancer Frequencies</td>
<td>Concordancer (Regular expressions)</td>
</tr>
</tbody>
</table>

**Table 1: Examples of Text Bases**
BaTelÒc: A text base for the Occitan language

The goal is to gather written literary texts (prose, drama, and poetry) and others genres, such as technical texts and newspapers, of modern and contemporary periods (from the 16th to the 21st century) and to embrace dialectal and spelling variations. There is enough material available to foresee a text base of hundreds of millions of words. The texts are well-documented with several types of metadata (genre, author’s name, year of author’s birth, dialect, year of publication, ...). The text base provides tools to allow feature-oriented selection of texts that can, thus, be gathered as a study corpus.

<table>
<thead>
<tr>
<th>Language status</th>
<th>Typical traditional corpus</th>
<th>BaTelÒc Text Base</th>
<th>Language Documentation corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Oral and written data</td>
<td>Written data</td>
<td>Recordings, transcriptions, translations</td>
</tr>
<tr>
<td>Resources gathered by</td>
<td>Team of native speakers</td>
<td>Teams of non-native speakers</td>
<td>One non-native speaker</td>
</tr>
<tr>
<td>Size</td>
<td>Millions of words</td>
<td>More than one million words</td>
<td>Less than one million words</td>
</tr>
<tr>
<td>Availability of data</td>
<td>Huge amount of data growing every day</td>
<td>Big amount of data</td>
<td>Few amount of data</td>
</tr>
<tr>
<td>Goals</td>
<td>Linguistic research</td>
<td>Preservation, linguistic and interdisciplinary research</td>
<td>Preservation, linguistic and interdisciplinary research</td>
</tr>
<tr>
<td>Compilation</td>
<td>Representative sample of one variety under study</td>
<td>All the possible opportunities</td>
<td>All the possible opportunities</td>
</tr>
</tbody>
</table>

Table 2: BaTelÒc Text Base between language documentation corpus and typical traditional corpus, inspired by Mosel (2013)

3. THE BaTelÒC TEXT BASE. For now, three million words have already been compiled within the text base. The texts are organized in nine genres (novel, poetry, memoir, short-story, fairy-tale, technical text, essay, song, treaty) and in five dialects (Gascon, Lengadocian, Lémosin, Provençau, and Auvernhàs). From a technical standpoint, BaTelÒc consists of texts encoded according to the international standard for sharing files on the internet, i.e. XML Text Encoding Initiative TEI P5, to ensure an accurate dissemination and the reusability of texts in the base and to provide high performance tools to select corpus (see Section 3.1).

BaTelÒc is designed to provide tools for linguistics studies, be it corpus linguistics or descriptive linguistics. The text base may be explored using a search engine that includes a concordancer to extract forms (word, part of word or sequence of words). It also includes more complex enquiries through the use of regular expressions (see Section 3.2).

3.1. BUILDING THE TEXT BASE. All the texts in the base are encoded according to XML TEI P5 format. XML is a computer language that defines a set of rules for encoding text segments using markups (recognizable by the use of angle brackets). This format is both human-readable and machine-readable. Every document is decomposed in two main parts. The head contains all the metadata about the document (author's name, year of author's birth, year of publication, dialect, spelling, ...). The body contains the whole document. Markups assign to text segments characteristics such as text formatting (bold, italic, ...) and document structure (paragraph, title, subtitle, ...). XML format is not a unitary format for all documents. It provides a markup syntax to be defined depending on needs and document types. Within BaTelÒc, we follow the Text Encoding Initiative P5 markup schema, which is a norm to encode digitized texts created for librarians and research in order to emphasize simplicity, generality, and usability of digitized texts and to ease search within texts. Figure 1 gives an extract of a BaTelÒc text in XML format.

```xml
<tei xmlns="http://www.tei-c.org/ns/1.0"
     xml:lang="ca"
     encoding="ISO-8859-1">
  <!--.DOCTYPE TEI SYSTEM "TeiP5.dtd">
  <teiHeader>
    <title>Un centenat de líneas de mi renuè</title>
    <text>
      <front>
        <docTitle>
          Un centenat de líneas de mi renuè
        </docTitle>
      </front>
      <body type="preface">
        <p>Sa nèu plan astrusac! Òc ben, plan astrusac! Així l'estilò negre de la pluma d'aur, l'estilò de totas las històries, de totas las jòias e de totas las lagremes, de tots los espènis e de tots las ràbias.</p>
        <p>Aqui i a tot lo tèxte ondrat de gavyèles.</p>
      </body>
    </text>
  </teiHeader>
</tei>
```

**Figure 1**: Extract of a BaTelÒc text in XML format

Several steps are needed in order to get a XML file from a rtf or doc file (see Figure 2 below): Manual pre-treatment to clean-up and mark-up for example the title of sections and sub-sections. The most important part of body markup is done automatically with a Perl program, for instance the markup of paragraphs, dialogs. All the metadata are saved in an Access database which generates the head of the file. Both elements together constitute the entire xml file.

3.2. BUILDING TOOLS TO SEARCH THE TEXT BASE. The exploration of the text base is done in two separated steps. For beginners, a “discovery corpus” is preselected as a default setting and it is searchable from the homepage. For advanced users, the basic consultation mode starts with building his/her own specific study corpus, based on various criteria. Once the corpus is defined, it may be used for the entire session. Tools are then provided to search for concordances.

3.2.1. CORPUS DEFINITION. Tools offer the possibility to select texts according to various criteria: author’s name, title of the book, year of publication, genres, dialects, spelling
Figure 2: Follow up processing to build a xml file

norms. This allows the definition and adaption of the corpus to specific research goals (for instance, a corpus organized according to genre for descriptive linguistic purposes, a corpus organized by time periods in order to study the diachronic evolution of Occitan, a corpus organized by dialects to study diatopic variation or a corpus organized by authors to study diastatic and diaphasic variation). Figure 3 shows an example where all the novels (roman) of the text base are selected.

Figure 3: Building a corpus of novels
3.2.2. Concordance Search. BaTelÒc provides tools to search for concordances of words or forms (for instance the word *çaquela* “however”). The results showing forms in context may be displayed in a narrow context (or KWIC), as in Figure 4, or in a larger context as in Figure 5.

![Figure 4: Searching for the form *çaquela* in a narrow context](image)

![Figure 5: Searching for the form *çaquela* in a larger context](image)

BaTelÒc also provides tools to look for two or three forms in a larger context. Figure 6 shows an example of *costat* “side” and *autre* “other” as search forms within an interval between 1 and 9 forms after the first form.

BaTelÒc search engine also includes more complex enquiries using regular expressions. Regular expressions are special text strings that allow more complex searches, i.e. through regular expressions it is possible to look for many different forms within one request. The disjunction (|) is used to look for alternative forms. For example, the following request *(esle|ei)* *(aquò|çò)* allows you to search for *es aquò, es çò, ei aquò, ei çò* – “it is that” (see Figure 7).

The functions “starts with” (*comença per*) and “ends by” (*s’acaba per*) can be used to look for regular forms inside of words such as prefixes and suffixes. Figure 8 presents an example of search for the suffix *–òt* (note that not all the found forms are suffixed words).
Figure 6: Searching for one form (costat) followed by another form (autre)

Figure 7: Searching for sequences of words es aquò, ei aquò, es çò, ei çò

as for example pòt "can" which is a finite verb). To look for all possible alternate forms (masculine, feminine, singular, plural), we can use the REGEXP function and the request may be formulated as follows: ..òta?s?$ (the two first dots aim at avoiding the very frequent verb form pòt) as is shown in Figure 9.

3.3. ENRICHING BATELÔC WITH LINGUISTIC ANNOTATIONS. Text base development logically involves enrichment of the texts with linguistic annotations. But Natural Language Processing (NLP) of endangered language such as Occitan is very challenging. It is not possible to transpose directly existing models for rich-resourced languages, partly because of the spelling and dialectal variations. We aim at providing corpora and lexi-
In the foreseeable future, we aim at enriching one part of the text base with POS and lemma annotations. Within the framework of the text base, lemmatization and POS will allow new request possibilities, for example the search of all finite verb forms or inflected forms of adjectives or the disambiguation of homographs such as *poder* (common noun “power”) and *poder* (verb “can”). One of the major difficulties to overcome is connected to the strong variation in written Occitan. The existence of numerous spellings is one of its cons in order to develop basic natural language processing tools, namely OCR (Urieli & Vergez-Couret 2013) and a Part-of-Speech (POS) tagger.\(^\text{16}\)

A Part-of-Speech tagger marks-up words in a text as corresponding to a particular part of speech with additional morphosyntactic information.

\(^{16}\) A Part-of-Speech tagger marks-up words in a text as corresponding to a particular part of speech with additional morphosyntactic information.
causes. The spelling used during the Middle Ages is called the “troubadour spelling”. This spelling disappeared gradually with the decline of the literary production. Since the 19th century, one can distinguish two major types of spellings: The Mistral’s spelling (created in the Provence and influenced by the French spelling) and the Gaston Febus’s spelling, which is used in Biarn. The last one appeared during the 20th century. It is a unified spelling, known as “classical spelling”, inspired by the “troubadour spelling”, and diffused in all Occitan territories (Sibille 2007). Another reason for the variation is related to the dialectal complexity of the language. The classical spelling naturally integrates the geolinguistic varieties (for instance lo filh vs. eth hilh “the son”, luna vs. lua “moon” or cabra vs. craba “goat”). Variations can also be seen as the result of the normalization process which is now in progress (for instance, differences in the spelling of conjugated verbs: avian vs. avián “they had”). Moreover, there are also phonological intra-dialectal variations reflected in the spelling (for example contes vs. condes “tales”). Because of spelling and dialectal variations in Occitan, it is difficult to simply apply the existing systems of POS tagging to create annotated corpora and large coverage lexicons as it is currently done with rich-resourced standardized languages.

In order to create morphosyntactic annotated corpora for Occitan, we first used a POS Tagger for Occitan available in the Apertium chain (Forcada et al. 2011). Apertium originally proposes open source systems for automatic translation, generally for related language pairs. Armentano I Oller (2008) developed a translation system for the Occitan/Catalan pair, which includes a POS tagger for Occitan. It is based on the use of one lexicon containing 36,500 entries. From our experience with this POS Tagger, three main difficulties can be raised:

a) If a word is not in the lexicon, no tag is proposed for this word. This results in strong performance variation depending on the texts used, especially because all possible spelling forms for all dialects are not included in the lexicon. Armentano I Oller announces an accuracy of 0.8 of correct tags for a text in Lengadocien (on an evaluation corpus of 600 tags). We made the same experiment on a text in Gascon (on an evaluation corpus of 1000 tags). We reached an accuracy of 0.6 of correct tags. Indeed, there were more words that could not be tagged in Gascon (19%) when compared to Languedocien (13%) (Vergez-Couret 2013).

b) The lexicon includes indiscriminately forms from various dialects. It should be evaluated if it is better to have one larger lexicon for all dialects or, on the contrary, one lexicon for each dialect.

c) The only way to improve the current performances of Apertium is to enrich the lexicon which can be a very time-consuming task.

To cope with this kind of problem for under-resourced languages, some researchers develop a system requiring a minimum of lexical resources. The main circumvention strategy is to use existing systems for a rich-resourced etymologically close language. Hana et al. (2011), for instance, use the proximity between Old Czech and Modern Czech (two successive states of the language) and Bernhard & Ligozat (2013) exploit the similarities between Alsatian and German (Alsatian being considered as a German dialect). We proposed to adapt the latter method for Occitan, using an etymologically related language – Castillan (Vergez-Couret 2013). It should be mentioned that the relation between these two languages is looser than the one described above. The method consists in running
existing morphosyntactic tools, here Tree Tagger for Castillan, on Occitan texts with a
pre-transposition of the most frequent words in Castillan first. This method only requires
the construction of a bilingual lexicon (Occitan/Castillan) of the 300 most frequent words
(based on a corpus) in Occitan\textsuperscript{17} and their translation into Castillan. We chose Tree Tagger
software (hereafter TT) (Schmid 1994) trained for Castillan. The TT’s performance relies
on the use of a large coverage lexicon. But unlike Apertium, TT predicts tags, using the
probability of POS tag sequences calculated on a manually tagged training corpus. We
assumed that probabilities of POS tag sequences would be fairly similar between Occitan
and Castillan.

– Transposition: This step consists in transposing the Occitan words which are in the
lexicon in Castillan (see Table 3). Only the bold words have been translated. The
other ones remain in their original form.

– POS tagging with TT: We used a simplified tag set (see Table 4) inspired by the one
used in Frantext. POS tagging was done twice, first on the original text and then
on the transposed text. Table 3 gives an example of tags for original texts and for
transposed texts. The symbol ✓ means known words and ✗ unknown words by TT.
Correct tags are greyed out.

– Running TT on original text: Some words are graphically similar in each pair of
languages. For example, the feminine singular definite article is la both in Occitan
and Castillan (see Table 3). TT will consider them as known words (✓) and will use
the available information about this word. Nonetheless, it does not insure that similar
words in the two languages will have the same POS.

– Running TT on transposed text: In the transposed text, similar words and transposed
words are considered known by TT.

The precision of Apertium on original text in Occitan is rather low, 0.65. As we ex-
plained above, Apertium only assigns tags to known words. As a consequence, the perfor-
mancess correlate significantly with the number of unknown words (19% for our evaluation
corpus). The Spanish TT precision for the original text is unsurprisingly low, less than
0.5. After transposition, the precision reaches up to 0.8. These experiments show that the
methodology first used between a language and one of its dialects (Bernhard & Ligozat,
2013) is exportable with similar results for pairs of languages, less close even if etymologi-
cally related.

Our approach is resource-free and gives a precision of 0.8 but this system is more
robust to deal with variations than Apertium. Nevertheless, improvements are required
to raise the precision up to 0.95, as usually expected for POS tagging. The strategy we
retained was, therefore, to use the two systems described above to foster the very long and
fastidious process of creating a gold standard for POS annotation in Occitan. The corrected
annotations were used for training more traditional supervised learning systems such as
Tree Tagger or Talismane (Urieli & Tanguy 2013), see Vergez-Couret & Urieli (2015) for
more details. Along with this comes the building of lexicons.

