

Discovering Resources in the VLO: 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

Common Language Resources and Technology Infrastructure (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 Virtual Language Observatory (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 involvement and user experience. The survey of user involvement in CLARIN was presented (Wynne, 2015) at the CLARIN Annual Conference 2015, showing user activities by discipline. Rather surprisingly, Translation Studies were not listed among the disciplines¹ (Wynne, 2015) of the users involved in CLARIN services.

At least some branches of Translation Studies, especially Corpus-based Translations Studies (Baker, 1993, 1995; Laviosa, 2002; Fantinuoli and Zanettin, 2015)² and Computational Translation Studies³, are carried out with digital methods and tools. They not only heavily rely on various languages resources, e.g. parallel and comparable corpora, translation memories, terminology resources, and lexica for research purposes, but they also generate both mono-, bi- and multilingual language resources (Budin, 2015). Language resources are also extensively used and generated by translation practitioners and by

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

¹ Probably included in the category *Other humanities*.

² Corpus-based theoretical and descriptive research in Translation Studies has investigated topics such as translation universals and norms, ideology and individual translator style, and corpus-based tools and methods are included in the curricula of translation training institutions (see Fantinuoli and Zanettin, 2015).

³ In this paper, Computational Translation Studies are understood as a paradigm and research methodology that takes place at two levels as proposed by Budin (2015): Translations Studies carried out with computational methods and Translation Studies investigating computational processes, for example from a sociological or cognitive perspective (for instance to optimize the human-computer-interaction in translation workflows).

trainers and students in translation training. For these reasons, the absence of documented users from the field of Translation Studies in the study mentioned above appears noteworthy and calls for further investigation.

The specific needs and requirements of this user group that could contribute to a higher uptake of the CLARIN services, in particular Virtual Language Observatory (VLO), by this user group are being investigated in this study. In no way suggesting that the study's results can immediately be generalised, as at this stage the investigation is a pilot study of real world research (Robson and McCartan, 2016) and didactic in action, this paper sets out to show the specific needs and requirements of a specific user group, and to give impulses for studies into user experience of the CLARIN service VLO.

The paper is structured as follows: First, we briefly present selected considerations that were vital for the design of this pilot study in section 2. In section 3, we elaborate on the actual pilot study and present the objectives of the study, the data collection process and settings, and participants. We also illustrate the task design in detail. This is followed by a presentation of results related to each task, the interpretation of results, and discussion. We conclude by summing up the key findings and give an outlook for potential further research.

2 Background

2.1 Language resources in Translation Studies and translation practice

The production, compilation, use, and re-use of various language resources, such as mono- and multilingual corpora, translation memories, terminology resources, and lexica is well-documented and explored in settings related to translation: in translation practice (*inter alia* Beeby et al., 2009; Bowker, 1998, 2002 and 2004; Gallego-Hernández, 2012; Wilkinson, 2005), in translation training (*inter alia* Kenny, 2007; Krüger, 2012, Kübler, 2003; Maia, 2003) and directly in the research in Translation Studies (*inter alia* Baker, 1993, 1995; Granger, 2003, Fantinuoli and Zanettin, 2015).

Computer-assisted translations tools (CAT tools) are used by 73 per cent of translation practitioners (Ehrensberger-Dow et al., 2016). This implies that these practitioners are also highly active in generating and reusing a particular type of language resources, namely translation memories, which are “a very specialised kind of parallel corpus, and are usually relevant, reliable and well integrated into the translation workflow. Of course, translators do not have a translation memory ready for all occasions” (Zanettin, 2012:247). In such cases, translation practitioners either build their ad hoc corpora or look for stable corpora (Sánchez-Gijón, 2003) and other language resources, such as terminological resources and lexical resources. Curated one-stop entry-points where these types of datasets could be found via search activities, addressing specific needs of translation practitioners and scholars, could, therefore, find a good acceptance by this user group.

2.2 Repositories and catalogues

Since creating digital resources from scratch is often time-consuming and expensive, re-use of the existing data and resources is recommended. To re-use the existing resources, researchers, and also other potential users, have to be aware of the existence of suitable resources and need “efficient ways to navigate to the language resources that matter, whatever the selection criterion is” (Van Uytvanck et al., 2012). Various portals, repositories, and catalogues that originate in various e-research projects and initiatives could also 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 tasks – as language resources originally created with a specific purpose are not always generally reusable – and 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 is one component of the CLARIN research infrastructure that falls into the former group, addressing a broad range of researchers in the humanities and social sciences. It 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 in a variety of CLARIN Centres (Van Uytvanck et al., 2010; Van Uytvanck et al., 2012). The 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).

2.3 Approaches to search activities

Search activity, i.e. search for new or additional data, whether due to a pertinent practical need or in the scope of the first stage of scholarly research (Kemman et al., 2014) is often the first entry point and contacts of users with the services of an infrastructure, and may have an impact on their further engagement with the infrastructure, and willingness to contribute resources, etc. Optimal alignment of the entry points of repositories and catalogues to user requirements also implicate that approaches to search activities and specific user group competences, and requirements are understood.

