

# Languagesindanger.eu - including multimedia language resources to disseminate knowledge and create educational material on less-resourced languages

Dagmar Jung<sup>1</sup>, Katarzyna Klessa<sup>2</sup>, Zsuzsa Duray<sup>3</sup>, Beatrix Oszkó<sup>3</sup>, Mária Sipos<sup>3</sup>, Sándor Szeverényi<sup>3</sup>, Zsuzsa Várnai<sup>3</sup>, Paul Trilsbeek<sup>4</sup> & Tamás Váradi<sup>3</sup>

University of Cologne<sup>1</sup>, Adam Mickiewicz University in Poznań<sup>2</sup>, Hungarian Academy of Science<sup>3</sup>, Max Planck Institute for Psycholinguistics<sup>4</sup>

Cologne, Germany<sup>1</sup>, Poznań, Poland<sup>2</sup>, Budapest, Hungary<sup>3</sup>, Nijmegen, The Netherlands<sup>4</sup>

E-mail: djung@uni-koeln.de<sup>1</sup>, klessa@amu.edu.pl<sup>2</sup>, durayzs@gmail.com<sup>3</sup>, oszkobea@gmail.com<sup>3</sup>, maria.sipos@gmail.com<sup>3</sup>, horeszocseme@gmail.com<sup>3</sup>, varnai.zsuzsa@gmail.com<sup>3</sup>, paul.trilsbeek@mpi.nl<sup>4</sup>, tavaradi@gmail.com<sup>3</sup>

## Abstract

The present paper describes the development of the languagesindanger.eu interactive website as an example of including multimedia language resources to disseminate knowledge and create educational material on less-resourced languages. The website is a product of INNET (*Innovative networking in infrastructure for endangered languages*), European FP7 project. Its main functions can be summarized as related to the three following areas: (1) raising students' awareness of language endangerment and arouse their interest in linguistic diversity, language maintenance and language documentation; (2) informing both students and teachers about these topics and show ways how they can enlarge their knowledge further with a special emphasis on information about language archives; (3) helping teachers include these topics into their classes. The website has been localized into five language versions with the intention to be accessible to both scientific and non-scientific communities such as (primarily) secondary school teachers and students, beginning university students of linguistics, journalists, the interested public, and also members of speech communities who speak minority languages.

**Keywords:** less-resourced languages, dissemination, resource localisation

## 1. Introduction: Digital archives and less-resourced languages

Even though there are about 7.000 languages spoken in the world, sufficient language resources are available for less than 10 percent of those. In addition, more than half of these languages are actually threatened to become extinct within a century. Due to globalization and technological innovation cultures and languages are subject to extreme changes. Many of them will become extinct in the coming decades, and with them, much knowledge about nature, history, and cultural practices will be lost. This loss has started to become internationally recognized in recent years, as e.g. seen in the UNESCO efforts to include endangered languages as integral part of its intangible heritage initiative<sup>1</sup>. At the same time, academic institutions have started to set up digital archives to preserve and document cultural and language material for future generations, as well as for academic purposes such as data collection for cross-linguistic and interdisciplinary research.

Under-resourced languages do not only encompass endangered languages from the Americas or Australia, but also most minority languages around the globe, including some European languages and dialects. The Language Archive in Nijmegen was the major technical partner of the DOBES-initiative of the VolkswagenFoundation

(2000-2013), and has developed archival standards and tools specifically for multimedia language documentation in collaboration with more than 100 research projects across the globe<sup>2</sup>. These language documentation projects have led to the creation of large multimedia corpora that consist of primary data such as video and audio recordings, as well as annotations containing translations, transcription, and linguistic analysis to some degree.

Until now, the primary stakeholders in these archives have been researchers from the linguistic and anthropological domain, as well as members of the speech communities represented in these archives. The possibilities for future corpus studies and the richness of these data of little-investigated languages have only been started to be explored (cf. Haig et al., 2011; Seifart et al., 2012). As Holton (2012) points out, there is a distinction between the larger language archives such as the DoBeS-archive or the ELAR-archive that represent a global repository for languages world-wide, and the regional archives that usually focus on particular areas and language families such as the Alaska Native Language Archive (ANLA)<sup>3</sup>. The latter are increasingly used as heritage language resource provider by native communities, as Holton describes: "First, the primary users of the Archive are Native speakers and their descendants, that is, members of Alaska Native language communities" (2012:106). However, the archived material is of interest to other user

1 [www.unesco.org/new/en/culture/themes/endangered-languages/](http://www.unesco.org/new/en/culture/themes/endangered-languages/)

2 cf. [dobes.mpi.nl](http://dobes.mpi.nl)

3 [www.uaf.edu/anla/](http://www.uaf.edu/anla/)

communities as well: non-scientific communities such as lay persons interested in diverse cultures, secondary school teachers, journalists, the interested public traveling around the globe, etc.