\textsuperscript{17} For the first experiment we used the Gascon dialect.
BaTelÔc: A text base for the Occitan language

<table>
<thead>
<tr>
<th>Original text</th>
<th>Tag</th>
<th>Trans_ca</th>
<th>Tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dab</td>
<td>X</td>
<td>Con</td>
<td>✓</td>
</tr>
<tr>
<td>la</td>
<td>✓</td>
<td>D</td>
<td>✓</td>
</tr>
<tr>
<td>complicitat</td>
<td>X</td>
<td>S</td>
<td>✓</td>
</tr>
<tr>
<td>de</td>
<td>✓</td>
<td>Pp</td>
<td>✓</td>
</tr>
<tr>
<td>la</td>
<td>✓</td>
<td>D</td>
<td>✓</td>
</tr>
<tr>
<td>lua</td>
<td>X</td>
<td>S</td>
<td>✓</td>
</tr>
<tr>
<td>que</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vau</td>
<td>X</td>
<td>A</td>
<td>✓</td>
</tr>
<tr>
<td>poder</td>
<td>✓</td>
<td>Vi</td>
<td>✓</td>
</tr>
<tr>
<td>adara</td>
<td>X</td>
<td>A</td>
<td>✓</td>
</tr>
<tr>
<td>tirar</td>
<td>✓</td>
<td>Vi</td>
<td>✓</td>
</tr>
<tr>
<td>camin</td>
<td>X</td>
<td>S</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Table 3:** Extract from original texts in Occitan and transposed texts in Spanish tagged with Spanish TT

<table>
<thead>
<tr>
<th>Dca</th>
<th>Cardinal number as article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pe</td>
<td>Enunciative particle</td>
</tr>
<tr>
<td>V</td>
<td>Finite verb (except Vi, Vpp, Vps)</td>
</tr>
<tr>
<td>Vi</td>
<td>Infinite verb</td>
</tr>
<tr>
<td>Vpp</td>
<td>Present participle</td>
</tr>
<tr>
<td>Vps</td>
<td>Past participle</td>
</tr>
<tr>
<td>Inj</td>
<td>Interjection</td>
</tr>
<tr>
<td>Np</td>
<td>Proper noun</td>
</tr>
<tr>
<td>S</td>
<td>Common noun</td>
</tr>
<tr>
<td>P</td>
<td>Pronoun</td>
</tr>
<tr>
<td>Pp</td>
<td>Preposition</td>
</tr>
<tr>
<td>Pr</td>
<td>Relative pronoun</td>
</tr>
<tr>
<td>&lt;sent&gt;</td>
<td>End of sentence</td>
</tr>
<tr>
<td>Cm</td>
<td>Comma</td>
</tr>
</tbody>
</table>

**Table 4:** Corresponding tag set

<table>
<thead>
<tr>
<th>Original text</th>
<th>Transposed text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apertium</td>
<td>0.65</td>
</tr>
<tr>
<td>Spanish TT</td>
<td>0.46</td>
</tr>
</tbody>
</table>

**Table 5:** TT precision

---

18 We deleted all the enunciative particles because there is no equivalent in Castillan and the tags for them would have been inevitably wrong. While these particles play a role at the enunciative level, the following propositions are still grammatical.
As far as the lexicon is concerned, available lexical resources for Occitan are mostly dictionaries, in paper and electronic format. But those resources do not conform to the standardized norm (Text Encoding Initiative, Lexical Markup Framework) and are not currently operable for NLP systems the way they are. We are constructing a lexicon, based on the entries of the Laus dictionaries (Laus 2001, 2005) and a list of finite verbal forms provided by Lo Congrès Permanent de la Lenga Occitana (Verbòc app). The lexicon currently includes around 700,000 inflected forms. But, as we have mentioned above, all the spelling, dialectal, and intra-dialectal variations are not included in the lexicon. It will be necessary to find strategies to enrich the lexicon, for example by using morphological similarities (Baroni et al. 2012). Other types of strategies to enrich the lexicon using existing resources for a related rich-resourced language would also help. For example, Scherrer & Sagot (2013) acquire German/Palatin (a German dialect) cognate pairs with unsupervised automatic learning methods. Such cognates could be used in our case for pairing dialectal and spelling variations.

4. Perspectives. The text base is now available online. The next step will be to design new tools for calculating frequencies and to gradually increase the amount of texts. The strategy we want to retain is to increase the range of genres, domains, dialects, neither following the lead of Frantext which pays greater attention to literature, nor the example of English balanced corpus, but rather gathering as many various data as possible to satisfy as many BaTelÒc users as possible. BaTelÒc consists in a two-fold effort: to document the language and to provide data for linguistic studies. We hope that providing linguistic data will help the research field of Occitan linguistics to gain new researchers, which is another way of playing a role in the conservation of the Occitan language.

References


19 BaTelÒc can be consulted on the REDAC website: http://redac.univ-tlse2.fr (20 January 2016). REDAC gathers the linguistic resources developed at CLLE-ERSS research unit and makes them available for download or browsing.


Myriam Bras
myriam.bras@univ-tlse2.fr

Marianne Vergez-Couret
vergez@univ-tlse2.fr
The first Mirandese text-to-speech system

José Pedro Ferreira\textsuperscript{a}, Cristiano Chesi\textsuperscript{bc}, Daan Baldewijns\textsuperscript{c}, Daniela Braga\textsuperscript{d}, Miguel Dias\textsuperscript{ce}, and Margarita Correia\textsuperscript{af}

\textsuperscript{a}CELGA-ILTEC (Universidade de Coimbra), \textsuperscript{b}Instituto Universitario di Studi Superiori (Pavia), \textsuperscript{c}Microsoft Language Development Center (Lisbon), \textsuperscript{d}DefinedCrowd, \textsuperscript{e}ISCTE - University Institute of Lisbon, \textsuperscript{f}Universidade de Lisboa

This paper describes the creation of base NLP resources and tools for an under-resourced minority language spoken in Portugal, Mirandese, in the context of the generation of a text-to-speech system, a collaborative citizenship project between Microsoft, ILTEC, and ALM – Associaçon de la Lhéngua Mirandesa. Development efforts encompassed the compilation of a large textual corpus, definition of a complete phone-set, development of a tokenizer, inflector, TN and GTP modules, and creation of a large phonetic lexicon with syllable segmentation, stress mark-up, and POS. The TTS system will provide an open access web interface freely available to the community, along with the other resources. We took advantage of mature tools, resources, and processes already available for phylogenetically-close languages, allowing us to cut development time and resources to a great extent, a solution that can be viable for other lesser-spoken languages which enjoy a similar situation.

1. INTRODUCTION. While the current discussion on NLP resources for some of the most widely spoken European languages focuses on theoretical proposals to further develop them and on how to bridge the gap with the most advanced systems available for English (Branco et al. 2012), some lesser-spoken ones, Mirandese being the prototypical example, still lack the most basic language resources, with severe consequences for research, teaching, and applications available to citizens. Prior to the 1990s, when this language was officially recognized and normalization efforts began (Barros Ferreira 2002), little attention had been given to this language – to this day, the most complete studies are still those done at the turn of the 19th century (Leite Vasconcelos 1899–1900). Lack of an established writing standard, little written production, and low perceived economic value were responsible for the near-inexistence of modern studies and language resources. This, in turn, contributed throughout the years to a low perceived sociolinguistic value of the language among its speakers and lack of access to modern technology in their native tongue.

This paper outlines the tasks that were carried out in a citizenship project, a collaboration between Microsoft, ILTEC, and the not-for-profit, speaking-community-led ALM – Associaçon de la Lhéngua Mirandesa (Mirandese Language Association), that aimed at developing basic NLP tools for Mirandese from scratch, having a high quality speech synthesis system generation process (Braga et al. 2010) as the final goal. This paper de-
The first Mirandese text-to-speech system

scribes such process, illustrating how the language resources development tasks evolved in relatively little time and with relatively few human resources and investment, through the adaptation and reuse of existing tools and resources available for phylogenetically-close languages. All the language resources that were developed will be made freely available for R&D purposes, through an open-access web interface, Casa de la Lhéngua, a portal which is to be managed by ALM.

2. ON MIRANDESE. Mirandese is a minority language spoken in Northeastern Portugal. It belongs to the Astur-Leonese group of West Iberian languages, being closely related to Asturian, spoken to this day in areas of the Asturias and Leon autonomous communities in Spain, with which Mirandese no longer retains a linguistic continuum. Unwritten for most of its history, Mirandese was first scientifically identified and studied in the late 19th century (Leite Vasconcelos 1899–1900).

Throughout the 20th century, strong demographic changes, namely the exodus of large numbers of people through emigration in the 1940s and in the 1970s, an influx of non-Mirandese speaking workers from various other parts of the country in the 1950s and 1960s, along with the rise of Portuguese-spoken-only media, led to inter-generational transmission to be gradually abandoned, leaving the language seriously threatened. Today, it is estimated that Mirandese is spoken by no more than 5,000 people as a first language, and by at most 15,000 in total, counting heritage and second language speakers, including those living outside of Terra de Miranda (Barros Ferreira 2002).

In the 1990s, strong efforts began to be made to make the survival of Mirandese possible: A group of linguists and native speakers managed to reach an agreement for a spelling convention common to different varieties, and the language was introduced into the formal education curricula locally, although with limited scope and only as an option. The Portuguese State finally granted the language co-official regional status in 1999 (Barros Ferreira 2002).

These initiatives had a strong impact on its speakers: What used to be perceived as a reason for shame by many in the diglossic community increasingly started showing up in book shelves, in the local and national media, and on the Web. Albeit seriously threatened as a mother tongue, it is currently learned and used by a large part of the population of Miranda in increasingly more formal contexts, currently enjoying a period of non-artificial revival.

3. LANGUAGE RESOURCES AS A MEANS AND A GOAL. Before the efforts this paper describes were made, although some initial efforts towards developing speech technologies had been set about (Trancoso et al. 2003, Caseiro et al. 2003), there were little or no available modern language resources for Mirandese. The most detailed linguistic descriptions are to this day those made by Leite Vasconcelos (1899–1900), more than 100 years ago, in part due to the lack of available data (Barros Ferreira 2002), leaving researchers in need of conducting original fieldwork to get in touch with actual large-scale data, and school pupils with little up-to-date base tools for the formal study of the language. Additionally, the fact that Mirandese is present in more and more support formats and usage contexts seems to be a decisive factor in the way its speakers perceive the language, granting it a higher sociolinguistic profile (Barros Ferreira 2002). These facts, along with the will to develop a Text-to-Speech (TTS) system for Mirandese, were the spark behind the effort to create the resources presented in this paper.
To achieve the end goal of creating a TTS system, a number of language resources that are usually available for more widely spoken European languages had to be developed from scratch. For the voice font generation, an existing and proven process could be followed, using previously existing tools designed for larger and better-resourced languages (Braga et al. 2008b). Work for this project encompassed the creation of a large text corpus, the definition of a complete phone-set, the development of tokenizer, inflector, text normalization (TN) and grapheme-to-phoneme (GTP) modules, and the creation of a large phonetic lexicon with part-of-speech (POS) classification.

The corpus was compiled from raw textual data collected by ALM, most of it generously provided by publishers, newspapers, and the authors themselves. Those data were then complemented with data crawled from the web using the work developed by Scannell (2007), increasing the total size of the corpus to over one million tokens.

The lexicon was built using the currently 25K-strong lemma list of the ongoing work on a Mirandese-Portuguese dictionary (Ferreira & Ferreira 2001–), complemented with the most frequent lemmas in the compiled text corpus, which was previously tokenized, lemmatized, and POS-tagged using customizations of García & Gamallo (2010) and Janssen (2012). The resulting lemma list was then inflected using a version of Janssen (2011), syllabified, stress-marked, and converted to IPA phonetic transcription using an in-house adaptation of an unpublished two-step Perl-based GTP tool originally developed for European Portuguese (Janssen & Santos 2012). This simple regular expression string replacement set of scripts starts by marking up stress and syllable divisions over the orthographic forms and, in a subsequent phase, applies an ordered set of grapheme-phone transformation rules based on syllable position and stress, taking advantage of the relatively shallow phonemic orthography of Mirandese.

A TN module for Mirandese was also put in place for the first time for this language, its rule-set being developed with the aid of previously in-house developed software (Cho et al. 2010) and taking advantage of the availability of a counterpart file for European Portuguese. The proximity between the two languages made it possible to reuse the pre-existing resource changing only minimally several hundreds of the thousands of rules and terminals that compose the module, thus greatly speeding up the TN module development process.

4. CREATION OF THE TEXT-TO-SPEECH SYSTEM. The Voice Font Building procedure that allowed us to create the TTS system discussed here can be described at three levels:

1. TTS front-end (preparation of language resources described in Section 3, text analysis);
2. Voice font building (creation of text prompts, voice talent selection and recording of uttered prompts, voice font compilation);
3. TTS back-end (voice font training using SPS – Statistical Parameter Synthesis, prosodic model creation and tuning).
4.1. LANGUAGE RESOURCES PREPARATION AND TEXT ANALYSIS. As detailed in Section 3, the preparation of language resources consisted of creating, collecting, annotating, and validating the linguistic information that is required for the voice font building procedure. This included the compilation of a fully annotated large lexicon, consisting of 124,360 word forms, for which standard orthography, pronunciation (syllabified, stress marked IPA transcription), POS (e.g. VER for verb, ADJ for adjective), as well as other morphological information (like mood, tense, person and number features) are provided, as exemplified in (1):

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
<th>POS</th>
<th>morphological features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abacelhe</td>
<td>ax - b ax - s eh 1 - lh aex</td>
<td>VER</td>
<td>subjunc., present, 3p, s</td>
</tr>
<tr>
<td>melhor</td>
<td>m aex - lh oh r 1</td>
<td>ADJ</td>
<td>qualifying, m, s</td>
</tr>
</tbody>
</table>

Another important resource that needed to be developed, was a complete phone set for Mirandese, consisting of 46 distinct phones, which takes us forward from prior work (Trancoso et al. 2003). This resource specifies the full list of available phones paired to distinctive features and parameters that are used to train the voice model (e.g. voiced, velar, dental, liquid, main stress, secondary stress, etc.) as exemplified schematically in (2).

<table>
<thead>
<tr>
<th>Phone</th>
<th>features</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>voiced bilabial plosive</td>
</tr>
<tr>
<td>t</td>
<td>voiced dental plosive</td>
</tr>
</tbody>
</table>

The text normalization module is composed of about 5,000 (contextual) rules dealing with the expansion of cardinal numbers (“12” > “twelve”) or date-time spell out (“12:00” > “noon”), for instance. The following TN categories were developed for Mirandese: cardinals, ordinals, percentage expressions, simple mathematical expressions, date and time expressions, currency, phone numbers, roman numerals, fractions, measurement expressions, titles, addresses, URLs and email, and file paths. Where needed, context-sensitive rules were made for instance to ensure the correct gender agreement in noun phrases containing a cardinal number.

