

Discovering Resources in the VLO: Evaluation and Suggestions from a Pilot Study with Students of Translation Studies

Vesna Lušicky

Centre for Translation Studies
University of Vienna, Austria
vesna.lusicky@univie.ac.at

Tanja Wissik

Austrian Academy of Sciences, Austria
and
University of Graz, Austria
tanja.wissik@oeaw.ac.at

Abstract

CLARIN provides access to language resources for scholars in the humanities and social sciences. In theory, scholars and students of Translation Studies may be assumed to be active data providers of language resources, as well as prolific users of the CLARIN services. However, data show that the uptake of CLARIN services by this user group is rather low. This paper investigates the needs of the students of Translation Studies and evaluates the CLARIN VLO from their perspective. It is based on a pilot study applying open and closed situated user assignments and an evaluation of the VLO service. The results provide insights into the needs of this user group and give suggestions to data and service providers that could increase the adoption of CLARIN services by the user group.

1 Introduction

E-research has transformed the process of research and has become a more ubiquitous research practice. CLARIN (Common Language Resources and Technology Infrastructure) aims at providing sustainable access for researchers in the humanities and social sciences to digital language data and tools. As observed in other service-oriented e-research infrastructures (Chunpir et al., 2015), the phase of development, setting-up and running of the CLARIN infrastructure and services was followed by the requirement to conduct studies into user needs and user experience. At the CLARIN Annual Conference 2015 the user involvement survey was presented (Wynne 2015), showing user activities by discipline. Rather surprisingly, Translation Studies were not listed among the disciplines¹ (Wynne 2015) in the diagram of the user involvement by discipline.

At least some branches of Translation Studies, especially Corpus-based Translations Studies and Computational Translation Studies, are carried out with computational methods. They not only heavily rely on various languages resources, e.g. corpora, translation memories, terminology resources, and lexica for research purposes, but they also generate both mono- and multilingual language resources (Budin 2015). In addition, language resources are also extensively used and generated by translation practitioners and students. For these reasons, the absence of documented users from the field of Translation Studies in the above-mentioned study appears noteworthy. The specific needs of this user group that focus on a higher uptake of the CLARIN services, in particular the CLARIN VLO, are being investigated in this study.

2 Discovering Resources in VLO

Digital data, especially language resources, often provide the basis for research projects in the era of e-research. Since creating digital resources from scratch is often time-consuming and expensive, re-

¹ Could be included in the category other humanities.

use of the existing data and resources is recommended. In order to re-use the existing resources, researchers have to be aware of the existence of suitable resources and need “efficient ways to navigate to the language resources that really matter, whatever the selection criterion is” (Van Uytvanck et al., 2012). Various portals, repositories and catalogues provide entry points to the datasets usable in the scope of Translation Studies. Among them are rather general repositories and catalogues that cater to diverse user groups, e.g. ELRA catalogue and META-SHARE, and catalogues curated for specific sub-types of user groups, such as commercial users in the field of machine translation in the case of the LT-Observe catalogue (Maegaard et al., 2016).

VLO (Virtual Language Observatory) is one component of the CLARIN research infrastructure that falls into the former group, addressing a wide range of researchers in the humanities and social sciences. VLO is a metadata-based portal for language resources, providing multiple views on metadata for linguistic data and software and trying to give a consistent online overview of the data that is available at a variety of CLARIN Centres (Van Uytvanck et al., 2010; Van Uytvanck et al., 2012). VLO offers faceted search (language, subject, collection, format, resource type, organisation, continent, national project, country, keyword, modality, data provider, genre) and string search (Odijk, 2014).

3 Pilot study

3.1 Objectives

As discussed above, scholars and students of Translation Studies use language resources in their research and practical work. For this reason, the objective of the pilot study was to investigate which selection criteria really matter to users in Translation Studies, when they try to navigate to the language resources that they need. Secondly, we wanted to test one of the services of the CLARIN infrastructure, namely the VLO faceted browser, as it offers the discovering functionality for language resources (Odijk, 2014), and to find out the perceived quality of the results in VLO by this user group. Lastly, the objective of this pilot study was to find out how students as prospective translators and researchers in Translation Studies would engage with the service and what is needed to ensure a higher uptake of the service by the user group.

3.2 Methodology

Modern translator training internalizes situated learning (Risku, 2016), by which it emulates the actual translation practice through the use of authentic resources, tools, assignments and processes relevant for translators. There is also a substantial overlap between translation competences and research competences, so translation courses can be expected to train some research competences indirectly (Vandepitte, 2013). As translator training caters to both translators-practitioners and researchers in Translation Studies, students as prospective translators and researchers in Translation Studies were investigated as users in this pilot study.

The findings of this study are based on qualitative and quantitative approaches combining open and closed assignments. This method allows the findings to be identified both through pre-formulated research questions and through the formulation of newly raised topics of interest that had not been anticipated during the planning phase of the study.