From the perspective of information retrieval, search activities are commonly divided into two broad categories: lookup and exploratory (Marchionini, 2006), although lookup tasks can be embedded in exploratory tasks and vice versa. Lookup search is assumed to have precise search goals, such as finding facts to answer a specific question. Exploratory search includes a variety of qualitative definitions, so this search is naturally faceted (Wildemuth and Freund, 2012), and has open-ended search goals and inexact task requirements. This approach to search activities corresponds to the functionalities of the VLO and seems to be suited as an underlying premise when designing tasks to investigate user search activities and requirements of the service.

Another factor that should be considered is the competence of users to identify information requirements, find relevant information, and evaluate the search results. Expertise plays a significant role in the approaches that users utilise to seek information. The distinction should be made between technical and domain expertise. Jenkins et al. (2003) observe that whereas each of the dimensions is valuable, users are most likely to succeed in search activities when both are present. These factors should be considered when developing tasks for a user-centred study or investigation of user requirements. Students, i.e. participants in this pilot study, are often classified as *experts-in-training* (Hagemann, 2016).

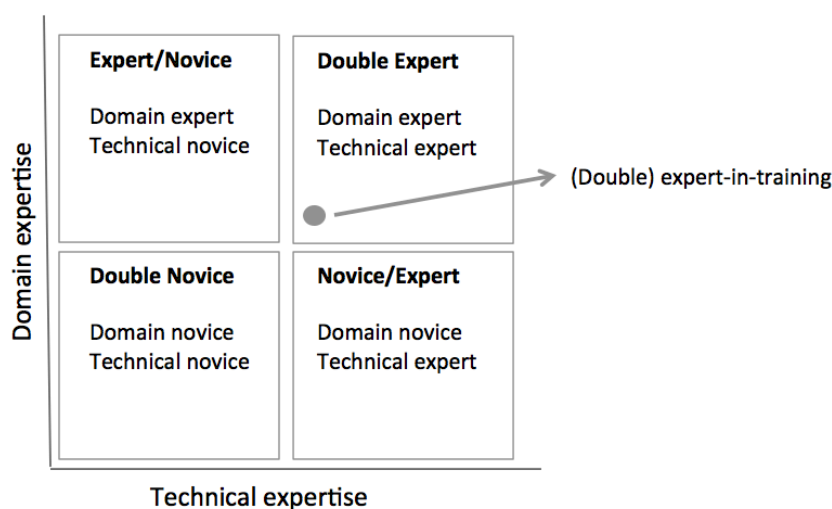


Figure 1: Two dimensions of expertise: domain and technical (adapted from Russel-Rose and Tate, 2013).

2.4 Users' search competences

Present day translation training is informed by both academic research and professional experience. The training incorporates 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. These curricula often follow one of the most exhaustive translation competence models, that of the European Master in Translation framework⁴, which covers six main competences (see Fig. 2), among others the

⁴ The network aims at promoting quality in translation training and is led by the Directorate General for Translation of the European Commission. The label is awarded only to academic translation programmes meeting admission criteria (currently 64 programmes). However, the competences developed in the framework are widely integrated also in curricula of the programmes that had not been awarded the official label. For more information, see https://ec.europa.eu/info/european-masters-translation-emt_en (accessed 1.4.2017).

information mining competence, and the technological competence. Information mining competence includes, inter alia:

- identifying information and documentation requirements,
- extracting and processing information for a given task,
- evaluating the reliability of sources,
- effectively using search engines (EMT Expert Group, 2009).

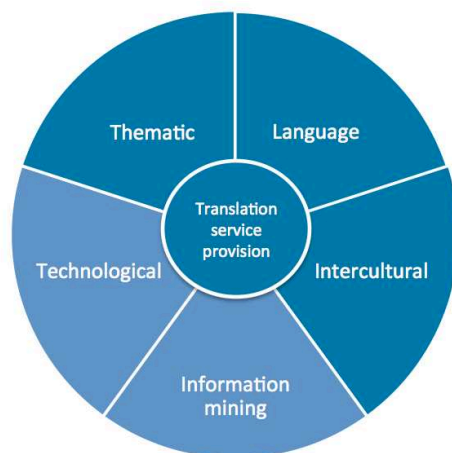


Figure 2: Competences in present translation training: Wheel of competence of the European Master of Translation (based on EMT Expert Group, 2009).

Although e-research practices (such as the role of digital curation practices) are not explicitly integrated into present day academic translator curricula, there is a substantial overlap between information mining and technological competences, and research competences, therefore translation courses can be expected to train some research competences indirectly (Vandepitte, 2013; Austermühl 2016). As translation training caters to both prospective translators-practitioners and researchers in Translation Studies, students were treated as *double experts-in-training* and investigated as users in this pilot study.

3 Pilot study with students of Translation Studies

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 matter to users in Translation Studies when they try to navigate through a research infrastructure 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 exploratory functionalities for language resources (Odijk, 2014), and to find out the perceived quality of the results in the VLO by this user group. Lastly, the objective of this pilot study was to determine 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 this user group.

