

# TeLeMaCo—A Collaborative Repository for Training and Teaching Materials in Linguistics

Marilisa Amoia, Hannah Kermes, Jörg Knappen, José Manuel Martínez, Elke Teich, Mihaela Vela

Department of Applied Linguistics, Saarland University  
{m.amoia,h.kermes,j.knappen,j.martinez,e.teich,m.vela}@mx.uni-saarland.de

**Abstract.** This paper presents ongoing work on a collaborative portal and repository for training and teaching materials in linguistics. It is implemented as a web service and allows access to a wide range of multimedia materials, including text, powerpoint presentations and videos. Supporting semantic tagging of learning contents, our framework allows for fine-grained structured queries and user-defined ranking criteria.

**Keywords:** Humanities and Social Science, Language Resource Repository, Digital Infrastructure

## 1 Introduction

Nowadays large amounts of data, resources and tools are available on the web for many specialized domains in principle accessible for researchers. However, since academic resources are distributed over several sites that are not necessarily interlinked and are often not represented in a common standard, gathering relevant resources still remains a cumbersome task for academic users.

As the range of academic materials available from the internet is steadily growing, there is an increasing awareness that discipline-specific “social networks” should be built up that support the exchange of scholarly materials. Here, some disciplines are more advanced than others in setting up such support. For instance, while the field of computational linguistics is fairly well equipped with platforms such as the ACL web registry or the ELRA catalogue and many others [5], the broader field of linguistics cannot draw upon a similar wealth of digital sources.

In recent years, there have been selected efforts to build digital repositories for linguistic purposes, such as The Language Archive (TLA) [3] or the Virtual Language Observatory (VLO) [6]. Both are platforms allowing to conveniently manage both collection and retrieval of large amount of linguistic resources and tools based on a common standard for metadata representation. For linguistic processing, there is, e.g., the WebLicht platform [4] that further provides facilities for text and speech processing and supports customization of processing pipelines ranging from annotation to parsing and semantic analysis. However, while it becomes much easier to find language resources (e.g., corpora) and tools (e.g., taggers, parsers) and to get some support in processing (as exemplified by WebLicht), it remains relatively difficult to find relevant training and

learning materials. While such materials are potentially available from the web, there are only few convenient portals that provide a centralized access to them.

In this paper, we present a system that aims at a centralized repository and portal for learning and teaching materials dealing with the application of language resources and tools to be used by linguists. We describe the basic functionalities of our system as well as the schema we employ for annotating training and teaching materials with metadata. Finally, we discuss the use of our system, focusing on search/query and conclude with a brief comparison with similar tools that have been developed in other fields.

## 2 TeLeMaCo

Our portal (*Teaching and Learning Material Collection*)—TeLeMaCo—provides access to the following types of material related to language resources and tools: technical documentation (including tutorials, quick starts etc), learning material for self study, short teaching modules (2–4 hours) that can be integrated in existing courses and full courses covering a broader spectrum of language resources and tools or focusing on specific topics of application of language resources and tools. We envision a hub for resources that are freely available or available to the academic community (e.g., in an identity federation). We do not target commercially available materials or services. Registered users may both contribute descriptions of learning and teaching materials they make available on the web and/or use TeLeMaCo to find relevant resources as well as provide feedback on the ones already referenced.

The prerequisite for successful search in a data repository is a meaningful annotation, which is true of course also for the types of content that we are concerned with here [1]. The categories of annotation used in TeLeMaCo are motivated from a teaching

Title:	CCP Query Language Tutorial
Institution:	
Pub year:	
Bittext type:	
Bittext:	
Language:	english
Objective:	Learning to use the CCP query processor
Type:	Manual
Target audience:	Users
Media:	PDF
Url:	http://cwb.sourceforge.net/files/CCP_TL
Owner:	jmartinez
Prerequisites:	Linguistics, command line use, regular ex
Level:	Beginners
Licence:	copyrighted
Access:	open
Old description:	This tutorial is essential reading for all CWB users, and serves as a comprehensive user manual for the CCP query processor (as an

(a) Metadata Annotation

### Feedback to EXMARALDA Partitur-Editor Handbuch Version 1.5.1

\* Thomas Scheidt \*

Keywords: Spoken Language, Transcription, EXMARALDA, Partitur Editor  
[http://www1.uni-bamberg.de/exmaralda/Use/PartiturEditor\\_Handbuch.pdf](http://www1.uni-bamberg.de/exmaralda/Use/PartiturEditor_Handbuch.pdf)  
 Dieses Handbuch beschreibt das EXMARALDA Partitur-Editor in seiner aktuellen Version (1.5.1 vom Oktober 2011). Was Sie den Editor zum ersten Mal benutzen, ist Ihnen eine Hilfe gegen dieses Handbuch zu lesen und bei der Einrichtung zu konsultieren.

**Feedback**

Was die described material relevant for me? No, not at all  Yes, very much

I think, the level of the described material is ... Basic  Expert

I reached the objectives given by studying the material Yes, we are all  Completely

I think, the prerequisites are ... Greatly wrong  Accurately stated

### Resource details

Institution: IDS Mannheim  
 Year of publication: 2011  
 Language: german  
 Type: Manual  
 Audience: Linguists, Social Sciences  
 Level: Intermediate  
 Prerequisites: None  
 Media:  
 Objective: Complete manual for the EXMARALDA Partitur Editor  
 License: open  
 Access: open  
 Creation date: Tuesday, 29 January 2013 16:55:48

(b) A Query Result

Fig. 1: TeLeMaCo

and learning perspective and include:

- standard categories like *author*, *title*, *publication year*, and *URL*;

- pedagogically relevant features, e.g. *prerequisites*, *target audience*, *objectives*, *level*;
- bibliography facilities, e.g., *bibtex*;
- topic-relevant features, such as *description*;
- technical features, e.g., *language* and *media type* of the resource;
- accessibility information such as *license* and *access*.