From the resources mentioned above, we managed to easily generate syllabification rules (using algorithms like the one discussed in Janssen & Santos (2012)) that allow us to segment words in the lexicon and pair them with their correct pronunciation and the most likely stress.

4.2. PROMPTS CREATION, VOICE SELECTION AND RECORDING. On par with these development tasks, a voice talent was selected for high-quality recording sessions of 5,132 prompts retrieved from the corpus. The prompts consisted of full sentences selected based on character length and phonological relevance (richness of phonological contexts), determined by an existing algorithm of the text TTS system software suite.

The voice talent was selected from a pool of candidates by a jury of 20 native speakers of varying ages, provenances, and sociolinguistic profiles. Public advertisement in the local media and speaking community networking helped greatly in getting a reasonable number of candidates with the correct profile: Native speakers, having at least undergone undergraduate studies and no older than 40. The jury listened to recordings of each candidate reading an expressive text, and filled in a short questionnaire. The two highest ranked candidates in this first phase underwent one hour of pure speech studio recording under
The first Mirandese text-to-speech system

loose scrutiny and were again ranked by the jury, who this time had to fill in a more thorough questionnaire developed for subjective pleasantness assessment, using a methodology published by Braga et al. (2008a).

Finally, the now elected voice talent was recorded over two weeks in a high-quality studio under the close supervision of a language expert, who monitored the clarity, accent, and completeness of the recording process, simultaneously checking the adequacy of each prompt and making textual corrections where needed. The recording process yielded over 7 hours of speech data.

Those data were semi-automatically trimmed and chopped into individual files using a standard acoustic marker inserted between prompts during the recording process, making it easier to map each individual recording file with a prompt. All the individual recordings were listened to by a language expert, and removed from the pool of available data when quality or conformity with the prompt was not met.

4.3. THE TTS BACK-END. After being properly segmented (both sentence segmentation and word segmentation is needed) and fully normalized, the prompts used for guiding the recording procedure were analyzed. This procedure was automatically carried out using rule-based sentence breakers, contextual text normalization (TN) rules (as discussed in Section 4.1), and POS taggers (e.g. Ratnaparkhi 1996). Once single words are normalized and categorized, the correct pronunciation is retrieved from the lexicon and assigned to the current word.

In the end, each prompt was enriched with several types of information, as shown in Figure 1:

![Figure 1](image)

Such fully annotated prompts made the TTS voice font training procedure possible: The approach used to train the font model is called Statistical Parameter Synthesis (SPS) and it is based on standard Hidden-Markov-Model (HMM) approaches to TTS (HTS, Zen et al. 2007, Zen et al. 2009).

The basic idea is that the waveform is stable during short time phrases and can be approximated by Gaussian models that represent the parameter distribution. Given a sequence
of observation \( (O_i) \), we expect an observation \( O_i \) at time \( i \) to belong to one state \( Q \) (e.g. 1, 2 or 3). In \( i+1 \), \( O_{i+1} \) might still be \( Q \) or a different state and this must be modeled depending on the transition probability built on the previous observation, based on the aligned wave – the prompt pairs we use for training. We used a decision tree based on relevant distinctive feature associated to a given state (expressed by Gaussian models), to better estimate the state sequence (Figure 2) (as well as its duration and excitation).

\[
\begin{align*}
\text{k + a + p} & & \text{L = voice?} \\
\text{p + a + n} & & \text{L = vowel?} \quad \text{R = silence?}
\end{align*}
\]

**Figure 2: Gaussian Models training, organized by Decision Tree**

Notice that the SPS procedure not only allows using distinctive phonetic features (linear spectrum pair, LSP model, cf. Zheng 2000) as parameters, but also prosodic cues, like pitch (F0). This allows us to both keep the advantages of having an HTS Voice Font\(^1\) (high flexibility, small font size) and to limit its disadvantages (muffle voice quality, flat prosody). The training process uses a gradient descent algorithm (Minimum Generation Error, MGE, cf. Yi-Jian & Wang 2006).

In the end, the (trained) decision tree is used in generation to select and concatenate the state models by maximizing the likelihood of the parameter sequence. The result of this process is a fully intelligible TTS voice font.

---

\(^1\) In generation, the model parameter is used to create the wave form signal.
5. CONCLUSIONS AND FUTURE WORK. In this paper, we presented the development of the first TTS system for Mirandese reaching intelligibility. The data and language resources developed to achieve this goal are being made available to the speaking and scientific community through the speech-community-led Casa de la Lhéngua, a free access web interface still under development. It rests to be seen if the secondary objective of granting Mirandese a stronger sociolinguistic profile within its speech community will be aided by the availability of these tools.

Future work should include the conversion of the NLP resources we developed to internationally standardized formats.

REFERENCES


The first Mirandese text-to-speech system


José Pedro Ferreira  
jpf@uc.pt

Cristiano Chesi  
t-chres@microsoft.com

Daan Baldewijns  
v-daanb@microsoft.com

Daniela Braga  
d-braga@hotmail.com

Miguel Dias  
Miguel.Dias@microsoft.com

Margarita Correia  
margarita@campus.ul.pt
Bridging divides: A proposal for integrating the teaching, research and revitalization of Nahuatl

Justyna Olko\textsuperscript{a} and John Sullivan\textsuperscript{abc}

\textit{\textsuperscript{a}University of Warsaw, \textsuperscript{b}Universidad Autonóma de Zacatecas, \textsuperscript{c}Instituto de Docencia e Investigación Etnológica de Zacatecas}

This paper discusses major historical, cultural, linguistic, social and institutional factors contributing to the shift and endangerment of the Nahuatl language in Mexico. As a practical proposal, we discuss our strategy for its revitalization, as well as a series of projects and activities we have been carrying out for the last several years. Crucial to this approach are several complementary elements: interdisciplinary research, including documentary work, as well as investigation of both the historical and the present state of Nahua language and culture; integration of both Western and native-speaking indigenous researchers as equal partners and the provision of space for indigenous methodologies; creation of teaching programs for native and non-native speakers oriented toward the preparation of language materials; and close collaboration with indigenous communities in developing community-based programs. The operability of this strategy will depend greatly on our ability to foster collaboration across academic, social, and ideological boundaries, to integrate theory, methodology and program implementation, and to efficiently combine grassroots and top-down approaches. An important aim is to restore the culture of literacy in Nahuatl through our monolingual \textit{Totlahtol} series, publishing works from all variants of the language and encompassing all genres of writing. We also strive to strengthen the historical and cultural identity of native speakers by facilitating their access to the alphabetical texts written by their ancestors during the colonial era.

\textit{As events of the past few years and the present in various parts of the world show, our global village must be truly multicultural and multilingual, or it will not exist at all.} \hfill [Nettle & Romaine 2000: 204]

1. INTRODUCTION. Nahuatl, the language of the Aztecs and still the largest indigenous tongue in North America with reportedly 1.5 million speakers would not seem to be in

\textsuperscript{1} This research was financed within the programme of the Minister of Science and Higher Education in Poland, under the name “National Programme for the Development of the Humanities” between the years 2013 and 2016. This paper builds on our two earlier publications dealing with the revitalization of Nahuatl (Olko & Sullivan 2014 a, b). In addition to expanding some of their major ideas, it discusses recent revitalization activities and their aims.
danger of extinction, but in fact it is. The prevailing attitude of racism in Mexican society along with a stepped up national educational and mass media policy of Hispanization has drastically slowed the intergenerational transfer of the language. We begin this paper by evaluating the current situation of Nahuatl, starting with its historical background and then discussing the problems and challenges it faces today, identifying the major cultural, linguistic, social and institutional factors contributing to its endangerment. We then formulate a proposal for revitalizing the language which integrates teaching, interdisciplinary research and concrete revitalization activities into a wide and diverse network of collaboration. A fundamental aspect of our methodology involves transforming the traditional academic division between the ethnographer and the language community under study: native speakers can be trained to do research, collaborate on projects with non-indigenous investigators as well as create and implement their own research and teaching methodology. This will not only empower native speakers, but will enrich ethnographic research with the addition of the insider perspective which it has always lacked. The important culture of Nahuatl literacy developed during the colonial period needs to be revived and extended in indigenous communities in order to strengthen their historical and cultural identity: we can promote creative and academic writing in Nahuatl, publish contemporary and older texts in standardized orthography, circulate them and encourage people to read and discuss them. The isolation of Nahuatl communities and the lack of interregional communication can be overcome by holding interdialectal encounters, both in person and using videoconferencing technology, and by promoting monolingual communication in indigenous languages in the social media. Finally, we need to tear down the existing ideological barriers to revitalization by widely disseminating the results of research showing the clear and irrefutable benefits that multilingualism offers to all of society.

An important framework for our research and revitalization activities is an international research project, *Endangered languages. Comprehensive models for research and revitalization*, that deals with three minority languages in two countries: Nahuatl in Mexico and Wymysióery’s and Lemko in Poland. Despite important differences, such as the economic, sociopolitical and cultural contexts in which these communities operate, many challenges and problems are shared by all the three minority groups: they include the diminishing role or virtual lack of intergenerational transmission, the absence of efficient support and monolingual spaces in the educational system, the unavailability of sufficient literary and educational materials, and a pervasive negative language ideology. The collaborative activities we have carried out in Wilamowice, Poland, during the last years are proving crucial for the development of strategies and activities aimed at the revitalization of Nahuatl. They have led to the implementation of new forms of academic and non-academic partnerships, including an efficient mode of collaboration between two leading Polish universities, local non-profit organizations and activists, municipal authorities, school authorities and inter-

---

2 While the Mexican National Institute of Statistics and Geography (INEGI) cites a population of 1,544,968 native speakers of Nahuatl, 5 years or older, based on the 2010 national census, there are no reliable statistics regarding active versus passive speakers, literacy, every-day language use, intergenerational transmission or access to Nahuatl education at school.

3 This project is financed within the National Program for the Development of the Humanities of the Polish Ministry of Science and Higher Education, and carried out at the Faculty of “Artes Liberales” of the University of Warsaw, with the direct participation of the Instituto de Docencia e Investigación Etnológica de Zacatecas (IDIEZ), in collaboration with Adam Mickiewicz University in Poznań and the “Wilamowianie” Society. It complements our major research project, *Europe and America in Contact. A multidisciplinary study of cross-cultural transfer in the New World across time (2012–2017)*, financed by the European Research Council within the Ideas Program, and focusing on language and culture change and continuity in Nahua culture across five centuries of interaction with Spanish/Mexican culture.
national groups of participating scholars. As a result, language instruction has begun at a local school; language transmission has been reestablished and several young neo-speakers have appeared; literary and teaching materials have been published; vivid artistic and dissemination activities related to the language have been launched with broad community participation; a notable positive change of attitudes toward the language has begun to manifest itself in the community and more broadly in Polish society; and finally, the local economy is exploring commercialization opportunities related to linguistic-cultural heritage, involving the creation of a touristic cluster in order to offer a broad range of activities promoting local language and culture. This experience is extremely valuable and useful for other projects and there are essential elements of an integral strategy that can be applied in both Polish and Mexican contexts. However, we have also become increasingly aware of specific differences and necessities regarding conditions and elements of language revitalization programs. One of them concerns the overt involvement of academic partners in collaboration with local institutions and agents: the successful model implemented in Wilamowice, fully complying with and supported by national and European legislation and academic practices, rules of funding and ethical concerns and procedures, cannot serve as a direct model for working with Nahua communities. Perhaps the fundamental characteristic of the Mexican context involves the colonial and postcolonial policy of dominance and discrimination over indigenous communities, and the way in which its enduring impacts and threats are currently perceived by the members of those communities. This context constitutes an unresolved challenge, both for indigenous people and for collaborating external partners who are interested in revitalization; it must be taken into account when planning and implementing each community-based project. The revitalization of literary languages in wealthy countries, especially those that possess a long tradition and enjoy a healthy degree of institutional support, differs greatly from projects dealing with unwritten languages in developing countries plagued by serious economic problems, migration, a colonial/postcolonial heritage and discrimination over indigenous communities, and the way in which its enduring impacts and threats are currently perceived by the members of those communities. This context combines grass-roots or community-based approaches with certain top-down forms of support, and allows for academic partners to play an important and positive role.

2. HISTORICAL BACKGROUND OF NAHUATL. Nahuatl, a Uto-Aztecan language, enjoyed great importance in pre-Hispanic Mesoamerica over a long period of time, and its speakers have survived to this day, inhabiting several regions of Mexico. Although it was recently suggested that the Proto-Uto-Aztecan community developed in Mesoamerica between the time when maize was first domesticated and 4500 B.P. (Hill 2001: 913–934), more tangible evidence of the preconquest history of Nahuatl appears in the first half of the first millennium AD during the time of the Teotihuacan empire when loanwords from Nahuatl appear in Maya script (Macri & Looper 2003). While identification of the dominant language of Teotihuacan remains a controversial issue, it is generally acknowledged that the Toltecs (ca. 800–1050 AD) spoke Nahuatl, but there was probably a major dichotomy characterizing its variants, consisting on the one hand, of the early arrivals identified as “Toltecs” and, on the other hand, of the later Chichimec migrants who came in several waves after ca. 1200 AD. In addition, and as a result of population shifts, move-
Bridging divides

162

ments, and influences branching in many directions in central Mexico, the distinguishing traits which at one time characterized these two major groups would have been modified, blurred or lost (Canger 1988: 63). Thus, it has been proposed that the first group of Nahuatl speakers, including the “Toltecs” in central Mexico and further south were the ancestors of today’s users of the variants of La Huasteca, Sierra de Puebla, Isthmus, and Pipil. The later incomers would have spread into the Valley of Mexico and to the east and south, including Tlaxcala, central Puebla, Morelos, and to a certain degree Central Guerrero, perhaps contributing to a three-way geographical split in the early group: La Huasteca, Sierra de Puebla, and Isthmus (Canger 1988: 64–65).

