

**D-SPIN Report R6.1:
Training materials and guidelines for
their application**

March 2010

D-SPIN, BMBF-FKZ: 01UG0801A

Deliverable: R6.1: Training materials and guidelines for their application

Editors: Svetlana Ahlborn and Frank Binder

Contents

Contents.....	3
List of figures	4
List of tables	4
1. Scope and purpose of this document.....	5
2. Introduction and (European) context.....	5
2.1. Strategic questions on training in the context of CLARIN	5
2.1. Placing D-SPIN's training activities	9
3. University courses: Two examples	11
3.1. Course "LRT for linguists", Giessen University, Summer 2009	11
3.1.1. Course participants	11
3.1.2. Prerequisites	11
3.1.3. Course elements	12
3.1.4. Course topics	12
3.1.5. E-learning elements.....	16
3.1.6. Course evaluation	17
3.1.7. Résumé.....	17
3.2. Goethe University Frankfurt Main: introductory course "Corpus linguistics"	18
3.2.1. Course participants	18
3.2.2. Course topics	18
3.2.3. Learning scenario	20
3.2.4. Module access approaches	21
3.2.5. Online exercises (with screenshots)	22
3.2.6. E-Learning modules	24
3.2.7. Course access information	25
3.2.8. Résumé	25
4. Creating e-learning materials	27
4.1. Planning an authoring pipeline.....	27
4.2. Choosing an authoring tool	27
5. Summary and outlook	28
6. Reports	29
7. References	30

List of figures

Figure 1: Mapping academic training activities according to scope and level	9
Figure 2: Placing D-SPIN's training activities on the map	10
Figure 3: Learning scenarios: The POP model	20
Figure 4: "Learning unit" and "module"	20
Figure 5: Multiple-select exercise with interactive elements	22
Figure 6: Input text exercise	22
Figure 7: Input text exercise	23
Figure 8: Drag&drop exercise	23

List of tables

Table 1: Participants of the course "LRT for linguists", Giessen University, Summer 2009..	11
Table 2: Session topics for course "LRT for linguists" at Giessen university	13
Table 3: Course readings for "LRT for linguists", Giessen University, Summer 2009.....	14
Table 4: Student presentations in "LRT for linguists", Giessen University, Summer 2009....	15
Table 5: E-learning elements for "LRT for linguists", Giessen University, Summer 2009	16
Table 6: Participants of the course "Corpus Linguistics", Frankfurt University	18
Table 7: Course topics for course "Corpus linguistics" at Frankfurt University	19
Table 8: Module access approaches	21
Table 9: E-Learning modules, Frankfurt University fall/winter 2009/2010	25

1. Scope and purpose of this document

This report covers the activities of D-SPIN work package 6 (WP6) “training and education”. Work in WP6 began in August/September 2008. This document, which is one of D-SPIN’s M24 deliverables, captures the state as of March 2010.

The document focuses on training activities and the creation of training materials.

The aims of this document are:

- To provide examples of how LRT-training can be integrated into university curricula by presenting two different courses on using language resources, held for different student groups at the universities of Gießen and Frankfurt;
- To describe the training materials that were used for these courses;
- To present the authoring process that was used to create the training materials
- To give access to the materials.

2. Introduction and (European) context

D-SPIN WP6 “Training and education” is a user-oriented work package of D-SPIN, Germany’s national complement to CLARIN (<http://www.clarin.eu>). Our focus does not lie on training the D-SPIN staff or technical community, but rather on training users of language resources and technology (LRT).

Since training activities involve interaction with the users of language resources, they can establish an important communication channel for the development of the D-SPIN / CLARIN infrastructure. Despite the potential of training activities as a communicative channel between CLARIN and its user community, CLARIN itself currently does not conduct any training activities (cf. Binder & Cristea 2009). Consequently, the national complements, such as D-SPIN, play an important role in that respect.

Nevertheless, there are ongoing related training activities on a European / international level. In 2009, the initial training network CLARA (Common Language Resources and their Applications) has been established, aiming at early stage researchers (PhD students) and experienced researchers (young postdocs). Two of the D-SPIN partners, the Max-Planck-Institute for Psycholinguistics in Nijmegen and the Eberhard Karls University Tübingen, are participating in CLARA.