The values of these features are directly assigned by the author of the learning content. Figure 1a shows the GUI used in TeLeMaCo for metadata annotation. A resource may be associated with user/learner feedback (see Figure 1b) that directly addresses these categories, in particular the ones relevant from learning or pedagogical perspectives. A user can thus express her/his opinion about a resource by answering questions like *What level has the resource in your opinion?* or *Did you reach the objectives given?* Furthermore, each resource is associated with a list of keywords.

In TeLeMaCo, we have designed an annotation scheme accounting for subject/topic classification as well as for pedagogical criteria. Criteria such as “level” and “prerequisites” can help users to decide which resource is appropriate for them with regard to their specific skills, while the field “objective” can help the user to judge whether the resource will provide the information he needs. The annotation scheme we designed is very close to that proposed by the Learning Resource Metadata Initiative (LRMI<sup>1</sup>). We plan to link the annotation scheme of TeLeMaCo to the Component Metadata Framework (CMDI) [2] that is being implemented within the CLARIN community. This will allow interoperability and an easy integration of our system into linguistic repositories such as the above mentioned TLA and VLO.

The database of TeLeMaCo can be queried through a web interface that allows fine-grained queries. So for instance, the user can retrieve learning material by setting search criteria such as author name, date, version, but also topic, difficulty level of the learning content or preferred media and whether the resource is open or requires a license.

In the future, we plan to automatically extract additional topic-relevant keywords from the resource description to be able to expand queries with synonyms and semantically related concepts and thus improve search results.

### 3 Discussion

There are indeed many platforms that are similar to TeLeMaCo in that they provide facilities for searching, sharing and managing open-resources for scholars and/or for learning purposes. For instance *Zotero*<sup>2</sup>, a tool that can be used as add-on to the browser for searching/managing automatically annotated resources that are saved locally. Academic social networks such as *academia.edu* and *Researchgate* or platforms for searching and sharing open educational resources such as *Jorum*<sup>3</sup> that is maintained by UK Further and Higher Education community. However, these solutions differ from TeLeMaCo in that they either do not have a collaborative aspect or do not have a pedagogical aim or do not provide mechanisms for evaluating the (pedagogical)

<sup>1</sup> <http://www.lrmi.net/>

<sup>2</sup> <http://www.zotero.org/>

<sup>3</sup> <http://www.jorum.ac.uk/>

quality of the shared resources, respectively. Further, most of such platforms structure resources in terms of subject/topic. On the contrary, TeLeMaCo will not only be a resource for sharing resources in a specialized domain, but rather a structured repository of specialized information with a pedagogical aim. The resources in TeLeMaCo are organized and annotated following subject/topic criteria as well as pedagogical criteria. Here the collaborative aspect of TeLeMaCo comes into play. The assignment of a pedagogical value to each resource is a social, collaborative process. It is started by the author of the resource and can be modified by the users, that want to use the resource. TeLeMaCo explicitly asks users for their opinion on a resource. In the current implementation, TeLeMaCo does not allow users to share a personal set of resources or to communicate with each other. This is certainly an interesting issue to be addressed in future.

## 4 Conclusion

We have described a portal for teaching and learning materials for linguistic applications of language resources and language processing tools that is based on a repository providing annotation categories for such materials. By providing annotation categories which allow to specify queries more precisely, we want to make teaching and learning material better and more easily accessible. We consider our system a complement to existing linguistic infrastructure components such as the VLO or WebLicht. Populating TeLeMaCo is a community exercise, of course. The system is currently in the testing phase in the CLARIN-D project and is not yet freely accessible on the web. We have started a test phase involving 12 academic institutions. The preliminary feedback we received shows that academics and scholars consider the system useful for personal research purposes and acknowledge that students will benefit from an integration of our platform in regular university courses.

## References

1. Bärenfänger, M., Binder, F., Lobin, H., Lungen, H., (editors), M.S.: Language resources and technologies in learning and teaching. *Journal for Language Technology and Computational Linguistics* 26(1) (2011)
2. Broeder, D., Schonefeld, O., Trippel, T., Uytvanck, D.V., Witt, A.: A pragmatic approach to XML interoperability – The Component Metadata Infrastructure (CMDI). In: *Balisage: The Markup Conference* (2011)
3. Drude, S., Broeder, D., Trilsbeek, P., Wittenburg, P.: The language archive: A new hub for language resource. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. Istanbul, Turkey (2012)
4. Hinrichs, M., Zastrow, T., Hinrichs, E.: Weblicht: Web-based LRT services in a distributed escience infrastructure. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valetta, Malta (2010)
5. Lemnitzer, L., Hinrichs, E., Witt, A.: Language resources, taxonomies and metadata. In: *Proceedings of the Text Mining Services - TMS Conference*. Leipzig, Germany (2009)
6. Van Uytvanck, D., Stehouwer, H., Lampen, L.: Semantic metadata mapping in practice: the virtual language observatory. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. Istanbul, Turkey (2012)