3.2 Data collection setting and participating users

Data collection for this study was carried out at two Austrian universities in five courses with students of Translation Studies at the BA and MA levels during two academic years (winter semester 2015, summer semester 2016⁵, and winter semester 2016⁶). The course at the BA level is recommended for students in their 5th semester. Therefore the majority of the students participating in the study were in their final year of the BA studies. The majority of the students at the MA level participating in this study

⁵ References to the VLO refer to the version 3.3.

⁶ References to the VLO refer to the version 4.0.2.

was in their second semester, or at even in more advanced stages of their MA studies. The courses that served as the platform for this study are obligatory for all students with all working language combinations in the program. This means that we had unique access to users with a broad range of working language combinations and could, therefore, cover a spectrum of languages with various degrees of support (Rehm and Uszkoreit, 2012). The participants' working languages ranged from rather traditional combinations (e.g. German, English, French, Spanish, Polish, Italian) to Arabic and Austrian Sign Language. Details on participants' working languages are given in the next section.

All five courses focused on practical aspects of specialised translation, emulating the actual practice through the use of authentic resources, tools, assignments and processes. The format of the face-to-face sessions was designed to cover selected topics on language resources, and an introduction to e-infrastructures, repositories, catalogues and similar services. None of the participants had previous experience with searching for language resources in repositories or catalogues.

In the study design, the length of time of the study and each task needed particular attention. This was a critical issue because the participants were performing activities that took some length of time to complete. Moreover, search tasks can be exhausting both mentally and physically (Kelly, 2009), and prolonged search activities may negatively impact the results. For these reasons, as well as due to time limitations (length of an academic session, availability of students limited to one semester) and logistical limitations (number of available computer working stations), the data collection was conducted in several phases, and with a different setup and number of participating users. Each phase focused on one specific task as elaborated in the next section.

Each phase of the study was initiated by a face-to-face information session, in which the overall objectives and logistical details of the study were presented to the students by the course trainers. A written task description was also presented to participants to read until they understood it thoroughly. Participants were given time to ask questions and to opt out of the study. The observance of the anonymity of the data collected was explained and stressed. To keep the search process natural, we did not ask the participants to think aloud, but collected self-reported research results and comments as described in detail in the section below.

3.3 Task design

Task design utilised qualitative and quantitative approaches combining mainly open and to a limited extent closed tasks to operationalize exploratory search activities by double experts. This method allowed 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.

We merged tasks into three task groups (one task group per study phase), two of them with two subtasks each to address specific research questions separately (see Tab. 1). We instructed the participants to inform us when they had completed each task; however, each task had a maximal time limit for completion.

Selection criteria can be understood as usability, i.e. a set of criteria that facilitate human decision-making (Maegaard et al. 2016). Based on this set of criteria, users conduct their exploratory search for language resources. As there is a little insight into the selection criteria for language resources preferred by the investigated user group, the first task (T1) aimed at establishing the set of selection criteria, based on which the users may decide if a language resource is relevant and operationally reusable for their purpose. The participants ($n_0=25$) were given an open assignment to identify the criteria without linking it to a specific service or e-infrastructure to ensure a minimal bias towards their selection of criteria. The users were asked to provide a weighted list of their selection criteria.

The tasks T2a and T2b were based on the criteria identified in T1. The objective of T2a was to establish how many unique language resources relevant for this user group could be found in the CLARIN VLO that cannot be found in other repositories and catalogues. In this task, the participants ($n_1=25$) were asked to query portals (CLARIN VLO, META-SHARE, and ELRA), and catalogues (LT-Observe, Opus) for language resources in their working languages (Croatian, Czech, English, French, German, Italian, Polish, Romanian, Russian) based on preselected criteria, identify them and document their research results in a form. The participants were asked to search for all relevant types of languages resources: mono-, bi-, and multilingual corpora, parallel and comparable corpora, translation memories, terminological resources, lexical resources, etc. This task also included self-reporting the information

where the language resource had been found, and where the metadata had been extracted from. The participants were instructed not to valorise the metadata by cross-pollinating them from various sources or improving them by additional research, but to 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 in different portals and catalogues in task T2b (What is the perceived quality of the metadata of the LRs found in the VLO?).

Phase	Task	Task objective	Type of task	Data collected
Definition of selection criteria	T1	When searching for LRs, which criteria are essential for the user group?	Open	Criteria sets with descriptors
Identification of language resources based on selection criteria	T2a	How many (unique) LRs relevant for this user group can be found in VLO?	Semi-open (task completion based on pre-defined criteria)	Number of resources; list of identified resources with descriptors of criteria based on found metadata records
	T2b	What is the perceived quality of the metadata of the LRs found in VLO?	Closed	Values 1-5 (1 for very high quality, 5 for very low quality) ⁷
Discovering language resources in VLO	T3a	When searching for multilingual LRs in VLO, what metadata is missing according to the user group?	Open	Metadata with descriptors
	T3b	How does the user group perceive the interface and the functionalities of the VLO while searching for (multilingual) LRs?	Open	Written self-reports

Table 1: Overview of the task design with task objectives, type of task, pre-task activities, and data collected in each task.