2.1. *Strategic questions on training in the context of CLARIN*

Training has been an issue at the CLARIN consortium meeting in Barcelona in May 2009. A session was devoted to the topic, as reported in the CLARIN newsletter #6 (cf. Binder & Cristea 2009). In order to stimulate the preparation of this session, Steven Krauwer (CLARIN coordinator) and Dan Cristea (CLARIN WP6 coordinator) circulated a number of strategic questions. These 13 questions are listed below, along with suggestions from our current perspective.

Q1: Do we see training and education as a CLARIN-EU responsibility or do we see it as the responsibility of the local research communities in the participating countries?

Within the preparation phase of CLARIN, training is seen as a task under the responsibility of the participating countries and their research communities, as indicated by the fact that training is not covered through CLARIN’s funding.

Q2: What is the best way to make our target audience aware of what CLARIN has to offer and to convince them to use it?

With this document, we show the way we went with D-SPIN's training efforts. We do not want to claim that we are on the "best way", but a start has been made. There is much room for improvement, and many things left to be done.

Q3: What are the instruments we have at our disposal, e.g. text books, journals, showcases, publicity in the media, webcourses, tutorials at humanities conferences, annual summer schools, special grants, helpdesks (for students and/or for researchers), centers of expertise in specialized areas?

In D-SPIN we conduct university courses that shall serve as a starting point within the German community. These courses utilize standard text books, which are available in German.

Furthermore we are creating e-learning materials as web based training for users of language resources. It is not feasible for us to write in-depth tutorials for all potentially interesting language resources in the community while at the same time planning and conducting our training activities. Hence, resources and sites that already provide training materials, such as tutorials, are included in our training activities.

A particularly important event will be the D-SPIN Summer School 2010, organized for the German community of (potential) users of language resources. Note that the summer school is not covered in this document. Information regarding the summer school is available at <http://www.dspin-sommerschule.de> (in German).

Developing a help desk concept is a further goal of D-SPIN WP6.

Q4: How can we enter humanities curricula?

In D-SPIN this is done by providing example courses and seminars on using language resources at two partner sites. Since five of D-SPIN's nine partners are universities, LRT-related topics are covered in courses at these partner sites as well. However, D-SPIN has not systematically gathered information on these activities and they are not within the scope of this report.

Q5: Should it be part of the bachelor's or master's programme?

This depends on the local context and curricula. The two example courses described below have different target groups. While the Giessen course is offered for bachelor and master students, the Frankfurt course was offered generally for researchers, many of them post docs in linguistics.

Q6: Should we develop common curricula (or components)?

The courses described below may serve as example components of curricula. They are, however, conducted in German. Given their dependence on German resources and text books, presumably they cannot be easily ported to other countries, unless the prerequisites are fulfilled there as well.

Q7: Would it be good to have a scheme whereby teachers (and/or) students can move around in Europe?

Independent of language resources, that is to say in general, this seems to be covered by the Erasmus Mundus programmes both for students and teachers¹ (life long learning programme², staff teaching assignment³), and several kinds of training actions under the 'People' Specific Programme⁴, of which Marie Curie actions, such as the 'initial training networks' figure prominently.

The CLARA initial training programme for early stage researchers is one example, where some of CLARIN's partners receive support for LRT-related training activities on a European level.

Examples of existing European master programmes are

- 1) The European Masters in Language and Communication Technologies⁵ with its partners University of Saarbrücken, Charles University in Prague, University of Groningen (Language Technology) and Free University of Bolzano-Bozen, University of Nancy 2, University of Malta, University of Melbourne, Shanghai Jiao Tong University (Computer Science)
- 2) The European Master in Lexicography (EMLex)⁶ with its partners Friedrich-Alexander-Universität Erlangen-Nürnberg, Universidade Nova de Lisboa, Károli Gáspár Református Egyetem Budapest, Univerza v Mariboru, Universität Stuttgart, Nancy-Université, Universitat Pompeu Fabra Barcelona, Göteborgs universitet, Universidade Santiago de Compostela, Institut für Deutsche Sprache Mannheim.
- 3) The European Masters in Language and Speech⁷ (funding period has ended)

Q8: Should we aim at the formulation of minimum criteria and maybe even issue certificates to students who have demonstrated that they have the skills?

Up to now, this has not been in the focus of D-SPIN activities. But the previous and current Master programmes could provide relevant information on what they regard as minimum criteria, key skills and required capabilities.