There are some initiatives that target regional language revitalization in Europe (e.g. Lule Saami Centre in Drag, Norway, Sorbian Schools in Saxonia, Germany). But given the prevalence of globalization and the huge dynamics of migration, we see it as an enormous opportunity to facilitate learning about language diversity in our educational systems. Hence the idea to the current project to integrate multimedia resources from language archives as well as from freely available materials, and integrate them into a general information and learning platform on less-resourced language with their cultural, geographical and linguistic characteristics. Ideally we would like to transmit the message that this diversity should be seen as a richness of human heritage that is worth maintaining, but also fun to learn about.

## 2. INNET

The INNET project – *Innovative networking in infrastructure for endangered languages* – is funded under the FP7 programme by the EC<sup>4</sup>. The partner institutions include the University of Cologne (Germany), Max Planck Institute for Psycholinguistics, Nijmegen (Netherlands), Adam Mickiewicz University, Poznań (Poland), and the Research Institute for Linguistics, Hungarian Academy of Sciences (Hungary). The technological backbone of the INNET project is provided by two international initiatives: DOBES and CLARIN. By employing the standards and tools developed in these initiatives, INNET strengthens the international impact of their results pertaining to eScience and infrastructure. At the INNET *Best Practice Conferences* (2012 and 2013), the main international digital languages archives participated and discussed the extension of the CLARIN network for language archiving, processing, and visualization.

To support the linguistic as well as technical work of digital language archiving, methodological experience as well as best practices are discussed on such topics as:

- 1) resources on technical infrastructure,
- 2) resources on best practices in language documentation and language technology,
- 3) blogs and news on (and in) endangered languages and their documentation,
- 4) archiving networks and resources,
- 5) software tools and technical issues.

The pertinent goals and objectives of the INNET-project relating to the dissemination of knowledge were framed as the following:

- Creating educational and scientific opportunities for young scientists and technicians in the language information and data domain to work with these archives

- One general objective of the INNET project is to go into schools and inform youngsters about the variety of languages and language systems and to transmit the message that this diversity should be seen as a richness of human heritage that needs to be maintained. (The development of in-depth materials for these purposes is demonstrated in Section 3 below).

In order to include the interested public as well as the educational sector in the discussion of minority or endangered languages and their documentation in technological settings, one of the foremost goals has been to conceptualize and implement an accessible website providing this information in an acceptable educational format. So far, there have been few attempts to make use of archival multimedia data to use in a classroom setting as teaching tool for endangered languages or linguistic diversity, one notable exception being Odé (2009). The conceptualization of the INNET website – <http://languagesindanger.eu/> – is described in the following section. The website itself was decided to be localized in five European languages as a start: English, Polish, Hungarian, German and Dutch. This way, its purpose to be a usable tool and resource provider in secondary schools can be fulfilled.

## 3. Languagesindanger.eu - design and implementation

### 3.1 Design and initial assumptions

Before the actual implementation of the website, the main functions of the product to develop were determined. These functions can be summarized as related to the three following areas:

- to raise students' awareness of language endangerment and arouse their interest in linguistic diversity, language maintenance and language documentation;
- to inform both students and teachers about these topics and show ways how they can enlarge their knowledge further; special emphasis on information about language archives and how they can be used by a general public;
- to help teachers include these topics into their classes.

The target groups of the product are:

- students of secondary schools, i.e. students aged 16-18 in the case of Europe,
- secondary school teachers, especially teachers of Geography, Social Studies, Cultural Studies, English as a Foreign Language, and also Mother Tongue.

Three basic requirements were kept in mind during the development of the package:

- attractiveness and multi-media environment: the design has to be visually attractive to both target groups; audios and videos are an important element;

<sup>4</sup> [www.innet-project.eu](http://www.innet-project.eu) (FP7 grant agreement no: 284415)



Figure 1: Languagesindanger.eu - the starting page

- interactivity: especially students should be encouraged to interact with the product; exercises are an important element;
- user-friendliness and applicability: the product shall be easy to use without previous knowledge and special skills; teachers shall find material ready for use in their classes.