The third task group (T3a, T3b) solely concentrated on the CLARIN VLO. The participating users were briefed on the aims and objectives of CLARIN, the main principles and search functionalities of the VLO (e.g. faceted browsing, textual queries, and advanced querying), and were given a demonstration. The third task group was conducted twice as the task was first carried out with the VLO refer to the version 3.3, which became obsolete less than a month later, and was replaced with a more sophisticated upgrade. Therefore, we decided to run the tasks T3a (When searching for multilingual LRs in the VLO, what metadata is missing according to the user group?) and T3b (How does the user group perceive the interface and the functionalities of the VLO while searching for (multilingual) LRs?) again in the VLO version 4.0.2. The participants ($n_2=14$) that performed the task in the VLO version 3.3 worked with the following languages: Arabic, Bosnian/Croatian/Serbian, English, French, German, Italian, Spanish, Russian, and Sign Language. The participants ($n_3=21$) that performed the task in the VLO version 4.0.2 worked with Bosnian/Croatian/Serbian, English, French, German, Italian, Polish, Spanish, Russian, and Welsh.

Both times the participants were asked to run a search for multilingual language resources in the VLO. They were given basic preselected search criteria to yield comparable results. In task 3a (When searching for multilingual LRs in the VLO, what metadata is missing according to the user group?) they were asked to list further categories of metadata that could 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”

⁷ The inverted score scale was adopted for practical reasons: The participants were already familiar with this scoring system as it mirrored the academic marks.

(Odijk, 2014). In addition, they were asked to provide a comment on the satisfaction with the search functionalities (T3b).

4 Results and discussion

In this section we present the results of all the tasks described in section 3, followed by a brief discussion.

4.1 Selection criteria as a basis for explanatory search in repositories

As the investigated users often need very specific types of language resources, e.g. when they work in a particular language combination or a particular domain, they need text and terminology from the domain, and language resources exactly in the languages in question, etc. The results of the task T1 thus unsurprisingly showed that the users highly valued the information on the *Language(s) covered* by the resource, but also the criterion *Representativeness of the domain* (see Fig. 3).

Related to this criterion are also two further criteria: the *Reliability of the resource* (for explanation see below) and *Up-to-dateness* of the resource. Participants described the criterion *Up-to-dateness* as the fact that a resource captures the latest data relevant to the domain, language, etc. This is deemed essential in specific domains (e.g.) legal, or after a spelling reform.

As certain language resources are preferably used directly in a computer-assisted translation tool, *Format* (e.g. tmx) was the third most identified criterion. A separate criterion *Downloadable* might appear redundant. However, it should be taken into account that the criteria should not be evaluated separately, but as a set of criteria to reflect the complexity of exploratory search activities.

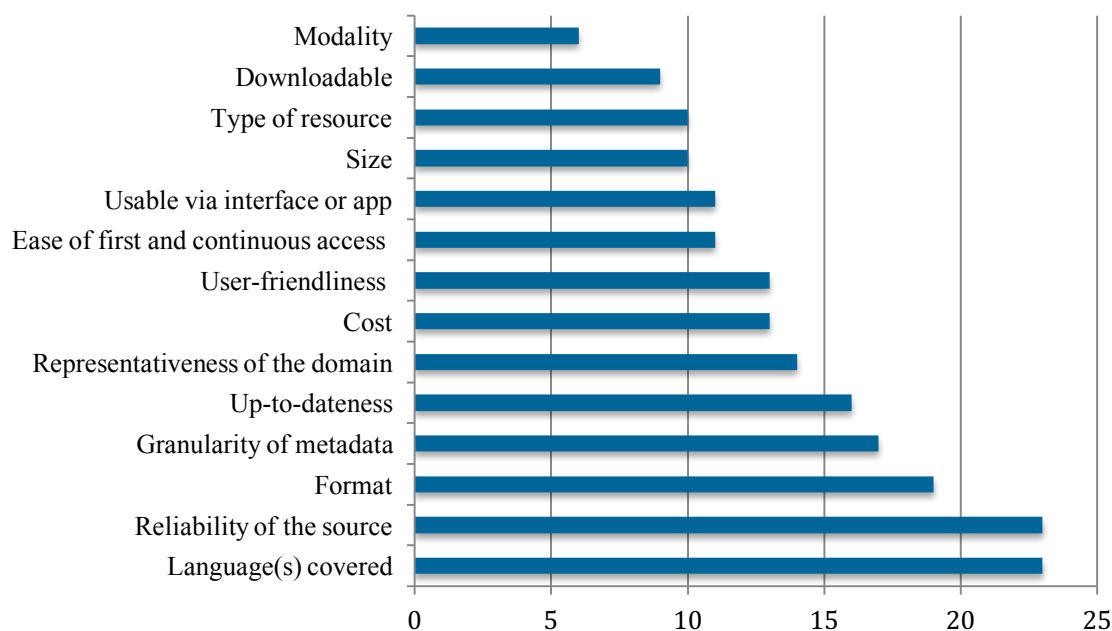


Figure 3: Selection criteria for language resources from the perspective of Translation Studies students (n=25) (results related to task T1).