Best documented are the Central Mexican Nahuas who at the time of the Spanish conquest populated numerous local ethnic states (altepetl), most of which before 1519 were in some way involved with the powerful organization of the Triple Alliance, often called the Aztec empire by scholars today. Nahuatl was used as a lingua franca throughout the empire and beyond. Although it collapsed upon the arrival of Spaniards, local Nahuas states (altepetl) survived, maintaining much of their political organization and many other aspects of their culture, in spite of becoming part of New Spain and thus the object of prolonged Hispanization. In the following centuries they continued to function as the seats of Indian municipal government based on European models. The Nahuatl language thrived in the new colonial contexts and was widely used for administrative and religious purposes across Spanish Mesoamerica, including regions where other native tongues prevailed.

Having their own preconquest tradition of books and glyphic records, the Nahuas were well prepared for the arrival of alphabetic writing. They immediately assimilated this tool and used it prolifically, producing an extremely rich and complex corpus of written texts which embraced historical annals, speeches, plays, petitions, assertions of local traditions and rights called “titles”, religious texts and a mass of everyday documentation, including wills, bills of sale, parish records, and censuses. The rapid development of the Nahuatl writing tradition was made possible by the adaption of the orthographic conventions of the Roman alphabet in the 1530s in such major centers as Mexico-Tenochtitlan and Tlatelolco. Beginning in the 1540s various kinds of writing in Nahuatl expanded quickly across the core area of Nahua culture and beyond. By the third quarter of the sixteenth century even small towns had a notary associated with the municipal government (Lockhart 1992: 330–331). The creation and development of Nahuatl orthography was a task undertaken simultaneously by several friars and their indigenous assistants. It was based on the Spanish values of the Roman alphabet representing similar sounds in Nahuatl, a process which was facilitated by the fact that Spanish had close equivalents for the majority of phonetic elements in the native language. In fact, it was Nahuatl that lacked more of the Peninsular sounds. Several phonological features of Nahuatl nevertheless posed a serious challenge. The glottal stop and vowel length were usually left unmarked, but other non-compatible elements were coped with quite well. The native sounds tl and tz were rendered as digraphs, while the double l, lacking in Spanish, was modeled on the Latin ll. Early orthographers also became aware of the fact that in Nahuatl voiced consonants are voiceless at the end of a syllable, so they changed prevocalic hu– [w] to –uh in the syllable-final position, doing the same with –uc and –cuh for the sound [kʷ].

This system, first developed by ecclesiastics, was immediately reshaped by native scribes and authors, whose primary concern, differing from the European priority given to standardized, conventional forms, was to reproduce not only orality, but also phonetic features that could change as a result of phonetic interaction with the sounds of neighboring words. Unlike for Spaniards, the word as such was neither an important nor easily recognizable en-
tity for the Nahuas, who tended to record sounds in an ongoing string of letters (Lockhart 1992: 336–339). This native adaptation and the relative flexibility governing use of the orthographic conventions did not disappear over time and never gave way to full standardization. Although there were further attempts at standardization undertaken by the Europeans, such as Horacio Carochi who published his outstanding Gramática de la Lengua Méxicana in 1645, these had little impact on the traditions of literacy and ways of writing in native communities. Toward the late seventeenth and into the eighteenth century, orthography in indigenous writing became more regionalized, reflecting local, unstandardized variants of spelling (Lockhart 1991: 122–134, Pizzigoni 2007: 35–39). In spite of the lack of standardization, Nahuatl writing and its associated orthography represented a native development, and constituted a long-standing literary tradition used by the Nahuas for their own purposes through the colonial period. As such, this orthographic tradition should be taken into account in the discussion of modern orthographic conventions of Nahuatl (see below), particularly because it has the potential for reinforcing the historical identity of today’s Nahuas, which is crucial for language revitalization programs.

Although colonial language policy and Hispanization is often blamed today as the main cause of language shift and the gradual displacement of Nahuatl, legal steps, such as the decision of the Spanish king Philip II in 1570 to make Nahuatl the linguistic medium for religious conversion and for the training of priests and friars working with the native people in different regions, no doubt contributed to its growing importance in Spanish Mesoamerica. It is becoming clear that the use of Nahuatl in the colonial world was not limited to a specially trained group of scribes, notaries and other officials. Members of the nobility belonging to other ethnic groups, as well as numerous non-elite figures of different backgrounds, including Spaniards, used spoken and written Nahuatl to facilitate communication in different aspects of colonial life (Yannanakis 2012: 669–670, Nesvig 2012: 739–758). One of the basic postconquest uses of Nahuatl beyond Nahuatl-speaking communities was Christian instruction carried out by friars and priests, who were allowed to be ordained a título de lengua, for the purpose of working as a kind of doctrinal interpreter in indigenous languages. Nahuatl was by far the most commonly spoken indigenous tongue among ecclesiastics. They used it as the language of instruction within linguistically-mixed communities whose members knew Nahuatl as a second language, and in regions dominated by other ethnic groups, such as the diocese of Oaxaca, where Nahuatl had already served widely as a lingua franca (Schwaller 2012: 678–687).

3. CURRENT SITUATION AND DEGREES OF ENDANGERMENT OF NAHUATL. From the first decades of contact, Nahuatl and other native languages began to evolve in response to the strong and long-term impact of Spanish, undergoing profound changes in a process that continues today. In spite of this heavy influence and a constantly growing number of bilinguals, now a clear majority in native communities, local variants of the language reveal a strong continuity with colonial Nahuatl, a fact which is often denied in mainstream and popular ideology. This continuity is also not sufficiently highlighted by existing scholarship, usually due to the simple reason that researchers specialize exclusively either in colonial or in modern data, making no attempt to connect these phenomena by seeing the language and cultural development over the long-term. As a result, views of modern

---

4 Carochi proposed the use of a system of diacritics to represent vocalic length and the glottal stop. Nevertheless, and as a rule, indigenous writers never considered the representation of these two phonetic characteristics important.

5 “by right of competence in an indigenous language” (Taylor 1996: 94–95).
Nahuatl in academic research have contributed to the current depreciated status of the language and its speakers. The notion of “Classical Nahuatl” has long been considered the only correct and original form of the language, while modern “dialects” are often still seen as little more than its corrupted, Spanish-influenced developments. Perhaps for this reason, Nahuatl dialectology has attracted surprisingly little attention among scholars dealing with diverse aspects of Nahua culture (Canger 1988: 29). The current state of language change in different regions varies considerably, depending both on the degree of contact and urbanization, as well as on more subtle cultural processes. There are also important distinctions within communities between proficient speakers and persons losing fluency and resorting to code-mixing due to the lack of language use. Once the language of empire and colony, one of the dominant languages in the entire pre-Columbian world, spoken by cosmopolitan elites and traders and widely used as a *lingua franca*, Nahuatl is today on the verge of becoming an endangered minority language. The numbers of speakers fall drastically every decade due to catastrophic educational and language policies, economic challenges as well as widespread practices of discrimination toward native speakers. And these adverse tendencies are exacerbated by current globalization processes and educational policies.

With the end of the Mexican War of Independence in 1821 the Spanish Imperial infrastructure that employed Nahuatl alphabetic writing as an official medium for documentation and communication disappeared. The new succession of governments did not make important advances toward integrating indigenous communities into national life, and writing ceased to link Nahua people and their communities to each other within and between regions where the language was spoken. As a result, these communities became more and more isolated from each other and the differences between regional linguistic variants increased. They also remained largely isolated from the rest of Mexican society. During this period, with the exception of a set of *ordenanzas* issued by the government of Emperor Maximilian I (Maximiliano de Hapsburgo) and the works of Faustino Galicia Chimalpopoca (1854, 1859, 1869 and 1870), Nahuatl writing became very scarce, and did not reappear in force until the second half of the twentieth century. At this time, a number of factors including economic integration, the extension of public education and the spread of communication media initiated a renewal of intense contact resulting in a steady loss of native speakers and the progression of Nahuatl toward endangered language status. Although the extinction and rapid fall of speakers are a threat to all current Nahua-speaking communities, the situation of particular variants and groups of speakers varies.

In addition, there are serious discrepancies between existing classifications and attempts to reconstruct the historical development and mutual relationships between variants of older and modern Nahuatl. The first classification covering close to the full geographical area where Nahuatl is spoken was proposed by Juan Hasler, who divided the area into four dialects: Eastern, Northern, Central, and Western (Hasler 1958, 1961), but his definition of dialects was criticized for not having been based on extensive and coherent linguistic data (Canger 1988: 39). Other scholars, such as Yolanda Lastra de Suarez, have emphasized the fact that the lack of data constitutes an obstacle to positing a historical classification; even so, they maintain that there is a basic division between Central and Gulf Coast dialects (Lastra de Suarez 1974). Later on, in her important work “Las áreas dialectales del náhuatl moderno” (1986) Lastra de Suarez analyzed and compared numerous phonological, morphological and lexical traits of the varieties of modern Nahuatl, proposing to distinguish four areas: Center, La Huasteca, Western Periphery and Eastern Periphery. Una Canger prefers to make a fundamental distinction between Central and Peripheral groups of Nahuatl, the latter being defined simply by their lack of a number of descriptive features present in Central varieties. Central groups would embrace dialects which share many important features spoken in the Valley of Mexico, Northern and Central Puebla, Morelos, and Tlaxcala. Huastecan and Central Guerrero Nahuatl are also classified as Central dialects, but possess features that are specific to the two regions they share with neighboring Peripheral variants. The latter include the Western Periphery, Northern Guerrero, Sierras de Puebla, Isthmus, and Pipil (Canger 1988: 45–59).
In the Ethnologue language cloud, several variants of Nahuatl are classified at grade 5 (developing): Central, Western and Eastern Huasteca, Southeastern Puebla, Northern Oaxaca, Western Durango and Mecayapan in southern Veracruz. In the EGIDS scale this status assumes that the language is “in vigorous use” and there is a standardized form of literature used by some of the speakers. The fact is, there is no widely accessible or commonly used literature in Nahuatl today. Written materials are limited to textbooks for the basic level school system and their orthography has not been standardized. Moreover, instruction in Nahuatl and its corresponding teaching materials form part of a school system geared toward overall instruction in Spanish, and it is present only in a portion of Nahuatl-speaking communities in the above-mentioned areas. The status of Nahuatl in Sierra Negra in Southern Puebla and in Orizaba (Veracruz), some communities in Northern Puebla, Morelos and Guerrero is classified as vigorous (6a), implying that it is used by all generations and in face-to-face communication. In fact, the situation varies from community to community and among different families in each community, because the number of passive speakers is growing rapidly. It is common to find fluently speaking generations between 20 and 40 years of age, who no longer use Nahuatl as their main language of communication inside or outside of the household, and passive speakers under the ages of 20 or 30. Much more common and widespread is the level of endangerment described as threatened (6b), in which native speakers diminish in spite of the fact that the language is spoken by all generations. According to Ethnologue, this is the case of Central Nahuatl spoken in Puebla and Tlaxcala, eastern Durango, eastern Central Mexico (Isthmus – Pajapan), Michoacan, and some parts of Morelos. However, this classification does not reflect today’s language situation in Nahuatl communities, which can be well illustrated by the example of Central Nahuatl. Except for a limited number of communities, where intergenerational transmission is intact but subject to widespread bilingualism and an entirely Spanish school system (e.g. San Miguel Canoa, San Isidro de Buensuceso, and Santa María Zoyatla in the municipality of Tepeojuma, all in the state of Puebla), the large number of passive speakers in the generations under 50 and 40 years of age threatens to totally disrupt language transmission. This is a prevailing phenomenon today in the region, and it corresponds with the disappearing status described by Grenoble & Whaley (2006: 18) as an observable shift towards Spanish in native communities (where it began to replace Nahuatl in a greater percentage of homes) and an overall decrease in the proportion of intergenerational transfer. Classified as shifting by Ethnologue (7 – with middle-aged adults still using the language, but lacking intergenerational transmission) are Huaxcalteca Nahuatl, the variants spoken in the Temazcaltepec and Coatepec regions in the State of Mexico, as well as that of the Ometepec region in Guerrero. However, this classification actually reflects the current situation of numerous Nahuatl-speaking communities across Mexico, where the speaker base is constantly shrinking. This status corresponds to the moribund level of Grenoble & Whaley (2006: 18). It is surprising that very few communities are classified in the Ethnologue cloud with the moribund (8a) and nearly extinct (8b) status, categories signifying that only those generations beginning with grandparents use the language.

Summing up, in many cases the Ethnologue classifications need updating and verification because they do not likely reflect the sudden decrease in language use that has occurred within the last two decades. Furthermore, the parameters of an individual community can be very different from those of neighboring native communities in the region and may not be representative even on a microregional scale. More and more, Nahuatl communities today are becoming reduced islands of speakers, with different degrees of transmission. Although the estimated number of speakers may seem high, all essential criteria of endan-
germent are met by today’s variants of the language. These include the low percentage and proportion of speakers within a population, the varying and quickly dwindling extent of language transmission, the loss of functions in language use and its failure to expand to new domains of modern life and media, as well as the proliferation of negative attitudes toward the language (Brenzinger 2007: ix).

The single most crucial factor contributing to language loss is the decrease in intergenerational language transmission within communities at the level of home / family / neighborhood, widely recognized as the key element of language maintenance and survival (e.g. Fishman’s Graded Intergenerational Dislocation Scale – Fishman 1991: 395). This is a widespread phenomenon in the Nahua-speaking world today, and it is aggravated by a lack of adequate educational support and adverse language ideology, both inside and outside the communities. Another essential factor contributing to increasing endangerment, that should not be underestimated, is the fact that materials for language education and literacy are scarce or non-existing. This situation becomes even more problematic due to the lack of consensus regarding standardized orthography and a common standard form for writing the language. Thus, Nahuatl can no longer be considered a “stable” language (Krauss 2007: 4–5), where the home domain remains essential and strong, not affected by the use of another language at school or work. Different sources of pressure, including all forms of discrimination and negative ideology have caused parents to cease speaking their native language, resulting in the destabilization of the linguistic environment at home. The failure of Nahuatl to expand into use with new technologies poses an additional threat. Functional domain differentiation between Nahuatl and dominant Spanish, the latter primarily associated with educational and labor opportunities, is one of the crucial factors of language shift. Nahuatl is seen as a language of limited potential, spoken only by elders, and lacking any utility in the modern world.