The data for the present pilot study was collected in one pre-study and two sample studies that were carried out at two Austrian universities in four courses with students of Translation Studies at the BA and MA levels in the winter semester 2015/16 and the summer semester 2016². In all four courses, the format of the classes was designed to include the topics on language resources, e-infrastructures, repositories and similar services. The course at the BA level is recommended for the students in their 5th semester, therefore the majority of the students were in their final year of the BA studies. The majority of the students at the MA level participating in this pilot study was in the second semester of their MA studies, or even in more advanced stages of their MA studies. The participants covered a wide range of working language combinations, as the courses are obligatory for all students with all language combinations offered in the curricula. The users’ working languages ranged from rather traditional combinations (e.g. German, English and Spanish) to Arabic and Sign Language.

² Therefore all references to VLO refer to the version 3.3.

The pre-study established the selection criteria, based on which the users may decide if a language resource is relevant and operationally usable for their purpose. In the pre-study, the users ($n_0=25$) were given an open assignment without any context or background of the services or e-infrastructure in question in order to ensure a minimal bias towards their selection of criteria. The users were asked to provide a weighted list of the selection criteria.

Based on the criteria identified in the pre-study, the first sample user group ($n_1=25$) was asked to query three pre-defined portals (VLO, META SHARE, and ELRA) for language resources in their working languages (cz, en, es, fr, hr, it, pl, ro, ru), and assign a score from 1 to 5 (1 for very high quality, 5 for very low quality) to the perceived quality of the metadata provided for each resource found throughout all of the portals. The second sample user group ($n_2=14$) was asked to solely concentrate on the CLARIN VLO. The users (working languages: ar, bs/hr/sr, en, es, fr, it, ru, sgn) were asked to run a search for language resources in the VLO. They were given basic preselected search criteria in order to yield comparable results. Based on the output, they were asked to describe which further categories of metadata would be useful from the perspective of Translation Studies to be included as “[...] one of the main purposes of metadata is to enable discovery of a resource” (Odijk, 2014). In addition, they were asked to provide a comment on the satisfaction with the search functionalities.

3.3 Results

Unsurprisingly, the results of the pre-study showed that the users highly valued the information on the language(s) covered by the resource and the format of the resource (see Fig. 1). The selection criteria that can be highlighted as rather specific for this user group are reliability of the source, timeliness, and representativeness of the domain.

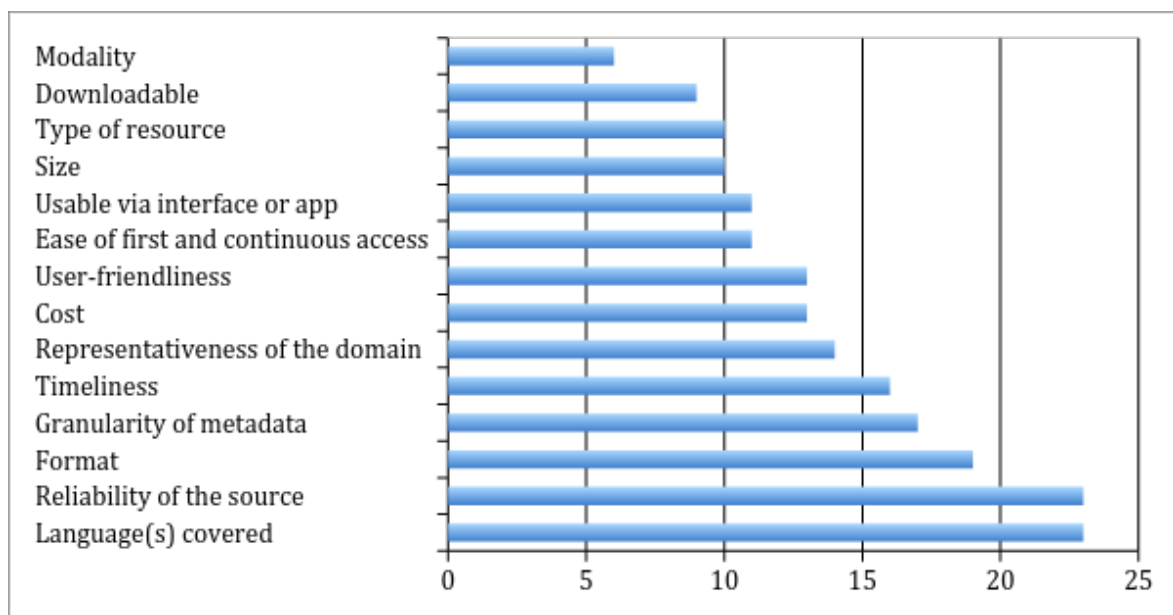


Figure 1: Selection criteria for language resources from the perspective of Translation Studies students ($n=25$)

Based on the criteria defined in the pre-study, the users in the first sample group identified 210 relevant resources³ in total, 20% of which were uniquely found in VLO. The median of the perceived quality of the metadata (1-5, 1 for very high quality, 5 for very low quality) for the resources uniquely found through VLO was 3.