It should also be emphasised that the selection criteria do not directly correspond to metadata describing a language resource. Metadata records are interpreted by the user and checked against the set of criteria to facilitate the decision-making and if needed, abort or continue the exploratory search. For example, reliability can be understood as the degree of authoritativeness of the originator and is perceived as an important indicator of the quality of the resource, especially for certain specialised translation assignments in legal, administrative, but also technical domains. In the absence of this metadata category, reliability could be derived from the combination of metadata records.

4.2 Relevant language resources and metadata

Based on the criteria defined in task T1, the participants queried repositories and catalogues (task T2a), which resulted in identifying of 210 relevant resources⁸ in total, found in all repositories and catalogues (CLARIN, META-SHARE, ELRA, LT-Observe). The raw dataset included multiple identifications of the same language resource as more than one language was queried in this task. Some multilingual resources were identified several times, for example, the resource *Europarl Parallel Corpus* fulfilled the criteria of the search for Spanish as well as for English. A subsequent clean up of the doublets covering several languages (e.g. *Europarl Parallel Corpus*, *FAO Glossary of Biotechnology for Food and Agriculture*, *JRC-Acquis Multilingual Parallel Corpus*, and others) was needed and removed 33 per cent of entries from the raw dataset. A comparison of all the identified relevant language resources in the cleaned up dataset showed that the majority of the identified resources could be found in more than one repository or catalogue. Ten language resources were uniquely found through CLARIN VLO, all but one of them being text corpora. The participants identified a wide variety of resources: monolingual, bilingual and multilingual corpora, translation memories, terminological resources, and lexical resources.

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 the VLO (task T2b) was 3. Overall, participants repeatedly assigned high quality scores for the quality of the metadata found in the catalogue LT-Observe⁹. It should be noted that this is a catalogue that consists of identified language resources for machine translation scenarios, based on evidence-based usability. The metadata in the catalogue has been validated and valorised by human experts (Maegaard, 2016).

4.3 Discovering language resources through CLARIN VLO and user requirements

The focus of the last task group was exclusively on the VLO, and bi- and multilingual resources (e.g. comparable corpora, parallel corpora, translation memories, lexical resources, terminological resources) found through the VLO. The answers of the participants were abstracted and grouped into categories (see Fig. 4 and Fig. 5)¹⁰.

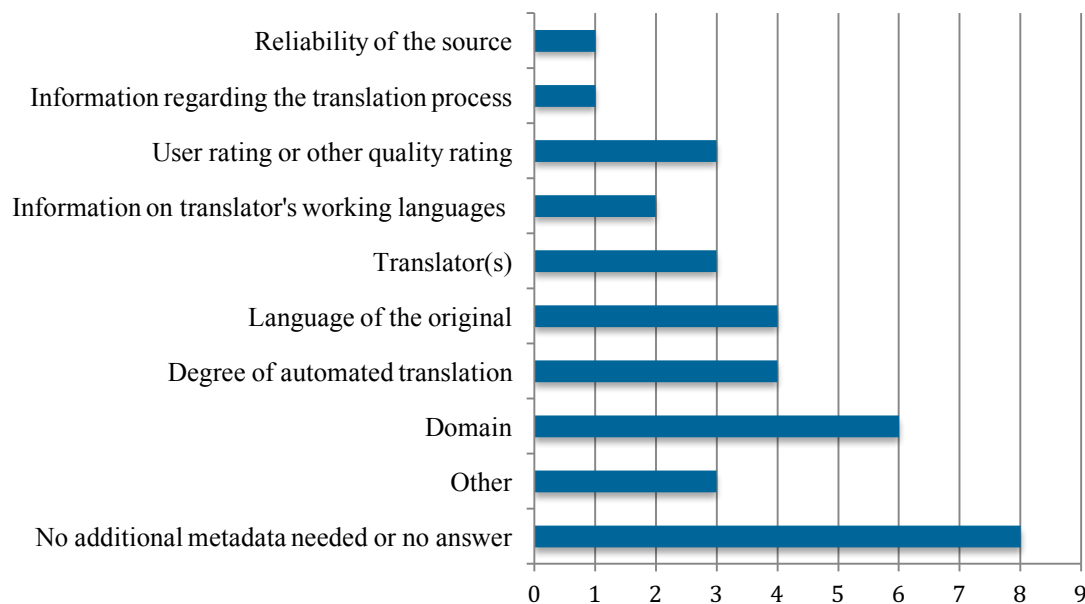


Figure 4: Desirable additional metadata for bi- and multilingual language resources from the perspective of Translation Studies students (n=35) (results related to task T3a).

⁸ Multiple selections in more than one language were possible.

⁹ <http://www.lt-innovate.org/lt-observe/resources-list>

¹⁰ Multiple answers were possible.

The most requested additional metadata information voiced by participants was *Domain* (see Fig. 4). The participants also listed additional metadata that they considered useful in relation to bi- and multilingual resources: for example, in parallel data the information what is the original text and what is the translation. Further requested information was the degree of automated translation, i.e. whether the resource had been translated by a human translator, by a human translator utilising computer-assisted translation tools, or by a machine translation system. In addition to this, the information regarding the translation process could be useful, e.g. whether the output of automated translation was post-edited, etc. Some participants would also find the information about the translator(s) (e.g. name, details about the translator’s working languages) useful. This metadata might not be applicable for all types of resources since we concentrated solely on bi- and multilingual resources in this task.

