17

Online presentation and accessibility of endangered languages data:
The General Portal to the DoBeS Archive[1]

Gabriele Schwiertz

University of Cologne

Data depositories containing language documentation corpora are generally well structured, well maintained, and include large collections of many under-researched languages. However, they are not yet conceived of as resources that can be easily consulted on scientific or non-scientific questions pertaining to one of those languages. A general portal to the DoBeS archive has been created to facilitate access to the data, to attract more users to the archive, and to lower the threshold for users outside the linguistic community to access the data. The structure and the main features of this portal will be presented in this paper.

1. INTRODUCTION. When considering the exploitation of language documentation data contained in language archives, three major user groups can be identified: The speaker community, the scientific community, i.e. linguists and scholars of related disciplines, and the general public. Each of these user groups has different interests and different needs, all of which are hardly satisfied by the IMDI-tree representation of the DoBeS archive. For the community users, community portals have been created in some projects. However, in order to attract users to the archive, facilitate access to the data, and generate new user scenarios and communities, we have created a general portal to the DoBeS archive.

This portal tries to address the needs of the different user groups with different sub-portals where selected information specific to typical queries of those users is presented in a straightforward way and shortcuts to the relevant media are included. In the future, this portal will serve as the main entryway to the archive and as a container of information about the DoBeS projects and language endangerment in general. In this paper, I will briefly discuss features of the portal and the motivation for the current structure. A preliminary

---

[1] The structure of the DoBeS portal has been presented in various settings and I wish to thank everybody for their contribution: This article has benefitted greatly from input of the audiences at the DoBeS meeting 2010 in Nijmegen, from a meeting with DoBeS teams and linguistic students in Cologne, from the Potentials workshop in Leipzig, and from discussion among a dedicated group of DoBeS members. I especially would like to thank Nikolaus P. Himmelmann, Dagmar Jung, and Iren Hartmann for their help. I also wish to thank Paul Trilsbeek and the Technical Team at the MPI for Psycholinguistics in Nijmegen for the actual implementation of the website.

version of the portal can be found here: http://ems03.mpi.nl/dobes/, in due course the portal will replace the DoBeS website at http://www.mpi.nl/dobes.

2. STRUCTURE AND FEATURES OF THE DOBES PORTAL.

2.1. MAIN PAGE. The first page of the portal has two main functions: It gives a first impression of what DoBeS is, and it constitutes the gateway to the sub-portals dividing the user communities. There is an interactive map in the middle of the page with all the DoBeS projects. From here, it is possible to go immediately to the single project pages. On the left side, appetizers are presented: Videos, audio recordings, or pictures that are freely accessible can be called up. On the right side, quick links and a Google-style search-box are provided. Under the map, users are divided into three main groups: The scientific audience, e.g. linguists, ethnographers, and musicologists; the general public, i.e. journalists, community members,[2] or interested laymen; and into future-depositors. As on all pages, a constant reminder of issues of access rights, access levels, and user registration is included with links to more detailed information on these topics.

2.2. RESEARCH PORTAL. The research portal[3] provides information on what kind of data is contained in the archive and what can be done with it. Access to statistics for all the projects in the archive is provided in a table showing how many sessions there are for each language, how many include video or annotations, and at what access levels[4]. In order to allow typologists to get a first impression before deciding whether the data in the DoBeS archive is suited to their needs, a link to all grammars and sketch grammars is provided together with a link to studies deposited in the archive. Furthermore, information is given on what kinds of searches are available, and what DoBeS related tools there are. Manuals are supplied for certain usage cases, e.g. writing an MA-thesis using data in the archive. In addition, more specialized information pages for various scientific user groups are linked here.

2.3. GENERAL INTEREST PORTAL. The general interest portal[5] is aimed at providing information about language endangerment and at simplifying access to the DoBeS archive for the general public, giving an initial impression of possible points of interest it contains. Again, there are appetizers on the left, and access to an overview of archive statistics is given. A link to a collection of community portals is supplied. Information is provided about the searching and browsing functions and about registration and access. 'Highlights' of the archive, items which we hope will be interesting to the general public, are collected here; these include subtitled videos, artworks, story collections, and other appealing material.

---

[2] Some teams have created community portals for the members of the communities they work with. These include information that is relevant for this audience, often in a language other than English. However, for the general portal, it remains to be seen with increasing use of the website whether there should be a third sub-portal just for community members.

[3] http://www.mpi.nl/dobes/research

[4] The data in the archive are marked for different access levels, indicating whether everybody is allowed to access them, or some subgroup of users or whether they are not accessible right now.

[5] http://www.mpi.nl/dobes/general

There are also plans for this page to contain other generally entertaining points in the future, such as a 'word of the day' from the DoBeS languages or a photo/video gallery which can be combined in many ways with services like Facebook, Twitter, and YouTube.

2.4. THE DEPOSITORS' PORTAL. For the time being, the depositors' page[6] contains basic information for future depositors to the DoBeS archive. At some point, this page could be a more powerful interface with tutorials, tools, and possibly links to pages with an upload functionality.

2.5. THE PROJECT PAGES. Each DoBeS project has its own page. The project pages contain the same information as the former project pages of the DoBeS website. An addition provided by the new portal is a summary page with the same structure for all projects; here, statistics will be displayed as well as a very brief statement about what the project is and a note on the team. Additionally, there will be a short list of highlights contained in the respective archives, such as a large number of videos about hunting techniques, a dissertation on the language, or a topical dictionary. More detailed information about the language, the project, and the team can be accessed from here.

3. SUMMING UP. The general portal to the DoBeS archive will facilitate access to users, create new user groups, and attract an interested audience. It will hopefully turn the DoBeS archive into a resource that various user groups will consult regularly. In the process, monitoring of users and uses of the page will allow us to adapt its structure and content to suit the needs of those users. In due course, more features can be incorporated, e.g. the possibility of working with the materials online, that is adding one's own annotations to the data in the archive.

Gabriele Schwiertz
gabriele.schwiertz@uni-koeln.de

---

[6] http://www.mpi.nl/dobes/deposit