### 3.2 Implementation

After conducting a curriculum research (Wójtowicz, Linda, 2012) and consulting the representatives of the target groups of users (teachers and students) followed by internal discussions within the INNET project consortium it was decided that technically, the product would be implemented as an interactive website. The English version of the information package was implemented and published at:

<http://languagesindanger.eu/> (a screenshot of the starting page is shown in Figure 1).

The website was developed using WordPress Open Source web software and its content management system (CMS). A new WordPress theme was created within the present project using a newly designed original graphic layout. The name of the site, "Languages in danger", appears in the header of the starting page in turn in English and the four other INNET languages (German, Dutch, Hungarian, and Polish). The Interactive Map section was initially created with the use of the Flash technology, however, after preliminary testing it was decided to additionally develop an HTML5 version of the exercises in order to enable the use of the service under all operating system platforms.

The finally implemented website consists of five main sections: the Starting Page, Interactive Map, the Book of Knowledge, Teaching Material, and the page "What can you do?".

#### 3.2.1 The starting page

Apart from an introductory text, the starting page at languagesindanger.eu includes the first two interactive modules in the form of a quiz question and a *featured language* information both alternating at random choice. The alternating quiz questions concern languages mentioned in the present website as well as to another languages and topics related to endangered languages. The featured language information was included for all languages that appear at the Interactive Map (Section 3.2.2) and contains a brief description of the languages, photographs, references to locations where more information can be found as well as the links to the Interactive Map exercises.

#### 3.2.2 The Interactive Map

A map of the world with five clickable areas (America, Europe, Africa, Asia, Australia and Oceania) opens a zoom view of the selected part and displays the locations of the featured languages for that region. Data for more than twenty featured languages were prepared, delivering materials including newly designed interactive exercises, videos, pictures and descriptions.

Apart from the regular view of the map, an anamorphic view (Figure 1) is included in order to illustrate the world's linguistic diversity and to draw students' attention to the distribution of the number of languages spoken in