¹ ERASMUS for Staff – weaving a web of knowledge and contacts:

http://ec.europa.eu/education/erasmus/doc1059_en.htm

² European Commission - Education & Training - lifelong learning programme:

http://ec.europa.eu/education/lifelong-learning-programme/doc78_en.htm

³ ERASMUS Staff Mobility - Teaching Assignments by Higher Education Institutions Teaching Staff and by Invited Staff from Enterprises: http://ec.europa.eu/education/erasmus/doc1067_en.htm

⁴ European Commission: CORDIS FP7: 'People' Specific Programme:

http://cordis.europa.eu/fp7/people/home_en.html

⁵ European Masters Program in Language and Communication Technologies (LCT) <http://lct-master.org/> and Master Courses: Erasmus Mundus - Action 1 - Erasmus Mundus Masters Courses (EMMCs) | EACEA:

http://eacea.ec.europa.eu/erasmus_mundus/results_compendia/selected_projects_action_1_master_courses_en.php

⁶ European Master in Lexicography (EMLex) <http://www.emlex.phil.uni-erlangen.de/>

⁷ European Masters in Language and Speech (funding period has ended):

<http://www.cstr.ed.ac.uk/emasters/index.html>

Q9: How do we define 'success' in education?

This question could not yet be covered by our activities at D-SPIN. The evaluation of training programmes is clearly a longitudinal process. According to Kirkpatrick (2006), success of training programs should be measured along different time scales, at the 4 levels of 1) reaction, 2) learning, 3) behavior, and 4) results. Horton (2006) shows how evaluation at these levels can be applied to e-learning activities.

Q10: How can we put our hands on existing training material?

So far, this question has not been in the focus of D-SPIN's training activities. Possibly this could be done by contacting existing training programmes, such as the Masters programmes mentioned above, as well as by contacting Professors from the national research communities, i.e. through workshops on academic LRT-training.

Q11: Are we aware of any existing stimulating success stories in the use of LRT and LRT services in research and education?

Such "success stories" of using LRT / LRT services in research and education are a good way of stimulating and motivating potential users. Many well established resources and tools can be seen as success stories in that sense. Their success is also reflected in publications of studies that make use of the resources. A very accessible area is that of corpus-based linguistics and its applications, which we use in our courses. As regards other areas of humanities, their respective experts and research communities might provide further examples of successful use of LRT and LRT services.

Q12: How do we deal with languages (everything in English, each in his own language)?

D-SPIN's training activities are currently carried out in German. There will be some mixed-language courses (English and German) at the D-SPIN summer school 2010. But apart from that, it is important that users be able to use their resources in their native languages, since it arguably allows for a wider audience. In short, localization is an issue.

Q13: What does education cost, and what should the education budget be for the construction and exploitation phase of CLARIN?

Budget reports could be collected for training activities. As regards D-SPIN, workshops or the summer school are covered by separate budgets, not as part of e.g. training material construction. Furthermore, information on personnel costs and work plans could be enquired from projects that include training activities.

As can be seen from these preliminary suggestions to these strategic questions, much work remains to be done, also on CLARIN's European level.

The following sections focus on D-SPIN's activities in the training domain.

2.1. Placing D-SPIN's training activities

In this subsection we document our perspective on academic LRT training and contextualize the D-SPIN WP6 activities. We consider two distinct types of academic training activities: university programmes and summer schools.

University programs

Summer Schools

Figure 1 shows how existing activities can be distinguished with respect to their scope (local, national, international), primary course language, and their academic level (Bachelor, Master, PhD). Figure 2 indicates the areas in which D-SPIN operates and places its offers.