Regarding the search functionalities, the majority of participants were satisfied with the given search functionalities. Considering the importance of the selection criteria *Language* and due to the fact that the investigations in Translation Studies are usually conducted contrastively between two specific languages, the participants were not satisfied with the fact that it was not possible to search for language combinations (Fig. 5). The participants suggested that a multi-value selection for a single facet (e.g. language) would be very desirable. Regarding functional desiderata, the users suggested that functionalities that would make storing of search queries and search results possible. Furthermore, localisation of the search interface into different languages would be useful.

Interface	Faceted search	Functional requirements
multilingual	multi-value selection for a single facet	storing of search query
	domain-specific search	storing of search results

Figure 5: Desirable functional and search requirements, and interface features from the perspective of Translation Studies students (results related to task T3b).

The repetition of the task T3b (How does the user group perceive the interface and the functionalities of the VLO while searching for (multilingual) LRs?) in the VLO version 4.0.2 generated almost identical results. Therefore, the results of the repetition of this task support the initial user requirements established in VLO version 3.3 towards the interface, faceted search, and functional requirements.

4.4 Discussion

What can we learn from the obtained user search and selection criteria, users’ performance in finding relevant language resources in the CLARIN VLO, as well as their comments on the search functionalities and requirements? To a large extent, the metadata fields provided in the CLARIN VLO (language, type, genre, licence, time coverage, etc.) help and support the users to decide whether the resource is relevant and useful for their task. Since relevant language resources seem to be also explored through other portals and services, the users might prefer them over CLARIN if they establish that they suit better their user requirements. This may result in a lower user involvement and may also have an effect on the involvement of this user group as data providers.

The information about user requirements of this user groups, specifically about desiderata in terms of metadata, could be of interest for providers of bi- and multilingual language resources, especially

intrinsically translation-related language resources, such as translation memories, and would allow them to fine-tune the metadata provided for a better searchability in VLO, and thus a higher re-use of the resource.

Based on the pilot study with the students of Translation Studies, we derived to the following desiderata of the features in the VLO:

- Multi-value selection within a facet: Multi-value selection for a single facet (e.g. language) could support a better user experience for those users, who want to narrow down their search to a certain language pair. Users in Translation Studies are often interested in a contrastive comparison of language resources in a specific language combination.
- Domain-specific search: User interested in specialised translation and languages for special (specific) purposes, often need language resources from a particular domain, for instance, legal or medical domain. The VLO faceted search currently supports the search by subject. The granularity and depth of facet *Subject* vary considerably (e.g. legal documents vs. the legal prescriptions concerning hunting). As observed in this study, the search does not support users unfamiliar with the topics covered, who want to explore the resources intuitively by domain, or by classification system. Clustering of subjects into domains would encourage a more explorative approach to the discovery of resources in the VLO.
- Multilingual (localised) interface: Users searching for language resources in languages other than English may not have a strong command of English, or may simply be inclined to search in the language(s) of the potential language resource. This seems to occur more often when searching for a multilingual language resource (e.g. Italian-German terminological resource) in languages other than English.
- Multilingual (localised) metadata: Multilingual metadata in languages other than English would support and complement the multilingual search interface. Multilingual metadata would be presumably needed for those language resources, in which English is not one of the languages of the resource. Reusing the data by national projects and national consortia, as well as automatization of the process, could support this endeavour.
- Storing of search queries or search preferences: This feature could support repetitive search sessions, e.g. for comparative or didactic purposes, but would require a user account.¹¹
- Storing of search results: This feature could support a repetitive or on-going discovery of language resource in the VLO but would require a user account.¹²
- Valorisation of metadata: Awareness for the multifaceted needs of various user groups should be raised among data providers. For instance, the information what is the original and what is the translation would be valuable.

Since the completion of the pilot study, new features have been introduced in the VLO version 4.1.0¹³. It is now possible to select multiple values within a facet, thus broadening the selection with the operator *OR*, or narrowing down the selection with the operator *AND*. Broadening the selection with the operator *OR* is the default setting of the multiple value selection behaviour¹⁴.

As deduced from the pilot study, the new feature will greatly improve the usability for users searching for language resources in a certain language pair or language combination. From the perspective of the users in Translation Studies, the search could be optimised by the default operator *AND*, as users typically want to narrow down their search to a specific language combination. Alternatively, the preference for the default setting could be stored in search queries or search query preferences as suggested above. We would also suggest making the advanced search feature more prominent (e.g. position it on the top of the interface) as it has a major impact on the user experience and the users' satisfaction with the search functionalities.

¹¹ Regarding sharing of the search query, the current version of the VLO 4.1.0. supports the following: bookmarking the search query, copying the link and sharing the link via email.

¹² Regarding sharing of the search results, the current version of the VLO 4.1.0. supports the following: bookmarking the search query, copying the link and sharing the link via email.

¹³ <https://vlo.clarin.eu/about> (accessed 8.4.2017).