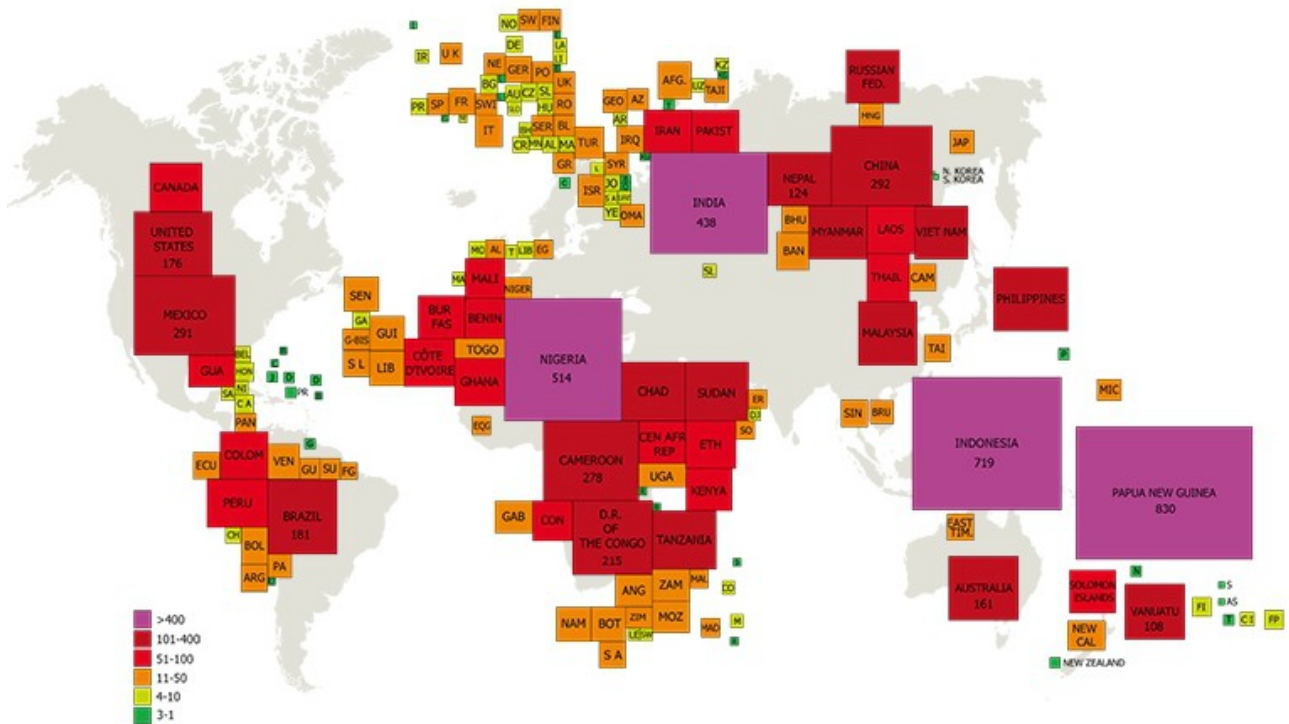


Figure 2: Anamorphic view of the interactive map at [languagesindanger.eu](http://languagesindanger.eu) (numbers based on Ethnologue, 2009)

various areas (the numbers were based on the Ethnologue, Lewis (Ed.), 2009).

### 3.2.3 The Book of Knowledge

The Book of Knowledge is an online text-book organized into ten chapters providing detailed information about linguistic diversity, language endangerment, language documentation and related topics such as the sounds of languages, language structures, cultural and political issues. This part of the website is dedicated mainly for more interested students as a self-study resource as well as for teachers. As a future work a downloadable and printable version of the Book of Knowledge is planned to be published (by September 2014).

### 3.2.4 Teaching Material

The Teaching Material section provides newly designed, detailed outlines of lessons in a form of downloadable materials including worksheets, background information and supplementary material such as audio recordings or ready-to-use presentation slides.

The teaching materials were developed in cooperation with teachers in each of the partner countries. Thus, the specific exercises and worksheets differ slightly across the website's language versions (see also Section 4 below), due to the differences in the education system curricula as well as using varying (localised) language examples.

### 3.2.5 The page "What can you do?"

The "What can you do?" section is aimed to encourage and stimulate students interested in the topics of language documentation, protection or revitalisation to start their own projects or studies, to broaden their knowledge by getting acquainted with the existing literature of the subject or studying linguistics. Among others, the section provides video interviews with young researchers, linguists or activists describing their own experiences in the domain of endangered or minority languages.

## 4. Localisation

### 4.1 Five target language versions

The first two language versions to be developed were the Polish and English versions (already published on-line). Hungarian, German, and Dutch are in the final phase of preparation. The reason for the development of the local language versions was to account for differences between the specificity of languages and cultures in the four countries participating in the INNET project and (perhaps primarily) to make the package more accessible for school pupils in particular countries. The need for a wider accessibility was confirmed by a series of curriculum studies aiming at investigation of the level of awareness of linguistic diversity and language endangerment among pupils and teachers (cf. e.g. Wójtowicz & Linda, 2012). The development of the local language versions was not only a simple translation of the contents from English or Polish but rather its adaptation and extension. Particular language versions might differ in terms of examples illustrating the topics discussed in the Book of Knowledge

with a view to better match the specificity of a particular region and to be better understood by the local audience. For example, in case of Polish version, the more focused languages included Latgalian (Nau, 2011), Wilamowicean (Wicherkiewicz, 2003) - i.e. languages that are/were spoken in the territory of the present or historic Poland (see also Klessa & Wicherkiewicz (2014) for a broader review of the Polish linguistic diversity). In the German version in contrast, the Sorbian language and German dialects are discussed as examples of minority languages and threatened language varieties.

It was decided that the English language version was meant to be the main one and thus it should contain all contents present in any other language version.

#### 4.2 Localisation case study: Hungarian

In the course of the adaptation of the contents for the Hungarian version., the first task was to test and adapt the course materials designed by the Polish partner to the needs of teachers and students at secondary schools in Hungary according to the requirements defined in the Hungarian National Curriculum (2012) and in the National Framework (2013) in order to enable a possibly non-problematic inclusion of the materials into the school teaching programme. The overall aim was to upload a collection of tasks and exercises on the *languagesindanger.eu* website for the teachers to be able to widen the linguistic horizon and attitudes of students towards endangered languages and, while developing language skills, to disseminate knowledge about linguistic diversity, endangered languages, especially about small Finno-Ugric minority languages and some minority languages in Hungary.

