

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: 1¹). 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.

¹ Himmelmann, Nikolaus. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation*. Berlin & New York: Mouton de Gruyter, 1-30.

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, Kiseliünaitè for Kursenieku, and Zarghooni-Hoffmann for Judeo-Spanish), language revitalization (cf. the papers by Olko & Sullivan for Nahuatl and Wymysiöeryś), 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.