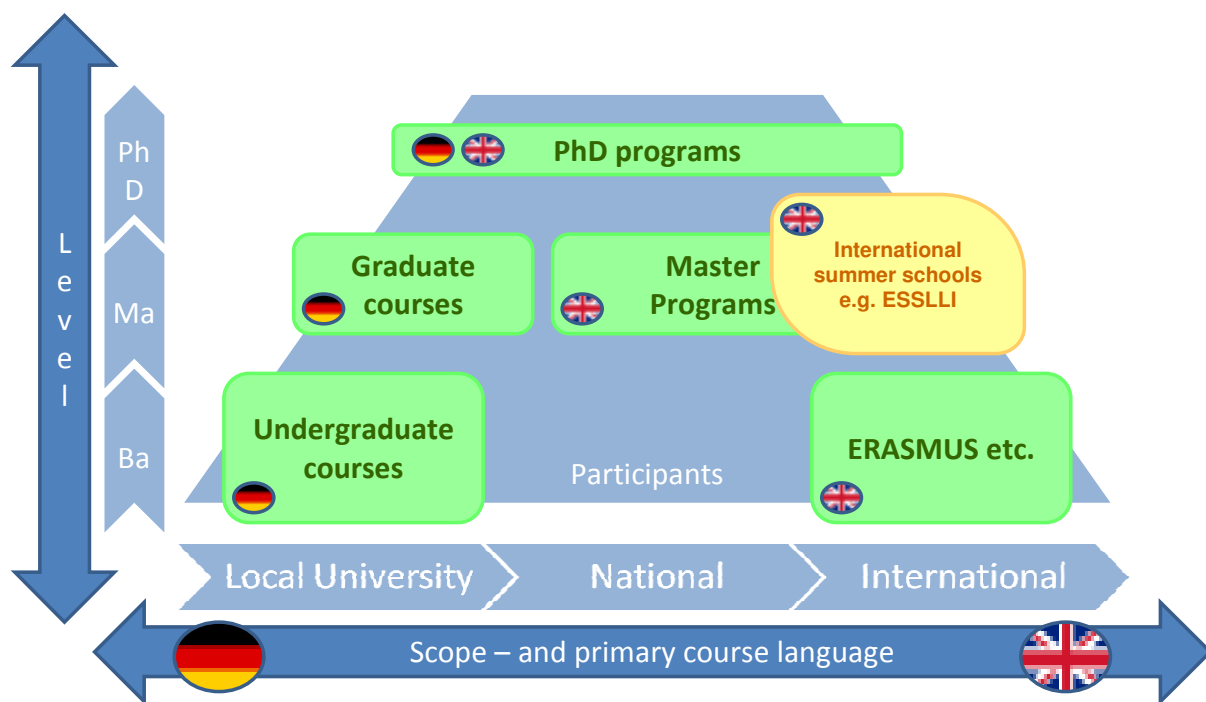


Figure 1: Mapping academic training activities according to scope and level

D-SPIN is currently placing its offers in a national scope. Please note that, although courses do take place at local universities, the resulting training materials are targeting a national audience, since they allow conducting similar courses at different universities in Germany.

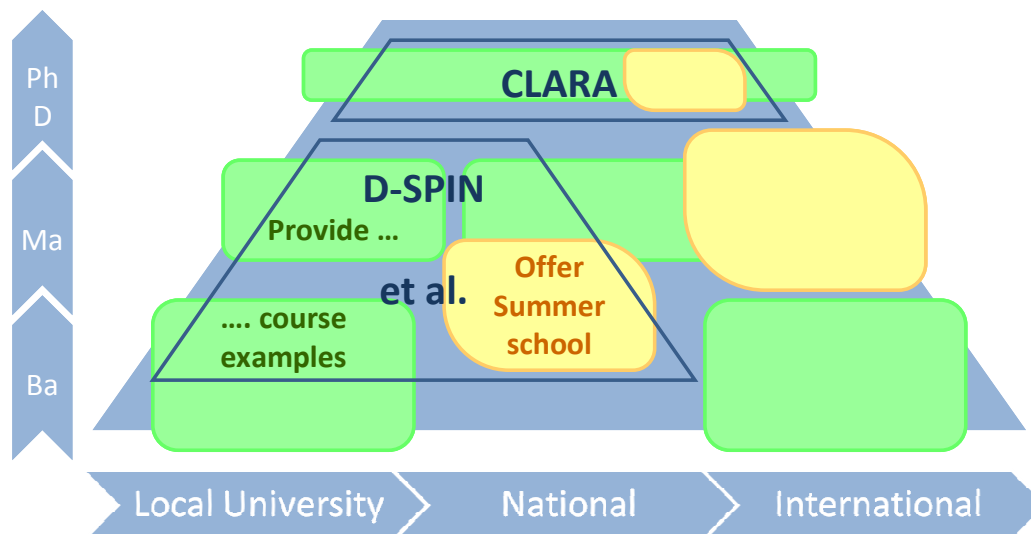


Figure 2: Placing D-SPIN's training activities on the map

D-SPIN's training activities follow the main goals summarized below:

- Offer example courses for training LRT users. Such courses can be integrated as elements of university curricula.
- Provide course materials for such courses, so that they can be reused at other universities.
- Conduct a summer school with a national scope (not covered in this document).
- Develop a help desk concept for D-SPIN (not covered in this document).

The following sections will provide information on the courses and training materials.

3. University courses: Two examples

As part of the D-SPIN activities in community training, university courses are being conducted. The following subsections provide information on two courses that have been designed and conducted at the universities in Frankfurt and Giessen, who are involved in training and education in D-SPIN.