The answers of the users in the second sample group were abstracted, grouped into categories⁴ and divided into the general metadata and the metadata specific to Translation Studies (see Tab. 1). These

³ Multiple selection in more than one language was possible.

⁴ Multiple answers were possible.

metadata might not be applicable for all types of resources, but may be decided case by case, for example for parallel corpora.

Desirable Metadata	
General	Specific to Translation Studies
Author(s) (of the texts)	Original language
Year of publication	Translator(s)
Number of downloads	Mother tongue of the translator(s)
Access (open, against payment, with registration)	Human translation or machine translation
Time span	Information regarding the translation process
Domain (e.g. environment, sport, medicine)	Reliability of the text source
Target group	

Table 1: Desirable metadata in VLO from the perspective of Translation Studies students (n=14)

Regarding the search functionalities, the majority of users were satisfied with the given search functionalities, but less so with the results that were shown. The users were not satisfied with the fact that it was not possible to search for language combinations. It was also suggested that functionalities to store prior search results and to localise the search interface into different languages would be useful. Since our investigation is a pilot case study of real world research (Robson and McCartan, 2016) and didactic in action, the above findings can only be generalised to a limited extent.

4 Conclusions

This pilot study addressed the needs of the students of Translation Studies as prospective translators and researchers in Translation Studies as a user group of the CLARIN service VLO and their assessment of the gaps in terms of the usability of the service. It was established that the resources found through VLO would need some additional metadata information in order to be better usable by Translation Studies scholars and students. Although the metadata that is not generated by the data provider cannot be added to the VLO by third parties, awareness for the multifaceted needs of various user groups should be raised among the data providers. This especially applies in cases when the resources provided had been generated by translators, translation scholars and translation students, in order to ensure a higher uptake of the VLO service as well as other CLARIN services by this user group.

Due to the specific nature of the modern translator training, which emulates the actual translation practice and trains research competences, the present pilot study could be a starting point for further research into the specific needs of the users from Translation Studies of the CLARIN services. Dissemination activities targeting translation scholars, students of Translation Studies, and translators would increase the visibility and the uptake of the CLARIN services by these user groups.

References

- [Budin2015] Gerhard Budin. 2015. Digital Humanities, Language Industry, and Multilingualism – Global Networking and Innovation in Collaborative Methods. In Forstner, Martin and Lee-Jahnke Hannelore, editors, CIUTI-Forum- 2014. Boston, USA: Peter Lang. DOI: <http://dx.doi.org/10.3726/978-3-0352-0290-8>
- [Chunpir et al2015] Chunpir, Hashim Iqbal, Thomas Ludwig, and Dean N. Williams. 2015. Evolution of E-Research: From Infrastructure Development to Service Orientation. In Aaron Marcus, editor, Design, User Experience, and Usability: Interactive Experience Design: 4th International Conference, DUXU 2015, Proceedings, 25–35. Cham: Springer. <http://dx.doi.org/10.1007/978-3-319-20889-3>.
- [Maegaard et al.2016] Bente Maegaard, Lina Henriksen, Andrew Joscelyne, Vesna Lusicky, Margaretha Mazura, Sussi Olsen. Claus Povlsen and Philippe Wacker. 2016. Providing a Catalogue of Language Resources for Commercial Users. In Proceedings of LREC 2016, Tenth International Conference on Language Resources and Evaluation [LREC 2016]. Pp.449-456.

- [Odijk2014] Jan Odijk. 2014. Discovering Resources in CLARIN: Problems and Suggestions for Solutions. Utrecht University Repository, Netherlands.
- [Risku2016] Hanna Risku. 2016. Situated learning in translation research training: academic research as a reflection of practice, *The Interpreter and Translator Trainer*, 10:1, 12-28. DOI: 10.1080/1750399X.2016.1154340.
- [Robson and McCartan2016] Colin Robson and Kieran McCartan. 2016. Real World Research. Chichester, West Sussex: Wiley.
- [Van Uytvack et al.2010] Dieter Van Uytvanck, Claus Zinn, Daan Broeder, Peter Wittenburg and Mariano Gardelleni. 2010. Virtual Language Observatory: The portal to the language resources and technology universe. In Proceedings of the Seventh Conference on International Language Resources and Evaluation, [LREC 2010]. Pp. 900-903.
- [Van Uytvanck2012] Dieter Van Uytvanck, Hermann Stehouwer, and Lari Lampen.2012. Semantic metadata mapping in practice: The Virtual Language Observatory. In Proceedings of the Eighth International Conference on Language Resources and Evaluation [LREC 2012]. Pp. 1029-1034.
- [Vandepitte2013] Sonia Vandepitte. 2013. Research Competences in Translation Studies. *Babel*, 59 (2): 125–148. DOI:10.1075/babel.59.2.
- [Wynne2015] Martin Wynne. 2015. User Involvement. Presentation at the Clarin Annual Conference 2015. <https://www.clarin.eu/sites/default/files/20151016-CAC-04-Wynne-User-Involvement-CAC2015-05.pdf>