¹⁴ <https://vlo.clarin.eu/help> (accessed 8.4.2017).

In addition to the suggested features for the VLO outlined above, outreach activities designed for users in Translation Studies (students, researchers, trainers, and practitioners) would help to make the VLO service and other CLARIN services more known in the community. A wider recognition of the CLARIN services among the Translation Studies community could result in bringing a new user group on board as well as gaining active providers of language resources.

5 Conclusion and outlook

This pilot study addressed the needs of the students of Translation Studies as prospective translators and researchers in Translation Studies, focusing primarily on Corpus-based Translation Studies and Computational Translation Studies, as one of the user groups of the CLARIN service VLO. Their assessment of the gaps in terms of the usability of the service was investigated, and suggestions were made for possible optimisation. It was established that the resources found through the VLO would need some additional metadata information, especially bi- and multilingual language resources, in order to be better suited for reuse by researchers, trainers, and students in Translation Studies. 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 data providers. This especially applies in cases, in which the resources provided had been generated by translators, translation scholars and translation students, to ensure a higher uptake of the VLO service as well as other CLARIN services by this user group. Moreover, outreach activities tailored for users in Translation Studies would help to make the VLO service and other CLARIN services more known in the community and would help to gain a user group as well as a data provider group.

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

The study also discussed and implemented considerations and preparation that should be taken into account when designing tasks for exploratory search in the VLO or similar services. User requirements of similarly under-represented user groups could take this pilot study as their departing point. In addition to task completion and self-reporting, further avenues of user search behavior could be explored to arrive at precise and complementary data, such as task completion time, query time, query length, scroll depth, cumulative click, etc. to better understand users' search activities and adapt the CLARIN services to evidence-based user requirements.

Acknowledgements

This work crucially uses data, tools, and services made available through the CLARIN infrastructure.

References

- [Austermühl2016] Frank Austermühl. 2016. Recherche und Arbeitsmittel. In Mira Kadric and Klaus Kaindl (eds.), *Berufsziel Übersetzen und Dolmetschen: Grundlagen, Ausbildung, Arbeitsfelder*. Tübingen: Gunter Narr, 200–217.
- [Baker1993] Mona Baker. 1993. Corpus Linguistics and Translation Studies: Implications and Applications. In Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds), *Text and Technology: In Honour of John Sinclair*. Amsterdam, John Benjamins, 233–250.
- [Baker1995] Mona Baker. 1995. Corpora in Translation Studies: An Overview and Suggestions for Future Research. *Target*, 72, 223–244.
- [Beeby et al2009] Allison Beeby, Inés Patricia Rodríguez, and Pilar Sánchez-Gijón (eds.). 2009. *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. John Benjamins Publishing.
- [Bowker1998] Lyanne Bowker. 1998. Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. *Meta*, 43.4, 631–651.

- [Bowker2002] Lyanne Bowker and Jennifer Pearson. 2002. *Working with Specialized Text: A Practical Guide to Using Corpora*. Routledge.
- [Bowker2004] Lyanne Bowker. 2004. Corpus Resources for Translators: Academic Luxury or Professional Necessity?. *TradTerm*, 10, 213–247.
- [Budin2015] Gerhard Budin. 2015. Digital Humanities, Language Industry, and Multilingualism – Global Networking and Innovation in Collaborative Methods. In Martin Forstner and Hannelore Lee-Jahnke (eds.), *CIUTI-Forum- 2014*. Boston, USA: Peter Lang, 423–448. DOI: <http://dx.doi.org/10.3726/978-3-0352-0290-8>.
- [Chunpir et al2015] Hashim Iqbal Chunpir, Thomas Ludwig, and Dean N. Williams. 2015. Evolution of E-Research: From Infrastructure Development to Service Orientation. In Aaron Marcus (ed.), *Design, User Experience, and Usability: Interactive Experience Design: 4th International Conference, DUXU 2015, Proceedings*. Cham: Springer, 25–35. <http://dx.doi.org/10.1007/978-3-319-20889-3>.
- [Ehrensberger-Dow et al2016] Maureen Ehrensberger-Dow, Andrea Hunziker Heeb, Gary Massey, Ursula Meidert, Silke Neumann, and Heidrun Karin Becker. 2016. An International Survey of the Ergonomics of Professional Translation. *ILCEA Revue de l'Institut des Langues et des Cultures d'Europe et d'Amérique*, 27.
- [EMT Expert Group2009] EMT Expert Group. 2009. *Competences for Professional Translators, Experts in Multilingual and Multimedia Communication*.
- [Fantinuoli and Zanettin.2015] Claudio Fantinuoli and Federico Zanettin. 2015. Creating and Using Multilingual Corpora in Translation Studies. In Claudio Fantinuoli and Federico Zanettin (eds.), *New Directions in Corpus-based Translation Studies*. Berlin: Language Science Press, 1–10.
- [Gallego-Hernández2015] Daniel Gallego-Hernández. 2015. The Use of Corpora as Translation Resources: A Study Based on a Survey of Spanish Professional Translators. *Perspectives: Studies in Translatology*, 23.3, 375–391.
- [Granger2003] Sylviane Granger. 2003. The Corpus Approach: A Common Way forward for Contrastive Linguistics and Translation Studies. In Sylviane Granger, Jacques Lerot and Stephanie Petch-Tyson (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam & Atlanta: Rodopi, 17–29.
- [Hagemann2016] Susanne Hagemann. 2016. (Non-)Professional, Authentic Projects? Why Terminology Matters. In Don Kiraly (ed.), *Towards Authentic Experiential Learning in Translator Education*. Mainz University Press, 33–51.
- [Jenkins et al2003] Christine Jenkins, Cynthia L. Corritore, and Susan Wiedenbeck. 2003. Patterns of Information Seeking on the Web: A Qualitative Study of Domain Expertise and Web Expertise. *IT & Society*, 1.3 (2003), 64–89.
- [Kelly2009] Diane Kelly. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends in Information Retrieval*, 3 (1–2), 1–224.
- [Kemman et al2014] Max Kemman, Stef Scagliola, Franciska de Jong, and Roeland Ordelman. 2014. Talking with Scholars: Developing a Research Environment for Oral History Collections. In Łukasz Bolikowski, Vittore Casarosa, Paula Goodale, Nikos Houssos, Paolo Manghi, and Jochen Schirrwagen (eds.), *Theory and Practice of Digital Libraries -- TPD 2013 Selected Workshops*. Springer International Publishing, 197–201.
- [Kenny2007] Dorothy Kenny. 2007. Translation Memories and Parallel Corpora: Challenges for the Translator Trainer. In Dorothy Kenny and Kyongjoo Ryou (eds.), *Across Boundaries: International Perspectives on Translation Studies*. Cambridge Scholars Publishing, 192–208.
- [Krüger2012] Ralph Krüger. 2012. Working with Corpora in the Translation Classroom. *Studies in Second Language Learning and Teaching*, 4, 505–525.
- [Kübler2003] Natalie Kübler. 2003. Corpora and LSP translation. In Federico Zanettin, Silvia Bernardini and Dominic Stewart (eds.), *Corpora in Translator Education*. Manchester: St Jerome, 2003, 25–42.
- [Laviosa2002] Sara Laviosa. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam – New York, NY: Rodopi B.V.
- [Maegaard et al2016] Bente Maegaard, Lina Henriksen, Andrew Joscelyne, Vesna Lušicky, 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*, 449–456.