The courses are described individually in the two following subsections. The course designs differ due to a) different groups of participants and b) partly different areas of interest.

Nevertheless, these descriptions might serve as templates for further courses. Besides e-learning content, we consider these course descriptions a central part of the training materials.

3.1. Course “LRT for linguists”, Giessen University, Summer 2009

This section reports on a course at the Justus-Liebig-University Giessen in the summer term 2009. The primary goal was to educate interested potential users of language resources and services. Available and well established resources both from the D-SPIN community and from other providers were used throughout the course. Correspondence regarding this course may be sent to Frank Binder frank.binder@germanistik.uni-giessen.de.

3.1.1. Course participants

The course took place in the summer term 2009 as a weekly seminar course, comprising 12 sessions of 90 minutes each, from April to July 2009. The participants were students from several degree programmes, thus resulting in a mixed group of participants, as shown in Table 1. All participants were students of language-related curricula at the German Department of Giessen University. 80% of the participants had no programming experience or interest.

Course participants	
Bachelor students in German language studies (2nd year):	10
Students pursuing teaching certification (2nd to 3rd year):	6
Master students (1st year) / Magister students (3rd year):	6
Total number of participants	22

Table 1: Participants of the course “LRT for linguists”, Giessen University, Summer 2009

3.1.2. Prerequisites

For general reading on session topics, chapters from textbooks were used throughout the course: Most chapters from Lemnitzer & Zinsmeister (2006) were used. Additional topics were covered through individual chapters from Carstensen et al. (2004), Gries (2008) and individual papers, as mentioned in the session descriptions below.

The course took place in a room with a computer pool of 10 personal computers, allowing for 2 people to share one PC. Throughout the course, both online and downloadable/installable resources were used. Required software was installed on those computers. Students who signed up for presentations on resources and software needed to install relevant software on their private computers. In general, this raised only marginal difficulties. Some support via e-

mail was necessary, both for online use of learning modules and for installing the required software.

Concerning support, the course would not have been possible without the very supportive pool administrators, the local E-Learning infrastructure team, and a co-worker to deal with technical issues of the course-specific learning modules. Furthermore, didactical training and counselling – part of the university’s vocational training offers – were most helpful throughout the course development.

3.1.3. Course elements

The following activities were integrated into the sessions, to cover the topics in ways that integrate and activate students and raise interest and awareness with regard to language resources.

- 5-minute-“short”-presentations of resources and related topics – this element was introduced to activate students and to give all students the possibility to contribute to the course, independent of whether they needed graded certificates.
- Longer “full”-presentations (20-30 minutes, plus optionally another 20-30 minutes for group activities, if suitable) on session topics – on which the presenters would later write their term papers.
- Practical exercises for selected topics (concordancing, plotting) were integrated in some sessions (should be intensified, cf. résumé on page 17), and as part of the preparations or group activities for “full” presentations.
- Using WIKI pages to prepare and summarize group discussions.
- In the final session: prepared and discussed scenarios of LRT-usage, connecting and integrating the topics and issues from previous sessions: Student groups were asked to represent publishing companies, research groups / institutions, schools and funding agencies. Their task was to explore the perspectives of each player, and to develop project ideas and between-group co-operations involving the usage and creation of language resources.
- Term papers for graded certificate (10 of the 22 students) – topics included reviews of corpus-based studies, legal issues and licensing, computational linguistics topics like word sense disambiguation and annotation formats, and last but not least, sketches on using language resources and tools in the context of German classes at schools, porting ideas from Mukherjee (2002), for instance when studying language use in politics.

3.1.4. Course topics

As the course description states, the course is designed to give an overview of existing language resources (data collections, software, and services) and to showcase how these can be used – based on practical exercises and selected research questions. It aims to enable students to use these resources in their own projects and be able to assess the potential and limitations of language resources.

As mentioned above in Table 1, the course was offered for several target groups at once, resulting in a mixed group of participants: Bachelor students of general and German

linguistics, Master students of language technology and language pedagogy, and students pursuing certification for teaching German literature, culture and language at German grammar schools / secondary schools.