It is not infrequent to find communities that meet the criteria of “endangered” and “severely endangered”, as defined by Krauss; that is, when Nahuatl-speaking parents permit their children to respond in Spanish. And in those communities where the youngest speakers are middle aged or belong to the generation of grandparents, parents cannot teach the language to their children. In fact, many members of native communities, can be classified as “ghost speakers” (Grinevald & Bert 2011: 51), who conspicuously deny any knowledge of Nahuatl in spite of evidence that they do have some level of competence. This happens both inside the community space and in interactions with outsiders, attesting to the prevalence of negative attitudes toward the language, and manifests itself in the form of people refusing to be identified as an indigenous speaker. A strong foundation for this adverse language ideology was formed after the Mexican Revolution, when intellectuals began to forge a new national identity, based in part on pride in a mythologized version of Mexico’s indigenous past. However, modern indigenous people, considered culturally backward and an obstacle to modernization, needed to be Hispanized, and their languages needed to be eliminated as quickly as possible. Thus, Mexican multilingualism can be characterized as a conflictive, substitutive and diglossic bilingualism, in which bilingualism/multilingualism is considered a historical stage leading to a new monolingualism (Flores Farfán 2002: 228).

School-based programs in multilingual countries include examples of the most successful cases of language revitalization programs. However, the potential of these strategies has not been applied to Nahuatl and other indigenous languages in Mexico. The use of public education as a focused instrument of Hispanization began in 1964 when the first generation
of bilingual educators was recruited by the federal Secretaría de Educación Pública for the purpose of assuring that indigenous school children gain literacy in Spanish. Nahuatl and other indigenous languages lack the rich and crucial repository of works found in libraries and online in Spanish and in all the major world languages. Indigenous people must have access to these kinds of materials in their own language, if they are to be truly educated for successful participation in an ever more global and multicultural society. A truly multilingual program of education would not seek to replace indigenous languages with Spanish or English; rather it would cultivate in children the unique perspective and cognitive tools available to them through their native language, and complement this with additional perspectives and tools from other languages. Mexican bilingual education grew after its creation and continues to expand to this day, but its goal of replacing indigenous languages with Spanish has not changed. Mexican elementary education is highly centralized, with materials and curricula produced almost exclusively by the federal Secretary of Education. Traditionally, individual teachers do not participate in curriculum development, but are trained as technicians who implement ready-made materials. Bilingual teacher preparation takes place, for the most part, in Spanish, and they are not encouraged to participate in innovating curriculum development and research in the language spoken by their students. This is particularly harmful for indigenous languages, considering that textbooks are only produced for a limited amount of their variants. When these are distributed in communities that speak another variant, they are often rejected. Further, the sons and daughters of bilingual teachers, most of whom are raised speaking Spanish, often inherit their parent’s job upon retirement. And new bilingual teachers are routinely given jobs in communities that speak variants and even languages different from their own. It is not uncommon for children to be encouraged to stop speaking their native language at school, while teachers advise parents to speak only Spanish to their children.

In fact, the situation of native-speaking children in Mexico and their Spanish-language proficiency, closely parallels the situation of Native American children who are likely to be stigmatized as “limited English proficient”. Whereas the United States has motivated bold new strategies for indigenous schooling that emphasize immersion in the heritage language and community-based planning (McCarty 2003: 147–158), this approach is virtually absent in Mexico. Immersion schools started to develop in the United States in the 1980s, based on the principle that the dominant language of the society should only be worked with in school as a foreign language (Hinton 2011: 298). This kind of immersion based program could begin to be implemented in Nahua-speaking communities, using the model of preschool language nests, in which the fluent speaking grandparent generation, often the last fully proficient generation of native speakers, would take care of young children using only the indigenous language.

Although nidos de lengua were established in Mexico, especially in Oaxaca, beginning in 2008, with at least ten language nests in existence by late 2009, serving the Mixtec, Zapotec, and Cuicatec languages (Meyer & Soberanes Bojórquez 2009), their small scale and limited distribution cannot meet growing challenges. While many adverse language attitudes prevail at the community level, more subtle forms of discrimination take place when the students enter junior high, high school and college. During the presidency of Vicente Fox (2000–2006) the federal government abandoned an initial proposal designed to promote spaces for indigenous education at public universities. Instead, a new system of intercultural universities was created. However, most of these underfunded institutions do
no more than offer traditional careers in Spanish to a largely indigenous student population. Curiously absent at all Mexican universities, including the Universidad Nacional Autónoma de México, with its flagship program “Mexico Nación Multicultural”, is the one mode of activity with which these institutions could trigger a national movement of indigenous linguistic and cultural revitalization: The large scale practice of curriculum development, teaching and research done entirely within an indigenous language.

2003 saw the creation of a federal law designed to protect the linguistic rights of Mexican indigenous people. Simultaneously, federal education legislation was modified, guaranteeing, at least in theory, speakers of indigenous languages access to basic education in their native tongue. Founded also at that time was the National Institute of Indigenous Languages (INALI), a state institution charged with overseeing the implementation of the law, within a context of national governmental decentralization. Its primary function was to promote and coordinate the foundation of indigenous language institutes, legislation, and most importantly, statutes providing means of enforcement of this legislation at the level of the individual states. To date INALI has carried out linguistic research and published a national catalogue of languages; it has produced numerous works in and on indigenous languages, including multiple translations of the Mexican constitution and other governmental documents; it has created norms for the preparation and licensing of translators and interpreters; and it has provided limited legal advice in individual cases of linguistic discrimination. However, state legislation in the area of linguistic rights is practically non-existent. Only a few of the thirty-one states have created indigenous language institutes. There are serious impediments to the implementation and execution of laws related to linguistic rights, given that the perpetuation of colonial attitudes is common among lawmakers (Zimmermann 2011: 22–23). INALI has not undertaken or sponsored concrete programs of massive language revitalization and it has been silent in regard to certain key issues: in spite of the aforementioned reform of national education legislation, the majority of native speakers of indigenous languages still do not have access to basic education in their native tongue; the implementation of national standardized testing (ENLACE and EXANI/EGEL) clearly discriminates against non-native speakers of Spanish.

Today in Mexico, the pervasive ideology shared by indigenous and non-indigenous people alike and crossing all professions and walks of like, contends that native languages are “dialects”, and cannot be considered languages such as Spanish and English. Nahua communities suffer from what has been called a social dislocation stemming from their lack of prestige and power, as well as from a closely related cultural dislocation, which results from modernization and globalization (Grenoble 2011: 34). Processes of urbanization linked to social and cultural dislocation and an increasing use of a national language, as well as migration to larger towns and to the United States, usually result in a complete disruption of language transmission. While government agencies, the educational system, and the mass media all participate today to some degree in the process of overall Hispanization, many independent individuals and organizations, such as the Escritores en Lenguas

---

8 National Autonomous University of Mexico
Indígenas, A.C.\textsuperscript{13} or the Fundación Cultural Macuilxochitl\textsuperscript{14} struggle to promote cultural and linguistic plurality.

4. Our methodology and activities. In a nutshell, our methodology consists of promoting native speakers of Nahuatl to the role of protagonist in academic and revitalization activities related to their language and culture. As a rule, Mexican educational institutions do not hire indigenous people to teach courses and conduct research related to their language and culture. Bilingual elementary school teachers are the exception. However, they receive their training in Spanish and do not participate in any aspect of curriculum development, and the few courses in indigenous languages taught at the secondary and university levels are not based on modern second-language instructional methodology. And indeed, the carrying out of “activist documentation”, the development of language revitalization methods, and the production of educational materials by native speakers themselves has begun to be seen as a powerful alternative to traditional approaches to the problem (Flores Farfán & Ramallo 2010: 13–14). The Instituto de Docencia e Investigación Etnológica de Zacatecas (IDIEZ) and the Faculty of “Artes Liberales” of the University of Warsaw have been working with Nahua immigrants from the Huasteca region who are studying at the Universidad Autónoma de Zacatecas. Offering an alternative to the general function of the Mexican university as the last step in the educational process of Hispanization, these students are provided with a monolingual space in which to continue practicing and developing their language and culture. Parallel to the careers they study at the university, they are trained to teach Nahuatl and they actively collaborate with Western academics in many types of research projects. Mexican education denies Nahua students access to the pre-Hispanic codices and colonial alphabetic texts written by their ancestors, and as a rule, discourages independent thinking. We provide young indigenous scholars with the opportunity to study these materials, as well as works written by contemporary authors, and they are encouraged for the first time in their academic lives to formulate and express their own opinions. And they do this, also for the first time since their childhood, in their own language. Additionally, we have begun to collaborate with other institutions in Mexico, such the Autonomous University of Tlaxcala and members of Nahua communities in Puebla and Tlaxcala where the need for revitalization is particularly urgent. In January of 2014, we sponsored an activity in which indigenous high school students from Zoyatla, Puebla, studied colonial manuscripts in Nahuatl alongside Mexican and foreign students and researchers.

Research consistently shows, on the one hand, that instruction in a child’s native language and additive bilingualism provide the best foundation for future academic achievement and, on the other hand, that submersion in the dominant language with no linguistic support for the second language learner and subtractive bilingual instruction can impact negatively on self-esteem and achievement (Austin & Sallabank 2010: 10). Mexico employs the latter two models for the basic education of its native population. Indigenous children, then, enter school and immediately cease to employ and develop their language as a tool for critical and creative thinking. Those who reach the university have accepted the idea that the value of their “mother tongue” is restricted to its function as a vehicle for practical and affective communication when they periodically return home to visit their families.

\textsuperscript{13} http://www.nacionmulticultural.unam.mx/eliac/menu/01quienes.html (10 December, 2015).

\textsuperscript{14} http://fc-macuilxochitl.blogspot.mx (10 December, 2015).
We have been able to develop a method for jump-starting these students back into the use of their language as an academic tool and allowing them to generate knowledge from the unique perspective that the cognitive structures of their language provide. Our aim of providing a safe space for language use and the development of indigenous research methodology stems from the importance we attach to academic expression in a native language. For almost a decade now, we have been working on a monolingual dictionary of Modern Huastecan Nahuatl. When a student begins to work with us, he or she is immediately given the task of formulating the monolingual definitions and example sentences that integrate the dictionary entries. Defining words is a basic cognitive exercise used in many elementary curricula around the world. But these university students have never done it in their native Nahuatl and they employ a unique process for accomplishing the task. In its initial phase, they experience headaches, and they all resort to a mental strategy they have developed over years of coping with a national educational system that is foreign to them: instead of resorting directly to decontextualized abstract thinking in order to produce a definition, they transport themselves mentally back into their homes and communities and imagine themselves in a situation in which the language they are seeking may occur. Collaborative thinking is natural for them, and they begin to gravitate away from the intellectual individualization of their formal education. James Lockhart, the ethnohistorian whose New Philology introduced the possibility of studying Mexican colonial society through sources written in Nahuatl, once said that when reading the early manuscripts written in this language he could sense the joy and excitement of the indigenous writers who had recently gained access to the new tool of alphabetic script. This is the emotion that permeates the working atmosphere in our programs, where native speakers of Nahuatl discover that they can use their language to reason and create new ideas.

Stimulating thinking from within Nahuatl is a point of departure for our research, teaching and revitalization activities. Nahuatl is a highly contextualized language. There are no infinitive forms of verbs, for example, and our researchers have created formulae for definitions that reflect this inherent specificity. The Nahua mind is not content to describe an action. It is important to specify which kinds of subjects, for example, can perform it. And since in Modern Huastecan Nahuatl there is no one term meaning “animal”, verb definitions must indicate whether deities, humans, wild animals, domestic animals, flora and/or grammatically inanimate entities can function as the subject of an action. Even this is problematic because mountains, celestial bodies, springs, land, and other natural phenomena are considered more animate than humans. The content of definitions is at times very different from that of a Spanish or English dictionary. For example, the second definition of *ahcuexoa* “to sneeze” is *Macehualli chicahuac quiquisitia ihyo pan iyacatzol quemman quihualillamiqui ce acahya* “A person expels air violently from his/her nose after feeling the sensation of missing someone who is far away.” Native speakers who begin working with us must thus reflect on and reevaluate what Mexican society and Christianity have taught them about the superstitious and valueless nature of their culture.

Creating grammatical and scientific terminology from within Nahuatl is another important and exciting activity we carry out. The grammatical manuals produced for use by indigenous bilingual elementary school teachers simply translate linguistic concepts and terminology from Spanish. So noun, *sustantivo* or *nombre* in Spanish, is rendered *tocayotl* “name” in Nahuatl. However, a Nahuatl noun does not have the same structure as its counterparts in English or Spanish, which do no more than provide a label for an entity. Nahuatl nouns consist of a nucleus surrounded by obligatory subject affixes and optional possessive affixes. In other words, a Nahuatl noun is actually a sentence. So, for exam-
ple, *nicihuatl*, with its first person singular subject prefix *ni–*, its root *–cihua–* (“woman”) and its singular absolutive suffix *–tl*, means “I am a woman”. Therefore a Nahuatl noun involves the process of providing a subject with a name, and for this reason we use the verb *tocaxtia* “to name someone or something” as the basis for creating a neologism, *tlatocaxtilitzli* “noun” or more literally “the process of providing a subject with a name”. One of the most challenging aspects of the monolingual dictionary project was defining the letters of the alphabet, a task that Joe Campbell, the authority in Nahuatl morphology, helped us accomplish during the IDIEZ Summer program in 2010. Obviously this implied describing the vocal processes involved in the production of the allophones associated with each letter. Some words, such as *copactli* “soft palate” were included in pre-Hispanic anatomical vocabulary and are attested in the Florentine Codex, a sixteenth-century encyclopedia of Nahua life. However, we had to create many neologisms. Nahuatl morphology is pretty transparent, and once one is familiar with the elements and the rules for their combination, creating new words that make sense is not that difficult. Our neologism for “vocal folds”, *totozaamayo*, is composed of the root *amatl* “paper(s)” with a suffix *–yo*, one of the functions of which is to use metaphor to create new meanings that many times are associated with the body. So *amayo* can be understood to mean “a part of the body that resembles a sheet or sheets of paper”. *Tozca–*, the combining form of *tozquitl* “voice”, is incorporated onto this, producing *totozaamayo* “a part of the body, associated with voice, that resembles a sheet or sheets of paper”. We then add the first person plural possessive prefix, *to–*, because in Nahuatl all body parts are possessed. So *totozaamayo* is “our vocal folds”. These simple kinds of activities empower and liberate native speakers by returning intellectual tools that were stripped from them during their formal education.