The evaluation and testing of the product was done with the participation of the product end-users. Having established contacts with the teachers interested in the project, the students' opinions were solicited through a preliminary survey enquiring about the knowledge students have about the concepts of multilingualism and language endangerment. The next phase in close cooperation with project partners both in Hungary and abroad as well as with Hungarian secondary school teachers was directed at the development of the course materials. Uploading of the materials is preceded by testing them in selected classes. During the adaptation of the teaching materials it was soon realized that the concept of linguistic diversity fits best in the curriculum of the school subject Hungarian language as a mother tongue. Thus, a new class on Hungarian language was added to the set of teaching materials that elaborated on the following topics: (1) Linguistic diversity; (2) The world's languages: multilingualism; (3) Endangered languages, Minority language situation, Language loss, Language change; (4) Linguistic minorities in Hungary – Hungarian in minority.

Thematic diversity is supplemented by exercises of various sorts. Besides individual reading and writing comprehension exercises students can work in groups to

collect data on the topics of the project and to disseminate the information in class. The materials compiled could also be used by teachers in extra-curricular classes. Many schools organize thematic days, e.g. the Day of the Hungarian Language, the Month of Hungarian Science, Finno-Ugrian Days, which are good opportunities for teachers and students to make use of the information available on the website when preparing for contests and to disseminate them.

### 5. Conclusions

The implementation of the *languagesindanger.eu* – website has received extremely positive feedback from the academic as well as the secondary school sector so far. Two aspects turned out to be crucial: 1) the multilingual perspective on languages around the globe presented in a multimedia format, and also 2) the localized perspective reflecting the individual country's situation of its language(s) and dialects. The conceptualization as well as implementation of the materials that make use of language resources have become a core area of the general research project. Based on the feedback received during the on-going tests and consultations with the users of the package, it appears desirable to further develop the website service, primarily by adding further detailed information and Interactive Map exercises for new languages, i.e. to continue the educational application of the INNET-project.

### 6. Acknowledgements

The authors wish to thank the INNET Project team members, especially: Marianne Bakro-Nagy, Nikolaus Himmelman, Michael Hornsby, Maciej Karpiński, Nicole Nau, Tomasz Wicherkiewicz, and Radosław Wójtowicz. We would also like to address special thanks to all persons who voluntarily contributed to the development of *languagesindanger.eu* website either by making available their multimedia resources or by participating in the testing phase. Thank you also to the anonymous reviewers and their helpful suggestions for this LREC-contribution.

### 7. References

- DoBes (Dokumentation bedrohter Sprachen / Documentation Of Endangered Languages) Project (2000-2013), available on-line at: <http://dobes.mpi.nl/>
- Haig, G.J.L., Nau, N., Schnell, S., and Wegener, C. (Eds.) (2011). *Documenting Endangered Languages: Achievements and Perspectives*. Berlin, New York: Mouton de Gruyter.
- Holton, G. (2012). Language archives: They're not just for linguists any more. In Seifart, F. et al (Eds.), pp. 105–110.
- Klessa, K. and Wicherkiewicz, T. (2014). Design and Implementation of an On-line Database for Endangered Languages: Multilingual Legacy of Poland. To appear in: *Proceedings of 6th International Conference on Corpus Linguistics (CILC 6)*, Las Palmas de Gran Canaria, 22-24 May 2014.
- Lewis, M. Paul (Ed.) 2009. *Ethnologue. Languages of the*

- World*, 16th edition. Dallas: SIL International:  
<http://www.ethnologue.org>
- Nau, N. (2011). *A short grammar of Latgalian*. München: LINCOM Europa.
- Odé, C. (2009). Teaching Materials on Language Endangerment. An Interactive E-learning Module on the Internet. In de Graaf, T. et. al. (Eds.), *Endangered Languages and Language Learning: Proceedings of the Conference FEL XII: 24–27 September 2008, Ljouvert/Leeuwarden*. Leeuwarden: Fryske Akademy, pp. 147-150.
- Seifart, F., Haig, G., Himmelmann, N.P., Jung, D., Margetts, A. and Trilsbeek, P. (Eds.) (2012). *Potentials of Language Documentation: Methods, Analysis, and Utilization*. (Language Documentation & Conservation, Special Publication No. 3). Honolulu: University of Hawai'i Press.
- Wicherkiewicz, T. (2003). *The Making of a Language. The case of the idiom of Wilamowice, Southern Poland*. Berlin-New York: Mouton de Gruyter.
- Wójtowicz, R. and Linda, K., (2012). INNET Innovative Networking in Infrastructure for Endangered Languages. Deliverable D4.1 Awareness Report, available on-line (accessed on 21 March 2014): [http://innetproject.eu/sites/default/files/D4.1\\_Awareness.pdf](http://innetproject.eu/sites/default/files/D4.1_Awareness.pdf)