**An inventory of European scholarly dictionaries**
A report on a Short Term Scientific Mission

Gerbrich de Jong
February 2015

In September 2014 I went on a Short Term Scientific Mission to make an inventory of European scholarly online dictionaries. I performed this mission at Trier University, Germany, as a member of the Center for Digital Humanities. The mission had a time span of seventeen working days. My supervisors were dr. Vera Hildenbrandt from Trier University, who is the chair of working group 2, and dr. Anne Dykstra, from the Fryske Akademy, who is the chair of working group 1.

The result of my mission, an inventory of online scholarly dictionaries in Europe, is the first step towards the goal of working group 1: to set up a dictionary portal. In order to create such a portal, someone needed to inventory what scholarly dictionaries could be in this portal and where they could be found. Although the aim was to provide an overview of online scholarly dictionaries, I also added other, high quality dictionaries. This is done because not all languages have a scholarly dictionary (yet) and because some high quality dictionaries simply cannot be ignored. We do not want the future users of the forthcoming European dictionary portal to miss any famous dictionary what they would like to use. Moreover, I searched for titles of printed scholarly dictionaries in case there was no high-quality dictionary available online. This was to meet the goal of working group 2, to make an overview of dictionaries which should be retro-digitized. My Scientific Mission resulted in a spreadsheet containing 125 dictionaries for 83 languages. This article is a clarification in order to understand how choices are made and how the spreadsheet’s content should be interpreted.

As the aim was to find at least one dictionary for each European language, the starting point was to list all European languages into the spreadsheet. This first step raised some difficulties, as it is not easy to determine which varieties should be in the list. Ursula Schültze, my supervisor’s student assistant, helped me listing European languages. Besides the exclusive national languages and the national languages with extraterritorial spread, we tried to list officially recognized minority languages as well. But it appeared to be too time-consuming to sort this out for all varieties spoken in Europe. Therefore, after a certain amount of time, we decided to leave the list (at that time containing 83 languages) as it was and to start doing what it was all about: finding scholarly dictionaries for these languages. The list of languages is of course open for discussion. There might be languages missing and there might be languages in which could be left out (for example because the language has only few speakers and there’s no decent online dictionary for it).

After the pragmatic choice was made to leave the list as it was, the search for online scholarly dictionaries could start. A complication was that there is no consensus about what a scholarly dictionary is. My working definition was:

A scholarly dictionary is a dictionary that has authentic illustrations with exact references to their sources. Such a dictionary is by definition descriptive and it explicitly strives to be an extensive and detailed synchronic inventory of a particular language.

During the workshop in Bled in the end of September 2014, this definition was slightly edited. During the conference in Vienna in February 2015 Dirk Kinable proposed a definition including the input of members of the different working groups. This definition was considered too detailed, and therefore the discussion about what a scholarly dictionary is, is still ongoing.

It was challenging to find scholarly dictionaries on the web for languages that I have no command of. I used Google Translate to translate some seeking terms, which I used in Google. Sometimes the website of an institution supporting the language was of help, especially when it provided information on language promotion projects including dictionaries. To be able to understand some of these websites, I used Google Translate again to translate the information.

Scholarly dictionaries are sometimes very well hidden on the web. Sometimes the website itself, when finally found, is difficult to explore. Some dictionary websites offer an English version of the website, but most dictionaries do not. They are not user-friendly if it comes to users who are unfamiliar with the language. If a dictionary was only provided in the target language, I used Google Translate to find out where to fill in the lemma. I also used Google Translate to translate the search terms and the entries for me. From these Google Translations I tried to deduce all I wanted to note about the internal structure of a dictionary.

That contained various things, which I will describe below. I filled in most details by using numbers. One can see what these numbers mean by hovering over the first box of each column (a legend will pop up). Besides the title of the dictionary and the country where the dictionary project is located, I tried to provide a translation of the title in English. If there was no English title provided by the dictionary, I translated the title myself (again using Google Translate). As Google Translate is not always reliable, and my knowledge of European languages is limited, these translations are often open for improvement.