Research collaboration between Western investigators and speakers of indigenous languages is also crucial to our methodology. Traditionally, Western ethnographic researchers have incorporated native speakers of indigenous languages into their work as informants whose role in the research process is limited to the passive transfer of raw linguistic data. A firm boundary is drawn between the informant (conceived of as a possessor of native cultural knowledge) and the anthropologist (the only participant capable of understanding and interpreting this knowledge at an academic level). We deconstruct this boundary by assigning an active role to students and researchers who are members of the communities under investigation. One of our goals is to train indigenous students to become independent teachers, researchers, and active collaborators with Western scholars. Given the continuing discrimination against indigenous people in Mexico, the fact that the Nahuatl scholars who collaborate with us are able to continue their graduate studies, including the preparation of Ph.D. dissertations in Europe under the auspices of the European Research Council funded project mentioned earlier, will be vital. In our combined projects, we do not “read over the shoulders” of the natives (Geertz 1973: 452), but strive to combine inside and outside perspectives in ways that are new to existing scholarship. In this way, we are also able to overcome certain limitations common to revitalization projects, such as the failure to recognize communities’ actual needs or the undervaluing of local attitudes towards language revitalization.

Individuals who wish to design and implement revitalization projects need to have a solid background in linguistics, but they must also have training in the theory and method-

---

15 The need to provide indigenous people with linguistic training, so that language documentation can take place from within their communities, taking into account local concerns and language maintenance, has already been emphasized (Grinevald 2007: 77–78).

16 See note 2.
ology of second language instruction and learning processes (Hinton 2011: 309). To meet this aim, we are planning to open a monolingual and international master’s degree program in Nahuatl language and culture in order to formally pilot the implementation of our teaching and research methodologies in the Mexican educational system. The majority of our students will be Nahuatl speaking bilingual elementary school teachers who will then use what they have learned and co-developed in our program to begin to make major changes in the way Mexico educates indigenous children. However, we also expect to have non-indigenous students and professors from Mexico, Europe and the United States participating in the program. Graduates will be able to teach, conduct research and work as activists in the implementation of concrete revitalization projects. The need to decolonize research methodologies and create spaces for developing and practicing indigenous methodologies has recently been emphasized, especially by native scholars in different parts of the world (e.g. Chilisa 2012, Kovach 2009, Tuhivai Smith 2012). For this reason, the new master’s program will establish a foundation and point of departure for indigenous students and researchers to develop methodologies from within their own language, processes for creating and transmitting knowledge, and traditional practices and concepts employed in their communities. Graduates will then apply and continue to develop these methodologies in their own teaching, research and revitalization endeavors.

5. Restoring historical identity and the culture of literacy in Nahuatl. By the fourth decade of the sixteenth century, the Nahuas had mastered alphabetic writing and initiated a written tradition in their own language. Beginning in the 1540s there was an explosion in the production of multiple written genres, including local municipal documentation, land sales, legal cases, petitions, personal correspondence, chronicles, theatre, dictionaries, grammars, and religious works. This writing expanded quickly across the core area of Nahuatl culture and by the beginning of the nineteenth century constituted the largest corpus of indigenous language texts recorded anywhere in America. Production tapered off after Mexican Independence and was not resumed in force until the 1970s, approximately ten years after the first generation of Mexican bilingual teachers was recruited. At this time, we begin to see works of literature in poetry, narrative, theatre, and essay published in indigenous languages. Many of these writers, such as Natalio Hernández, whose first books were authored under a pseudonym for obvious reasons, emerged from the ranks of these teachers who had become disillusioned with the system.

At no time during this period of almost five hundred years has literacy or the practice of reading and producing literary works been widespread among the Nahuas. However, the tradition and its corpus is a fact, and we believe that unless it is reactivated, spread and developed, no attempt at revitalization will be successful. Our strategy for carrying this out includes emphasizing the continuity between older and modern Nahuatl language and culture; encouraging indigenous people to create monolingual spaces in their communities and educational institutions in order to read and discuss works that their ancestors and contemporaries have written, and to create works of their own; and finally, publishing, through the University of Warsaw’s and IDIEZ’s monolingual Totlahtol series, older and contemporary written works, as well as reference materials, in standardized orthography. Refugio Nava Nava’s book of children’s literature, Malintzin itlahtol (2013), written in Tlaxcalan Nahuatl, was the first work to be published in this series, quickly expanding with other genres by indigenous authors (Zapoteco Sideño 2014, De la Cruz Cruz 2015, Nava Nava & Cuahutle Bautista 2015, Bueno Bravo et al. 2015).
Bridging divides

Nahuatl language and culture has evolved in a continuous and uninterrupted fashion from pre-Hispanic times to the present. However, this fact is obscured by the structure of Western academics and by Mexican National ideology. Linguists and anthropologists study the culture of people who are currently alive. Historians and archaeologists study the culture of people who are dead. Works produced from these perspectives give the impression, on the one hand, that present day indigenous culture has no past, and on the other hand, that past culture has nothing to do with the lives of indigenous people today. This illusion is appropriated into Mexican National ideology, which states that the nation has roots in the great indigenous civilizations of the Mayas and the Aztecs, but that indigenous people today are culturally backward and constitute an obstacle to progress.

We believe that it is important to stimulate indigenous students to participate in activities, carried out in Nahuatl, that focus on the reading and commentary of texts written by their ancestors. This is essential if native people are to reconstruct their historical identity and be able to promote the survival and growth of their culture. We are implementing this by involving indigenous students and researchers at IDIEZ as well as Nahuatl collaborators from Puebla and Tlaxcala in the reading and analysis of older Nahuatl manuscripts and modern Nahuatl texts through participation in our major research projects. We began these activities in 2014 at the Winter Nahuatl Institute in Cholula and continued them at the first Nahuatl Document Analysis Workshop (XVI-XVIII Centuries) for Native Speakers held in the Mexican National Archives from August 19 to 21, 2015. Thirty speakers of Nahuatl from diverse communities in Mexico City and the states of Mexico, Puebla, Tlaxcala, Guerrero, Oaxaca and Veracruz took part in the workshop activities, which were conducted entirely in Nahuatl. The participants had the opportunity to personally examine three original manuscripts and then worked together to transcribe, read and analyze the materials in a setting charged with emotion, because these documents constitute a key component in the formation of the historical identity of these modern heirs to the ancient Nahu culture and tradition. The workshop provided a window into the literature written by their ancestors, the existence of which had been previously unknown to many of the participants. Such activities allow indigenous people to directly experience the fundamental relation of continuity between older and modern Nahuatl language and culture.

Crucial to this enterprise is the restitution of literacy in the native language and the unification of its orthography in close relationship to the older tradition of writing in Nahuatl. These goals encounter several major obstacles in Mexico today. There are currently two different types of orthographies used for modern Nahuatl. One group has developed independently of the earlier colonial conventions, grounding itself in linguistic considerations that seek to rationalize spelling: digraphs originating in Spanish orthography are eliminated whenever possible; glottal stops and vocalic length are represented. These systems confuse the concept of everyday writing with that of phonetic documentation and constitute an obstacle to language revitalization and native literacy in several important ways. First, no attempt has been made to standardize any of these systems by means of monolingual dictionaries that could codify the spelling of all words; and this lack of standardization prohibits native speakers from using writing to communicate across variants. Second, their attempt to distance themselves from the earlier writing system widens the artificial academic division between older and modern Nahuatl language and culture. But perhaps more

---

17 As Lüpke correctly pointed out, "[i]t is widely assumed by linguists that the basis of the ideal orthography is phonemic. If this was the case, the main difference between a phonemic transcription and an orthography would be the inventory of symbols used" (2011: 331).
importantly, it discourages indigenous people from reading and studying the great corpus of older works that constitute the written cultural legacy of the Nahua civilization.

Four factors have contributed to a general feeling of animosity toward older writing conventions that exists in Mexico today. The modern resurgence of Nahuatl writing actually began in the middle of the twentieth century when Protestant missionary linguists, working under the umbrella of the Summer Institute of Linguistics and later in cooperation with the Mexican Secretaría de Educación Pública, began producing bible translations in various indigenous languages. Missionary and governmental goals coincided for a time, for each group believed that indigenous people needed to be redeemed, on the one hand from their pagan religion and on the other hand from their backward culture. Older spelling conventions were considered bridges to the past that needed to be burned. Many people see the use of the modern linguistic conventions as a political statement in favor of the independence of indigenous languages with respect to the Mexican hegemonic culture of Hispanization. And finally, academics who work with older Nahuatl have also contributed to the problem, alienating indigenous people by stating that their modern culture and language is no more than a deformed and pauperized version of the glorious civilization of the past. Two other schools of thought use what can be called enriched traditional orthographies. Both are based on older writing systems and include modifications, such as the use of the <h> to represent the glottal stop or aspiration. Members of the Asociación de Escritores Indígenas, A.C. base their system on Alonso de Molina’s dictionary. And we use the enriched traditional orthography based on Horacio Carochi’s grammar and modified by Richard Andrews in his Introduction to Classical Nahuatl (2003), Frances Karttunen in her Analytical Dictionary of Nahuatl (1992), and Joe Campbell and Frances Karttunen in their Foundation Course in Nahuatl Grammar (1989). We are also preparing a monolingual dictionary of Modern Huastecan Nahuatl in order to codify this orthography. It will be published in 2016 as the first reference work within the Totlahtol series, one of the purposes of which is to extend this orthography to other variants.

6. COMBINING AND SHARING EXPERIENCE. The success of Nahuatl revitalization efforts in the coming years will depend to a large degree on the ability of native speakers from different regions of Mexico to communicate and collaborate with each other in the planning and implementation of projects for the development of their language and culture. The lack of contact between different isolated Nahua communities makes them even more susceptible to rapid language shift (Flores Farfán 2002: 229). However, international cooperation will also be needed if indigenous people are to overcome the general tendency toward Hispanization in Mexican society.

Until recently, geographic distance and the differences between linguistic variants constituted what was considered an insurmountable barrier to the possibility of interregional communication. However, in December of 2011, as part of a research project funded by the US National Endowment for the Humanities, IDIEZ brought together twenty native

---

18 A relatively recent justification for the use of modern orthographies can be found in Anuschka van’t Hooft’s The Ways of the Water. A Reconstruction of Huastecan Nahua Society Through Its Oral Tradition (Hooft 2007: 11-12).
19 Some of the members of the Asociación de Escritores Indígenas, A.C. have included Librado Silva†, Francisco Morales, and Natalio Hernández, all of which have been participants in the Seminario de Cultura Náhuatl that Miguel León Portilla has directed for over fifty years at the National Autónomous University of Mexico.
20 The project An Online Nahuatl (nci, nhe, nhw) Lexical Database: Bridging Past, Present, and Future Speakers was directed by Dr. Stephanie Wood from the University of Oregon from 2009 to 2012 (http://whp.uoregon.edu/dictionaries/nahuatl, 10 December, 2015).
speakers representing approximately ten variants of Nahuatl for a five-day workshop in Zacatecas. A second Interdialectal Encounter of Nahuatl, financed by the Mexican National Commission for the Development of Indigenous Peoples (CDI) and organized jointly by IDIEZ and the University of Warsaw, was recently held over the weekend of January 18 and 19 of 2014 in the city of Cholula with the participation of sixty native speakers and thirteen non-native speakers. Both events were recorded and broadcast by XECARH “The Voice of the Hñahñu People”, a radio station affiliated with the CDI. Normally in Mexico when native speakers of indigenous languages are involved in discussions and conferences, the agenda is predetermined by the organizers, who are usually non-indigenous administrators of government cultural institutions. In our two Interdialectal Encounters, the proposed topics of discussion were reviewed and ratified or modified by the indigenous participants at the beginning of each event, and the actual discussions were held monolingually in Nahuatl. Generating a monolingual environment was problematic at the beginning of the 2011 event and continued to be an issue for the new participants of the 2014 event. Indigenous children learn quickly that Mexican society will not tolerate the use of their language outside of their villages, and in most cases, in their local schools. As a rule, native speakers of indigenous languages converse with each other in Spanish outside of their community. And if they must use their language in a public situation, they will immediately translate what they have said into Spanish. Not surprisingly, this behavior was replicated by many of the indigenous participants at the beginning of each Interdialectal Encounter, probably reinforced by the belief that speakers of different variants of Nahuatl would not be able to understand each other. In fact, before the 2011 event, which probably constituted the first time in hundreds of years that speakers of multiple variants of Nahuatl had had the opportunity to converse monolingually with each other, it is probable that no one really knew if interdialectal communication would be possible. In both events as well as during the AGN workshop in 2015, it immediately became apparent that a high enough degree of intelligibility existed to permit fluid and animated monolingual discussions on a diverse array of topics, including identity, revitalization, rituals and local festivals, ways of greeting, education, immigration, grammatical terminology, linguistic policy, migration, intergenerational language transmission, gender issues, and interculturality. Perhaps most important is the shift in attitude that occurs among the participants as the discussions progress and they are able to experience interdialectal communication for themselves. An initial environment of timidity bordering on distrust gives way to an atmosphere of joy and solidarity, as well as the desire to continue the communication after the end of the event. We will also soon begin to organize live interdialectal discussions using videoconferencing platforms, including members of Nahua immigrant communities in the United States.

We have also started to spearhead the formation of an international consortium of institutions of higher education to foment the teaching, research and revitalization of Nahuatl language and culture. Instruction in indigenous and other minority languages, both for native and non-native speakers, is few and far between at universities in Mexico, the United States and Europe. Reductions in public spending and the general tendency to reduce course offerings in the humanities do not paint a bright future for these areas of instruction. Universities were designed to be self-sufficient islands of knowledge, and to this day the majority of them resist to a more or lesser degree sharing costs and human resources, not to mention the new distance instruction platforms that would make this kind of efficient collaboration possible. We promote these strategies of inter-institutional cooperation in order

---

to make Nahuatl instruction available to anyone in the world who wishes it, acknowledging the importance of neo-speakers for research and revitalization projects.

The division between the roles of the investigator and the informant that constitutes one of the major methodological underpinnings of ethnographic research, hinders the production of knowledge. A few years ago, IDIEZ added the component of individual tutoring to its Summer Program curriculum in colonial and modern Nahuatl. Foreign students were required to bring a research topic with them to the program and work on it for an hour per day during the six-week program with a native speaking tutor. The students were invited to consider these sessions as an exercise in mutual instruction. They would explain to their tutors how they had set up their research project so that the tutor would not be restricted to simply supplying information, but instead would be able to participate, along with the student/researcher, in its analysis and interpretation. Every year, more and more of our students involve their tutors in the preparation of their theses and dissertations, as well as long-term research projects. And they are producing a new kind of knowledge generated by combining the perspectives of Western science and the specifically indigenous ways of collecting, organizing and interpreting data. The long-standing colonial Nahuatl teaching program carried out at the University of Warsaw since 2000 has been enriched since 2012 by a course in modern Nahuatl taught by native speakers, making it the only permanent full academic year Nahuatl program of its kind. Yet another complementary endeavor is our revitalization website dealing with three endangered languages: Nahuatl in Mexico and Lemko and Wymysiöeryś in Poland. Its three domains of research, culture (including literature), and education describe, document, and recreate the universe of each of the endangered languages. And they are all presented in monolingual interfaces in each of the three endangered languages, plus English, Polish and Spanish. The website has been designed, on the one hand, as a space available for writers in Nahuatl, Lemko and Wymysiöeryś to publish their works, and on the other hand, as a resource repository for scholars and students working on those languages and their communities. Its target user groups include speakers of endangered languages, students, and scholars.