Quote: German title of the course (and rough translation)

„Einführung in die Arbeit mit computerbasierten Sprachressourcen“

(Introduction to working with language resources and technology)

Quote: German description of the course, as published in the university calendar:

„Diese Veranstaltung gibt einen Überblick über vorhandene computerbasierte Sprachressourcen (Datensammlungen, Software und Dienste) und deren Nutzungsmöglichkeiten in verschiedenen Anwendungsfeldern. Anhand praktischer Übungen und exemplarischer Fragestellungen aus Sprachwissenschaft und Sprachdidaktik wird der Umgang mit diesen Ressourcen am Computer vermittelt. Neben internationalen Beispielen stehen Ressourcen aus dem deutschsprachigen Raum im Vordergrund. Ziel ist es, die Studierenden zu befähigen, diese Sprachressourcen in wissenschaftlichen Projektarbeiten oder didaktischen Szenarien selbständig einzusetzen. Dabei lernen die Studierenden Möglichkeiten und Grenzen dieser Ressourcen kennen und können die Entwicklung und Qualität der Ressourcen wissenschaftlich diskutieren und anhand praktischer Erfahrungen einschätzen und bewerten.“

Table 2 provides an overview of the session topics. Table 3 lists the texts that students were asked to read in preparation for the respective sessions. Table 4 provides a list of student presentations, as they took place throughout the course.

Session	Session topics – “LRT for linguists” – Giessen, Summer 2009
1	Introduction and overview
2	Foundations: Empiricists, Rationalists & language data
3	Basic corpus linguistics – search and concordances in remote data
4	Basic corpus linguistics – concordances with local data
5	Quantitative corpus work: Collecting and organizing data
6	Quantitative corpus work (continued)
7	Annotation: Using annotation (in remote data)
8	Annotation: Using and creating annotation (in local data)
9	Lexical Resources
10	Resources for educational purposes and language teaching
11	Legal issues, copyright & licences
12	Final discussion

Table 2: Session topics for course “LRT for linguists” at Giessen university

Course readings		
Session	Session topic Preparatory readings	Number of pages to read
1	Introduction and overview (none)	-
2	Foundations: Empiricalists, Rationalists & language data Lemnitzer & Zinsmeister (2006): Chapter 1 & 2 “Einleitung” & “Die Quellen linguistischer Erkenntnis”	34
3	Basic corpus linguistics – search and concordances in remote data Lemnitzer & Zinsmeister (2006): Chapter 3: „Linguistische Korpora“	22
4	Basic corpus linguistics – concordances with local data Lemnitzer & Zinsmeister (2006): Chapter 5: „Deutschsprachige Korpora“	25
5	Quantitative corpus work: Collecting and organizing data Gries (2008): Chapter 1, “Grundlagen empirischer Forschung” “Design und Logik quantitativer Untersuchungen” pp. 1-35	35
6	Quantitative corpus work (continued) Lemnitzer & Zinsmeister (2006): Chapter 6 “Korpuslinguistik in der Praxis”	48
7	Annotation: Using annotation (in remote data) Lemnitzer & Zinsmeister (2006): Chapter 4: “Linguistische Annotation und ihre Nutzung”	40
8	Annotation: Using and creating annotation (in local data) (none)	-
9	Lexical Resources In Carstensen et al. (Eds.) (2004): Section 4.3: Kunze (2004) „Lexikalisch-semantische Wortnetze“ Section 4.4: Gibbon (2004) „Lexika für multimodale Systeme“	17
10	Resources for educational purposes and language teaching In Carstensen et al. (Eds.) (2004): Section 5.10: Ludewig (2004) “Sprachlehr- und -lernsysteme”	7
11	Legal issues, copyright & licences In Rehm et al. (Eds.) (2007): Lehmberg et al. (2007) “Rechtsfragen bei der Nutzung und Weitergabe linguistischer Daten”	10
12	Final discussion (none)	-

Table 3: Course readings for “LRT for linguists”, Giessen University, Summer 2009