- [Maia2003] Belinda Maia. 2003. Some Languages are More Equal than Others. Training Translators in Terminology and Information Retrieval Using Comparable and Parallel Corpora. In Federico Zanettin, Silvia Bernardini and Dominic Stewart (eds.), *Corpora in Translator Education*. Manchester: St Jerome, 43–53.
- [Marchionini2006] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Communications of the ACM - Supporting Exploratory Search*, 49(4), 41–46.
- [Mellange2006] Mellange. 2006. MeLLANGE Corpora & E-learning Survey Results. *ITI Bulletin*.
- [Odijk2014] Jan Odijk. 2014. *Discovering Resources in CLARIN: Problems and Suggestions for Solutions*. Utrecht University Repository, Netherlands.
- [Rehm and Uszkoreit2012] Georg Rehm and Hans Uszkoreit. 2012. *Europe's Languages in the Digital Age. White Paper Series*. Springer.
- [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.
- [Robson and McCartan2016] Colin Robson and Kieran McCartan. 2016. *Real World Research*. Chichester, West Sussex: Wiley.
- [Russell-Rose and Tyler2013] Tony Russell-Rose and Tate Tyler Tate. 2013. *Designing the Search Experience: The Information Architecture of Discovery*. Morgan Kaufmann Publishers Inc.
- [Sánchez-Gijón2002] Pilar Sánchez-Gijón. 2002. Aplicaciones de la Lingüística de Corpus a la Práctica de la Traducción. Complemento de la Traducción Asistida por Ordenador. *Terminologie et Traduction*, 2, 84–106.
- [Van Uytvanck et al2010] Dieter Van Uytvanck, Claus Zinn, Daan Broeder, Peter Wittenburg, and Mariano Gardelloni. 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*, 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*, 1029–1034.
- [Vandepitte2013] Sonia Vandepitte. 2013. Research Competences in Translation Studies. *Babel*, 59 (2), 125–148. DOI:10.1075/babel.59.2.
- [Wildemuth and Freund2012] Barbara M. Wildemuth and Luanne Freund. 2012. Assigning Search Tasks Designed to Elicit Exploratory Search Behaviors. *Human Computer Information Retrieval (HCIR) Symposium*, 4, 1-10.
- [Wilkinson2005] Michael Wilkinson. 2005. Using a Specialized Corpus to Improve Translation Quality. *Translation Journal*, 9.3.
- [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> (accessed 1.4.2017).
- [Zanettin1998] Federico Zanettin. 1998. Bilingual Comparable Corpora and the Training of Translators. *Meta: Journal des Traducteurs / Meta: Translators' Journal*, 43, 616–630.
- [Zanettin2012] Federico Zanettin. 2012. *Translation-Driven Corpora Corpus Resources for Descriptive and Applied Translation Studies*. Taylor and Francis.