It is probably safe to say that there has never been a successful indigenous language revitalization project in Mexico. Racism is a structural aspect of Mexican society that is not recognized and addressed in the public forum: unhindered by criticism, it generates pervasive discrimination against indigenous people that cannot be countered by limited and isolated revitalization efforts. We believe that an international consortium of committed institutions and individuals can provide the independent funding, experience, creative theories and strategies, and prestige that may catalyze these projects and assure their success. Even more importantly, however, these initiatives should inspire and support but also provide autonomous space for community-based programs.

7. BENEFITS OF LANGUAGE REVITALIZATION FOR SPEAKERS AND THEIR COMMUNITIES. “At the dawn of the twenty-first century, the world’s linguistic and cultural diversity is under assault by the forces of globalisation – cultural, economic and political forces that work to standardize and homogenize, even as they stratify and marginalize” (McCarty 2003: 147). The processes of globalization and homogenization are exacerbated by the still pervasive worldwide belief that the establishment a national language and culture is a fundamental requirement for political stability. Therefore, as in Mexico, most bilingual education systems exist solely for the purpose of humanely transitioning speakers
of minority languages to the monolingual use of the dominant national tongue. Language and educational policy-makers ignore the fact that languages reflect the most fundamental human experiences, while “their decline will result in the irrecoverable loss of unique knowledge that is based on specific cultural and historical experience”, thus weakening considerably the ethnic and cultural identity of speech communities (Brenzinger 2007: ix).

But more importantly, they fail to understand how the cultivation of linguistic and cultural diversity benefits everyone. The human capacity for solving problems and creating new ideas, which today is the motor of economic production, is distributed in a form resembling that of a living, evolving mosaic within the structure of each different language. UNESCO emphasizes in the *Universal Declaration on Cultural Diversity* that creation is rooted in cultural tradition, but is nourished by contact and genuine dialogue with other cultures. Therefore, multicultural heritage fosters creativity in all its diversity because “in the face of present-day economic and technological change, opening up vast prospects for creation and innovation, particular attention must be paid to the diversity of the supply of creative work” (UNESCO 2002: 5).

This idea provides a strong rationale for language revitalization that should be grounded in the social, cultural and economical benefits of cross-cultural transfer, as well as the advantages of extending linguistic diversity. An important aspect of Whorfian thought associates the benefits of ethno-linguistic diversity with pan-human creativity, problem solving, and mutual cross-cultural acceptance (Fishman 1982), while differences in language structure result not only in differences in construal, but are also significant for the performance of nonlinguistic tasks and activities (Gumperz & Levinson 1996, Kövecses 2006). Revitalization should be aimed to strengthen what David Dalby (1999–2000) calls the linguasphere: the dynamic, evolving global structure of interacting and interdependent languages. The processes of globalization weaken communities’ ability to maintain, develop and use the unique tools of perception, reasoning, and creation coded within their languages. This tendency not only debilitates the cognitive resources with which they reproduce themselves, but also prohibits them from contributing to the enrichment of the dynamic, evolving global sum of these intellectual tools, which Dalby calls the logosphere. Drawing from the resources available in the logosphere depends on the vitality of the linguasphere, the diversity of the tongues that integrate it, and society’s attitudes regarding the value of multilingualism. The continuity of local knowledge, along with other aspects of empowerment of communities, is a key to sustainable development, which, in turn, is needed for preserving local ecosystems, and through them, the global ecosystem. Language maintenance is an essential part of these local and global ecosystems (Nettle & Romaine 2000: 176–177).

However, on the other hand, language planning based on the value of linguistic and cultural diversity is often an initiative of Western scholarship focusing on languages as abstract entities rather than communities of speakers with their realities and postulating what non-Westerners should be doing with their mother tongues (Coulmas 2013: 221). Therefore, the understanding and sharing of deeper benefits related to language maintenance by the community members themselves is in fact one of the fundamental challenges and preconditions of any revitalization project.

While language revitalization is crucial for local ecosystems and traditional resource management in the hands of specific communities, it can also be argued that the preservation and development of an endangered language offers benefits that extend far beyond its community of speakers. The unique perspective that is coded within its structure provides anyone who speaks the language with a distinct set of critical and creative tools that can be applied to solving the problems and enriching the lives of both native and non-native
speakers of a given language. Thus, revitalization strategies should embrace the creation of an important category of neo-speakers (Grinevald & Bert 2011: 51) who may include both younger generations of community members where language transmission had been broken and extra-community speakers, including activists, teachers, students and researchers. This approach is further strengthened by the most recent results of psycholinguistic research that demonstrate a strong correlation between multilingualism and enhanced non-verbal processes. It has been demonstrated that bilingual and multilingual children and adults have expanded cognitive potential, which manifests itself in greater flexibility and capacity for task-solving and in generally higher intellectual and social skills (Bialystok 1999, 2001, Bialystok & Martin 2004, Bialystok & Senman 2004, Bialystok et al. 2012, Costa et al. 2008, Kovács 2009). Current research also suggests that multilingualism enhances executive control, the brain’s capacity for staying on track in the selection, organization, and utilization of the data we need to solve problems and achieve goals. And it seems to be particularly important in inhibiting the interference of non-essential information in this process. Advantages for executive control can be seen already in 24-month-old children (Poulin-Dubois et al. 2011: 567–579). On the other hand, bilingualism also offers significant advantages to the elderly, protecting against cognitive decline and possibly delaying the onset of symptoms of dementia (Bialystok et al. 2012: 240–250). Thus, the use of more than one language, which in many cases simply consists of continuing with the “natural” multilingualism and cultural pluralism that has historically characterized many regions of the world, offers benefits for all the age groups of a given society. This exceeds its practical advantages in the normal processes of daily communication.

These findings seem to be backed up by direct outcomes of revitalization programs: it has been repeatedly shown that children coming out of strong immersion models always match or surpass their counterparts participating in the dominant-language programs, in both classroom performance and standardized testing (Hinton 2001: 298–299). It seems highly significant that the success among the Nawaho and Hawaiian immersion programs based on a bottom-up language planning that provided a means of empowerment for native teachers, children, and communities, can also be measured by the fact that its students demonstrated much better academic performance than those receiving school instruction in English (McCarty 2003: 151–157). In addition, recent research has shown a strong correlation between language loss, deterioration in indigenous health, symptoms associated with post-traumatic stress, and elevated suicide rates (e.g. Chandler & Lalonde 1998, McIvor et al. 2009, Ball & Moselle 2013). We believe dealing with health issues should become an essential part of integral revitalization programs.

Thus, the revitalization of Nahuatl and other languages should not be seen as an aim in itself. The strength and vitality of intellectual infrastructure, at the level of both local communities and the larger society, depends on preserving the local ecosystems which assure the quantity, the quality, and especially the diversity of the ideas the society can cultivate. The speaker of a minority language, then, is one of humanity’s most valuable assets, for this person perceives the world, its problems and its possibilities for development uniquely, and possesses a special set of cognitive tools for hypothesizing ways to transform it. Revitalization can catalyze the potential of native speakers of minority languages for making unique contributions to the enrichment of life in society as a whole. On the other hand, non-native speakers can also benefit from the acquisition of a rare language, and the application of its specific cognitive tools to their professional activities, including research work. In other words, “cultural diversity widens the range of options open to everyone; it is one of the roots of development, understood not simply in terms of economic growth, but
also as a means to achieve a more satisfactory intellectual, emotional, moral and spiritual existence” (UNESCO 2002: 4). We believe that this approach provides a rationale for creating integral, efficient strategies for revitalization programs of endangered languages, but it also unlocks an important general social and economic potential. Integral strategies of revitalization, education, and usage of endangered languages, if successfully applied, will generate enormous societal and cultural benefits in today’s world where cross-cultural transfer has become a powerful, but at the same time very challenging political, social and cultural phenomenon.

8. OUR APPROACH: SUMMARY AND CONCLUSIONS. It is essential to recognize that languages are based on local systems of knowledge and ways of life. Preserving these systems is crucial for the preservation of local languages (Nettle & Romaine 2000: 165, Bergier 2014). Traditional cultural and economic systems, however, can only survive and prosper where their members retain control over the resources, the mechanisms of knowledge transmission, and the key activities that constantly reconstitute, redefine, and integrate their communities. Factors such as the state economy, poverty, migration, Western-style education, and discriminating language policies have undermined, displaced, and even disintegrated the forms of organization and knowledge systems employed by the Nahuas and other indigenous groups in the Americas. While many of these processes are irreversible, new strategies are needed to counter and reverse widespread language shift. These may include supporting stable multilingualism in language-contact situations and searching for alternative, sustainable forms of development based on local languages and traditional knowledge. But this also poses special challenges for and demands heightened awareness from external institutions that engage in revitalization programs and community-based activities. While some of the proposed strategies may be applicable to European contexts, working with languages in the postcolonial circumstances of developing countries should address their specific challenges and highly sensitive ethical, ideological, cultural, economic, and psychological issues.

In order for our strategy to work, we need to foster collaboration across academic, social, and ideological boundaries, integrate theory, methodology, and program implementation, and efficiently combine grassroots and top-down approaches at the different interrelated levels (comp. Fishman 2001: 467). We must also address and bridge the gap between theory and practice in revitalization, i.e. between the study and planning of revitalization in academic circles and the implementation of concrete programs, be they community-based, educational or the direct result of governmental language policy. The urgent need to combine different levels of activities in the revitalization of the native languages of Latin America, including the pedagogical, public, and sociolinguistic spheres, has already been emphasized (Zimmermann 2011: 34–36), but it now needs to be put into practice. In the case of Nahuatl, we need to make essential contributions to linguistic knowledge by compiling extensive documentation of both a historical (archival texts) and a contemporary (audio and video recordings) nature. We use these collected resources to create, expand, and enrich dictionaries, grammars, and pedagogical materials, adapting the products of linguistic research for use in revitalization. Indeed, lexical and structural data from historical documentation can be reintroduced into modern language in order to enrich the linguistic tools available to native speakers. We also aim to strengthen the historical and cultural identity of native speakers by making research results available to members of speech communities and facilitating their access to the texts written by their ancestors throughout the colonial era. We further strive to raise the prestige of endangered languages in academic circles and
the broader society by promoting teaching and research, and by harnessing promotional campaigns in the mass media.

Education has an important place in our activities and is linked to our research and publication projects. It is extremely necessary to foster the teaching of Nahuatl on all academic levels. Researchers need to incorporate or facilitate the incorporation of their data into innovative and efficient resources for instruction. We plan to establish a monolingual university program grounded in international collaboration, and strive to work efficiently and productively with state educational institutions in order to improve teaching methodology and extend the presence of native languages in primary and secondary education. Crucial to the fortification and development of Nahuatl education and literacy is the implementation of a standardized orthography that preserves the richness of varietal differences.

It is equally important to disseminate and apply the results of psycholinguistic research on multilingualism in order to stimulate positive attitudes on the part of parents, teachers, politicians, social workers and other service professionals toward minority language transmission and multi-language education.

Our approach involves direct collaboration with members of the language communities we are studying and working to revitalize. Native speakers work with us as students and researchers, not informants. And we provide them with training to assure that they may successfully carry out any number of educational, social and political tasks essential for guaranteeing linguistic and cultural growth. They are encouraged to actively develop and extend the use of their language into more and more sectors of modern culture and social life, especially through the creation and expansion of spaces for monolingual language practice. Other essential aspects that need to be incorporated into community-based revitalization projects are master-apprentice and language-at-home programs. These essential components of language revitalization can draw on psycholinguistic research and experience. “Children are perfectly capable of growing up bilingual, trilingual, or even quadrilingual. But parents and mentors must create an environment where both (or all) of the languages can thrive” (Hinton 2013: 230). The success of these approaches depends on training, which provides knowledge of the psycholinguistic and health benefits of multilingualism, helping to overcome fear of discrimination and the resulting failure of children in school or in the job market, negative attitudes toward lack of fluency and language-change, and the pitfalls associated to language purism. In Wilamowice, the formal instruction in the community’s endangered language which came about as a result of our collaboration with local teachers of the language, school administrators, and municipal authorities was preceded by special psycholinguistic and educational workshops for children, their parents and teachers. In the Nahuatl context, a similar initiative has been undertaken by activists from San Miguel Xaltipan (Beatriz Cuahutle Baustista, Refugio Nava Nava) in the form of a special community-empowerment workshop and informal Nahuatl teaching to children.

As has been argued, stable transmission at home and the presence of daily speakers in an ethnolinguistic community is much more crucial for language maintenance than any virtual community of speakers or external prestige enjoyed by the language (Fishman 2001: 465).

Revitalization cannot be achieved without the construction of a positive language ideology that is crucial for building native speakers’ self-confidence, strengthening their historical and cultural identity, enhancing their professional performance, and assuring external/international recognition of their language. Positive language ideology should make both native speakers and the broader society aware of the benefits of harnessing an endangered language as a unique cognitive tool. It is essential to overcome the isolation of groups that struggle to preserve their languages, such as different Nahua communities in...
many regions of Mexico. Communities can be brought together by focusing their attention on shared problems relating to administrative barriers, discrimination and marginalization, and providing channels for interdialectal discussions and the expression of individual native speakers’ voices. Finally, these discussions and activities must be included within the context of a complex and diverse global cultural heritage through the implementation of collaborative research projects, web portals, meetings, conferences, and special courses all involving the participation of both native speakers and international students and researchers.

REFERENCES


De la Cruz Cruz, Eduardo. 2015. Tototahuan inixtlamatiliz. Totlahtol Series. Warsaw: Faculty of “Artes Liberales”, University of Warsaw & Instituto de Docencia e Investigación Etnológica de Zacatecas.
Galícia Chimalpopoca, Faustino. 1869. Epitome o modo fácil de aprender el idioma náhuatl o lengua mexicana. Mexico D.F: Tipografía de la V. de Murguía é hijos.