List of student presentations

Session	Session topic	URLs
	<ul style="list-style-type: none"> ○ “Full” presentations given by students (20-30 minutes, plus optionally another 20-30 minutes for group activities) ○ “Short” presentations on resources given by students (5 to 10 min each) 	
3	<i>Basic corpus linguistics – search and concordances in remote data</i> <ul style="list-style-type: none"> ○ Linguistic corpora for German ○ Google’s advanced search http://www.google.de/advanced_search ○ WebCorp – Linguist’s search engine http://www.webcorp.org.uk/ 	
4	<i>Basic corpus linguistics – concordances with local data</i> <ul style="list-style-type: none"> ○ Selected scientific journals that relate to the course ○ AntConc http://www.antlab.sci.waseda.ac.jp/antconc_index.html 	
5	<i>Quantitative corpus work: Collecting and organizing data</i> <ul style="list-style-type: none"> ○ Organizing data in spreadsheet software ○ Gnuplot http://www.gnuplot.info/ 	
6	<i>Quantitative corpus work (continued)</i> <ul style="list-style-type: none"> ○ Presenting and discussing corpus-based studies on <ul style="list-style-type: none"> ○ “neologisms in newspapers” and ○ “anglicisms in advertisement” 	
7	<i>Annotation: Using annotation (in remote data)</i> <ul style="list-style-type: none"> ○ Using linguistic annotations in online corpora ○ TigerSearch http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/ ○ Associations and data providers http://www.elra.info/ http://www ldc.upenn.edu/ 	
8	<i>Annotation: Using and creating annotation (in local data)</i> <ul style="list-style-type: none"> ○ TreeTagger http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/ ○ TextGrid http://www.textgrid.de/ 	
9	<i>Lexical Resources</i> <ul style="list-style-type: none"> ○ Lexical resources – foundations and examples ○ Wikipedia as a lexical resource Zesch et al. (2007) + http://de.wikipedia.org ○ WordNet http://wordnet.princeton.edu/ ○ World Atlas of Language Structures http://wals.info/ 	
10	<i>Resources for educational purposes and language teaching</i> <ul style="list-style-type: none"> ○ Computer-assisted language learning ○ Curiosities: ChatBots / Eliza http://de.wikipedia.org/wiki/ELIZA ○ Grammis http://www.ids-mannheim.de/gra/grammis.html ○ Goethe Institut: Jetzt Deutsch lernen http://www.goethe.de/z/jetzt/deindex.htm 	
11	<i>Legal issues, copyright & licences</i> <ul style="list-style-type: none"> ○ Copyrights and licenses ○ Netiquette http://www.netplanet.org/netiquette/ ○ Mailing lists for linguists ○ Linguistic conferences and scientific communities ○ DOBES project http://www.mpi.nl/DOBES/ 	
12	<i>Final discussion</i> <ul style="list-style-type: none"> ○ Curiosities: PersonalNovel http://www.personalnovel.de/py/ 	

Table 4: Student presentations in “LRT for linguists”, Giessen University, Summer 2009

3.1.5. E-learning elements

Throughout the course, e-learning activities were included in three stages (cf. Table 5).

The content of these e-learning materials was closely related to the reading materials. Interactive test questions and exercises were used to illustrate key concepts from the texts.

One objective was to see how the authoring pipeline would perform and which issues would come up when students try to use the materials. The available formats are described below. The authoring process is described in section 4.

Since the course language is German, all materials are available in German.

Session	Topic	SCORM-Package (SCORM version)	Source-Package	Web-Preview	Revision	Comments / known issues
1	Testing the E-Learning access	zip (1.2) (3 MB)	zip (0,17 MB)	web	20090423-1551_FB	Layout-problems in Firefox (cross-browser compatibility issues), support for printing not implemented
7	Linguistic Annotations	zip (1.2) (3 MB)	zip (0,13 MB)	web	20090603-1126_FB	Improved browser compatibility, support for printing not implemented, overflow in text fill-in questions unhandled
11	Legal issues	zip (1.2) (3 MB)	zip (0,17 MB)	web	20090703-2030_FB	support for printing is not correctly integrated

Table 5: E-learning elements for “LRT for linguists”, Giessen University, Summer 2009

This table and the referenced materials are publicly available at

<http://www.uni-giessen.de/~g91214/SeminarSprachressourcen2009/>

Since there are still a number of known issues, the e-learning materials for this course can not yet be considered stable enough for independent productive use.

The materials are provided in two versions, along with an online preview. The first format is a zip-file that serves as a SCORM package (Sharable Content Object Reference Model)⁸. The SCORM packages can be uploaded into any SCORM-compliant learning management system (LMS). They were tested in Moodle and Ilias. The packages are currently provided as SCORM version 1.2.

The source packages can be downloaded, extracted, opened, and modified with the authoring tool CourseLab⁹, which is available free-of-charge for Microsoft Windows. As explained in

⁸ An Introduction to SCORM 2004 3rd Edition:

http://projects.aadlcolab.org/scourse/2004_3rdEd/viewer/index.html

⁹ CourseLab - free e-Learning authoring tool: <http://courselab.com/>

section 4, we chose CourseLab as our authoring tool, and our packages were created with CourseLab version 2.4.