I only noted the abbreviation if it is used on the website to refer to the dictionary, or if the abbreviation is clearly depicted on the website. A careful observation is that dictionaries with an abbreviation are mostly high-quality, well-known dictionaries. It goes too far to say that there is a link between popularity and the existence of an abbreviation, but there seems to be some relation between it.

For every online dictionary I tried to fill in the organization publishing the online version. For each printed dictionary I filled in the publisher instead. Not all online dictionaries explicitly state the institution behind it, so I was not always to fill in this information. I only noted the editor when his or her name was explicitly stated on the website or when I thought It might be useful to distinguish between different dictionaries bearing the same title. There are, for example, many dictionaries bearing the title ‘Deutsches Wörterbuch’, but there is only one Deutsches Wörterbuch von Hermann Paul. Nevertheless, this column often remained empty, as many dictionaries do not mention their editor.

My main goal was to make an inventory of online dictionaries, but if there was no such thing available, I listed the title of a printed dictionary. This was in order to meet the goals of working group 2, which is to provide an overview of dictionaries, which should be retro-digitized. To make it easy to filter the printed dictionaries from the online ones, I tagged them with ‘0’ in the column ‘Accessibility’. Most online dictionaries are fully online available, others are partly or fully behind paywall. Dictionaries which were explicitly stated to be under construction or which are soon to be online, are tagged with ‘2’. Many dictionaries inform the users that the website is frequently being updated with new entries. I understood this kind of construction and maintenance as typical for an online dictionary.

For online dictionaries I noted the way it is been digitized. The first option is full text digitization, which applies for digital born dictionaries. The other options are page image scan with Optical Character Recognition (OCR) and without OCR. In one case there was a combination of two options. The Czech dictionary offers full text search as well as a page image scans of cards from the card-index system. Another technical thing I tried to capture is how a dictionary can be searched. If it was possible to search on lemma (0) and/or on meaning (1). Capturing this demands a high knowledge of the internal structure of a dictionary. Moreover, it was very difficult to find this out for dictionaries written in a language that I do not speak at all. Google Translate was a great help, but did not always enable me to fill in details.

Regarding the catagories: audio-fragment (column L), grammatical characterization (M), etymological information (O), and usage information (Q); I listed if this information was present (1) or not (0). Where (1) means that I found this phenomenon at least once in the dictionary, and where (0) means that i did not find it. This does of course not mean that it is surely not there.

The category meaning (N) is divided into four different options: synonyms (1), paraphrases (2), translations (3), and images (4). In this category various combinations of these options are possible. As can be seen in the spreadsheet, almost every dictionary gives meaning by paraphrases and only few provide images to depict the meaning. The category examples (column P) is filled in with three possible answers: no (0), without source (1) and with source (2). Concerning Cross-references (column R) I distinguished between two types. The first type, to other entries (1), can be found in many online dictionaries. The other type, to other dictionaries (2), is more exceptional. If I did not find any cross-references I filled in 0 (for no). Some dictionaries provide links to a page on which one can see the inflection of a verb. I did not interpret this as a cross-reference.

If cells are empty, this is because I was not able to fill them in. This was the case when I could not make use of Google Translate or when the website of the dictionary did not work. Sometimes Google Translate was unable to recognize the script (for example for an Albanian dictionary) or I could not copy the text to Google Translate as the text in the dictionary was only available in page image scans. If complete rows are left empty, this is because I could not find a dictionary (online or printed) for this language.

Every user of this inventory must realize that this spreadsheet is made by someone who does not speak (most of) these languages. It is therefore necessary that the information in the spreadsheet is checked by the members of the management committee. They have to check as well whether the list provides the best dictionaries available. Currently, the spreadsheet is being improved by using the comments of the management committee. As soon as this process has been completed, the edited and improved spreadsheet will be online available.

I am happy that I contributed to the forthcoming European dictionary portal, by making this inventory. I hope this article provides a clear overview of how to interpret the spreadsheet.