Justyna Olko
jolko@al.uw.edu.pl

John Sullivan
idiez@me.com
Brief considerations about language policy: An European assessment

Francisco Carvalho Vicente and Paulo Carvalho Vicente
CIDLeS - Centro Interdisciplinar de Documentação Linguística e Social

The rising of language policy worldwide is a consequence of a globalized world and the openness of borders. Even countries with a relative cultural homogeneity face nowadays new challenges regarding massive migration fluxes and the results of growing awareness for endangered languages and cultures, notably in Europe. This is being noticed around the Old Continent where diversity proves to be a distinct value since ever. In this paper we reflect on the scope of cultural identity and multilingualism to shed new light on language policy and consequently refresh our understanding of a key policy, which is already a decisive public policy for the European peoples.

1. MANY ANGLES OF ANALYSIS IN LANGUAGE POLICY. In the context of strong migration movements, language policy has been highlighted as one of the most important cultural policies, receiving, as an investment area, a major development. Indeed, the actions of promotion and internationalization of a given language, as well as preservation of languages, allows the maintenance of old relations and the implementation of new ones, both in the present and in the future. These initiatives permit the design of world representations, creating an identity reinforcement of what one language gets and produces, which operates as “self-vision” and “worldview” (symbolic value). The promotion of languages, such as the Portuguese, reflects a strategic value, since it is assumed that it is an instrument of national and transnational political unity, clarifying its position amidst other dominant languages and heightening the knowledge of the language as a reflection of a specific way of living and being of its speakers (Gama 2009, Gama 2010, Pinto 2001).

A strict approach to the concept and practices of language policy requires the attention to a wide range of aspects that, even being apparently parallel to their study, influence significantly the systematic apprehension of a linguistic reality, spatially and temporally well located. Thus, language policy constitutes a strongly rooted study field. Bernard Spolsky characterizes language policy as the study of regularities on the choices among the varieties of a language and adds:

[it] includes not just the regular patterns of choice, but also beliefs about choices and the values of varieties and of variants, and also, most saliently, the effects made by some to change the choices and beliefs of others. (Spolsky 2005: 2152)

The author clarifies that language policy exists even when it is not clearly defined and put into force by an authority. Many countries, institutions, and social groups do not have
formal or written language policies, whereby the nature of a language policy must result from the study of language uses and beliefs. Even when a recognized language policy exists, its effects on language use cannot be guaranteed or consistent (Spolsky 2005: 2153).

Language policy and planning were initially linked to the literary policy in the post-colonial states, in particular the selection and standardization of a national language. Those policies were criticized for treating multilingualism as a problem, promoting national languages as building tools and mechanisms for the unification of the nation, ignoring and discouraging linguistic diversity and minority languages. Table 1 summarizes the understanding of language policy and language planning, making clear the differences between the two concepts.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Reach</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy</td>
<td>positions, principles, decisions, strategies</td>
<td>top-down, official policy for languages</td>
</tr>
<tr>
<td>Planning</td>
<td>specific measures, practices</td>
<td>bottom-up, first level measures to support languages</td>
</tr>
</tbody>
</table>

**Table 1: Definitions of language policy and planning**
(adapted from Sallabank 2011: 278)

In the language planning domain two types of activities can be identified: Attempts to modify the language itself (*corpus planning*) and attempts to modify the environment in which the language is used (*status planning*). The latter refers to the attempts made to ensure political and official recognition of a language. It also includes the expansion of the domains in which the language is used, such as in the legal and governmental spheres and in the new media (Sallabank 2011). Following the argument of Ruíz, Sallabank (2011) recalls three “orientations” regarding language policy:

1. **Language as a problem**: In this perspective, multilingualism can lead to a lack of social cohesion and ethnic conflict. Following this orientation, endangered languages are associated to poverty and disadvantage;

2. **Language as a right**: Full participation in the society through the mother tongue. This requires the provision of educational resources, translators, etc., and can therefore be expensive and cause conflicts;

3. **Language as a resource**: Multilingualism increases the capacities of the society as a whole, enhances the status of subordinate groups, promotes local economy and culture, encourages awareness of other views and mutual respect, rather than domination. In this case, endangered language communities are seen as sources of uniqueness.

**2. Language Policy in the Global Era.** Spolsky (2005) presents four factors as key determinants of a nation’s language policy: The sociolinguistic situation, national ideology, recognition of English as a global language and the knowledge of linguistic rights.

Gender issues, social status, and provenance (city vs. countryside) determine people’s views about endangered/minority languages and likewise influence language policies. Policies for the preservation of a language need to take into account gender, age, and economic
constraints and thereby find ways to help minority groups to promote themselves economically, while maintaining the typical social structure of the community to which they belong.

The Constitution of India, for instance, recognizes 18 languages (scheduled languages), the other languages that are not mentioned in that document are seen as “minority languages”. Defining the parameter for the classification of minority languages on the basis of their numerical strength (number of speakers) is not appropriate in the context of India. The numerical criterion (size of the speech community, i.e. the number of speakers) is inappropriate to describe the status of minority languages. One language may be spoken by a low percentage of people in a given state and be classified as an official language. In the Indian context, the domain criterion also fails because different languages are dominant in different domains. English, for example, is dominant in higher education and business (as the lingua franca of globalization), but not in religion (Pandharipande 2002). Another fact that complicates the definition of minority languages in India is related to the major reorganization of the states in accordance with the concentration of languages in different parts of the country. The reason behind this reorganization is connected to the purpose of minimizing the number of linguistic minorities; however, it produced the opposite effect, i.e. new minorities have emerged (Pandharipande 2002). In general, languages that do not have cultural, economic or political power tend to be included in the list of minority languages. Other authors recall that attention needs to be paid to the fact that a language policy is not an autonomous factor and that what appears to be ostensibly the “same” policy can lead to different results depending on the space where it occurs (Romaine 2002).

Language policies, as regulatory tools which address how languages should coexist and be used in specific economic, political, and social contexts, are always representations of different language ideologies, i.e. beliefs, visions, and conceptions of the role of a certain language held by the various social actors (most commonly institutional ones) (Krzyzanowski & Wodak 2011). For these authors, language ideologies must also be rebuilt and negotiated in debates in which language is central, like a topic, a target, and, in doing so, ideologies can be simultaneously articulated, formed, and implemented. These language ideology debates are taking place at different levels of public and semi-public spheres (Krzyzanowski & Wodak 2011: 119).

Language change is ultimately an adaptive response to changes in a given culture, what Salikoko Mufwene (2002) identifies as a socio-economic ecology. Arguments for language maintenance without arguments of corresponding changes in the current socio-economic ecologies of speakers seem to ignore the centrality of native speakers in the overall situation. Mufwene writes that in a globalized world “English is certainly a threat to other languages in policies where it functions as a vernacular, but not at all in countries where it has been adopted only to help the local economy interface with the worldwide economy. Thus it is not a threat to Japanese nor to Putonghua in Taiwan, although it seems to be a threat to French in francophone African countries, where French also has a hegemonic status” (Mufwene 2002: 189). For the author, globalization highlights the role of vitality of the socio-economic structure of a given language. It exists at least a partial correlation between the type and range of globalization in a scenario and the fact that the primary language of the economy is threatening other languages (Mufwene 2002: 189). Conversely, the attempt, for instance, to overestimate the space of Portuguese and the enforcement of its use by international, national, and municipal agencies of the official Portuguese speaking countries, mainly Brazil and Portugal, is part of the very logic of globalization that relegates Portuguese to a very particular localism (Breitenvieser 2010: 196).
Language policy is closely related to “the right to speak [one’s] own language” (Wright 2007). The linguistic rights of human beings are crucial for the understanding of language policy, since the effects of power in language practices represent an essential factor for the development of the language (Ricento 2006). Following Skutnabb-Kangas et al. (1995: 2), on the individual level, the observation of linguistic rights implies the right to:

- identify positively with the respective mother tongue (whether a minority or majority language) and see this identification respected by others;
- learn the mother tongue;
- use it in official contexts.

And, on the collective level:

- the right of existence of minority groups (the right to be different);
- the right to develop and enjoy one’s own language;
- the right of minorities to establish and maintain schools and other educational institutions with control of the curricula;
- autonomy in group-internal administrative issues.

Considering the Portuguese reality, Portugal has always been characterized as a culturally homogeneous country (Monteiro & Pinto 2005); a characterization supported by the fact that Portugal has a geographically reduced area, it has the oldest borders in Europe, and it does not accommodate significant ethnic minorities. There are, however, clear linguistic examples that contradict this homogeneity. For instance, the Minderico language, which emerged as a sociolect in a very specific context in Minde and turned into a full-fledged community language, and which is now a linguistic heritage whose importance and value is internationally recognized (cf. Ferreira in this volume), is a testimony of the weight of very specific socio-economic ecologies. Beyond that legacy, its recognition and preservation is a matter of rights, specifically of linguistic rights and by extension of human rights. The Minderico community is increasingly aware of this situation. Besides the revitalization work, the academic, public, and politic debates must continue and be strengthened in order to increase the national consciousness about the linguistic reality that Minderico represents.

Mirandese, an Astur-Leonese minority language also spoken in Portugal, almost exclusively in the Municipality of Miranda do Douro (cf. Ferreira et al. in this volume), had to make its own route (prolonged, indeed) to obtain recognition as a language. It came to pass with the Law 7/99 of January 29 (1999). According to the Portuguese Constitution, Mirandese is the official language in the Municipality of Miranda do Douro and the linguistic rights of the Mirandese community are now officially recognized.¹

At this point, it is important to mention that Portugal has not ratified the European Charter for Regional or Minority Languages (1992), which has entered into force in 1998. This fact shows on the one hand the lack of awareness of language diversity in the country and, on the other, makes the promotion of and support for minority languages in Portugal even more difficult. Other countries with a higher number of linguistic minorities proceed similarly, as is the case of France, Italy, Greece, or Russia.

¹ In 1997, the Portuguese sign language has been officially recognized in the Constitution of Portugal (art. 74).
According to Gal (2010), the European Union gradually became a typical socio-linguistic regime of the late modernity, which, although officially multilingual, interprets its multilingualism in a limited way. Therefore, Gal proposes that the EU should be considered a “top down regime of multilingual standardization that tries to manage increased diversity in the same way nation states managed non-standard varieties” (quoted by Krzyzanowski & Wodak 2011: 119). When analyzing the implications of the Lisbon Strategy in the EU’s multilingualism policy, Krzyzanowski & Wodak (2011) argue that policies for multilingualism oscillate between economic values (Knowledge-based Economy), ideologies, and traditional European cultural values, such as diversity and education. There is not a clear one-way development but rather the opposite, namely, multiculturalism policies are clearly close to macro strategies in the EU, which depend on the global economy and political circumstances. Often, multilingualism is correlates directly with political interests, which are themselves the targets of change because of other complex influences. Defined primarily as indispensable skills for the development of a European Knowledge-based Economy, language and multilingualism assumed the same role as other skills related to Knowledge-based Economy, such as the knowledge of information and the handling of information and communication technologies. Nevertheless, Krzyzanowski & Wodak (2011) conclude that the European approach to multilingualism and language policy has become increasingly a top-down and autopoeitic approach, considering the period from its inception in the late 1990s to the present.2 Non-economic arguments concerning these policies have been removed and references to society and social cohesion at different levels in the context of multilingualism and language policy have been rare. “Democratic” elements receive the same kind of treatment. This happens surprisingly in the period after 2006, when the democratic aspects of European communication with European citizens were deeply creased in many other EU policies (Krzyzanowski & Wodak 2011).

3. CONCLUSION. In the context of a global and connected world, with open borders and subject to a number of influences, language policy and planning in each country is intimately related to multilingualism, more specifically to multilingual networks, and not just to a single language and its development, as a necessary step to build a nation or global trade (Annamalai 2003: 116). Thus, the network should be built upon the functional relationship between languages. A language policy that does not accommodate social changes will be ineffective, while a policy that is not an instrument of social change will not be worth it. Annamalai argues that “a policy for multilingualism must recognize the changes in multilingual networking induced by social, political and economic forces and must at the same time arrest any trend towards monolingualism by the same forces” (Annamalai 2003: 127).

Beyond the work of politicians and bureaucrats (though important and relevant) concerning one language policy in a given territory, Fishman (2002) draws attention to the fact that the official documents and the Constitution per se are insufficient to achieve the respect for linguistic diversity. Indeed, more important than that is the observance of spontaneous and informal uses of a language – a concrete indicator of its social recognition.

---

2 The EU has currently 24 official languages, however only a limited number of those languages is considered working language in the European institutions. Focusing the analysis on the period until 2013, Gazzola (2006) makes an economic assessment of the publishing and translation costs the European Parliament would have in the case of including more working languages and concludes that multilingualism does not imply unsustainable increases in the budget of the institution.
Brief considerations about language policy: An European assessment

REFERENCES


Breitenvieser, Camila. 2010. A Lusofonia em Questão: o “Imaterial” e o Político na Orde-


New Challenges for Multilingualism in Europe in Dubrovnik, April 2010.

Gama, Catarina. 2009. A identidade cultural portuguesa – instrumentos promotores: o Insti-
tituto Camões. Master thesis in Political Science and International Relations. Lisboa: Univer-
sidade Nova de Lisboa.

Gama, Catarina. 2010. Da cultura à identidade cultural. Revista Portuguesa de Ciência
Política 0. 191–194.

evaluation for the European Parliament. Language Policy 5. 393–417

Krzyżanowsk, Michal & Ruth Wodak. 2011. Political strategies and language policies: The
European Union Lisbon Strategy and its implications for the EU’s language and multilin-


Mufwene, Salikoko S. 2002. Colonisation, globalisation, and the future of languages in the

Pandharipande, Rajeshwari V. 2002. Minority Matters: Issues in Minority Languages in

Lisboa: Editorial Estampa.

Oxford: Blackwell.


Cambridge: Cambridge University Press.

Rights. Berlin: Monton de Gruyter.

Rolstad & Jeff MacSwan (eds.), Proceedings of the 4th International Symposium on

Wright, Sue. 2007. The right to speak one’s own language: reflections on theory and practice.
Language Policy 6. 203–224.

Francisco Carvalho Vicente
fvicente@cidles.eu

Paulo Carvalho Vicente
pvicente@cidles.eu