The web preview is simply the extracted SCORM-package, uploaded to a web space, i.e. without a “surrounding” LMS. When extracted, the SCORM packages - as exported by CourseLab - are completely functional (navigation, questions, etc.), except that no performance data is stored. Hence, the exported SCORM packages work both locally (for testing), online at some web space, as well as when uploaded to an LMS.

3.1.6. Course evaluation

At two stages, aspects of the course were evaluated. First, a mid-term evaluation was performed to elicit feedback from the students about their opinion on the course so far, and about their expectations for the second half of the course. Second, the usage of e-learning materials was evaluated through an online-survey, which was conducted by the local E-Learning-infrastructure team as part of their regular evaluation of e-learning activities at Giessen University.

The mid-term evaluation was conducted directly in the course. A small feedback sheet with two questions was handed out, asking about positive aspects and suggestions for improvements with regard to the overall mission of the course. The results made obvious that a) the course atmosphere was positively perceived as open and lively, apparently related to the short presentations on various topics and b) that the broad range of topics was well accepted. However, it was asked that more time be devoted to practical work with resources.

The e-learning evaluation was conducted by the university’s E-Learning infrastructure team as part of a university-wide anonymous survey among all participants of those courses that provided e-learning. The survey was placed into the LMS during the final weeks of the term. Along a general tendency in the survey, this course received positive feedback on electronic communication and on using wikis to document and prepare session topics and group activities. However, the overall scoring of the course’s e-learning activities – compared to those of other courses – is not satisfying and indicates the need for improvement. Most importantly, it was demanded that the e-learning tasks should relate more closely to practical work with resources and tools. Additionally it was criticised that technical difficulties were an unnecessary distraction from the course topics. However, it must be said that during the course the majority of the participants were able to work through the e-learning materials without problems, and that only 2 to 5 students responded to the individual questions of the survey. In that sense, the results from the survey may not be considered representative of the whole course.

3.1.7. Résumé

The course was perceived as a valuable offer in the curriculum. The course will be offered again in 2010. Some suggestions for improvements and further development have been discussed in the previous subsections. Comments and suggestions – be it minor revisions or complete redesigns – are highly appreciated.

One intention behind the course was that it should serve as a guiding path and test case for developing e-learning materials on using language resources. This intention, however, was not fully served. So far, the e-learning materials cover only a few parts of the course.

It is expected that such courses are already in place at other universities and that establishing contact with the instructors could prove valuable, in order to collect feedback and requirements regarding the provision of teaching materials for such courses.

3.2. Goethe University Frankfurt Main: introductory course “Corpus linguistics”

The course provides a wide range of material, which presents efficiently besides the theoretical knowledge in corpus linguistics, a rarely available information how to find the relevant information from the digitalized resources such as historical, diachronic corpora and corpora of modern languages.

The course was created with the help of the authoring tool CourseLab, saved in SCORM format and integrated in the E-Learning platform Moodle. Details of the authoring process are laid out in section 4.

3.2.1. Course participants

Participants	Number
Magister students (1 st year general and comparative linguistics)	1
Postdocs (general and comparative linguistics)	5
Guest students (general and comparative linguistics)	2
	8

Table 6: Participants of the course “Corpus Linguistics”, Frankfurt University

3.2.2. Course topics

Goethe University Frankfurt, fall/winter 2009/2010 Introductory course “Corpus linguistics” (level: beginners; module access: successive) Svetlana Ahlborn		
N.	Topic/ Module	Description
1.	1. Entering guide: work with the E-Learning course. 2. Development of corpora	1. Guidelines for the work with the E-Learning platform Moodle. 2. The module describes the development of corpus linguistics starting with projects from the 60-s till the infrastructure initiatives.
2.	Types of corpora	Different types of corpora are presented.
3.	Historical and diachronic corpora	The module presents different types of current historical and diachronic corpora

4.	1. Historical and diachronic corpora. 2. Practical guide: TITUS	The usage of the text archive TITUS is presented.
5.	Introduction to query systems	The module treats such topics as „Linguistic search engines“, “Web as a corpus”, “Web linguistics”.
6.	Concordances and concordancers: application	The module surveys the practice-oriented contact with concordancer software. The notion “concordance” is accompanied with the demonstration of tools from DWDS Project and the COSMAS Project.
7.	1. Corpus search and retrieval software. 2. Search in own corpus.	The module shows different types of retrieval software. The students get acquainted with such notions as “collocates”, “concordance plot”, “word lists”, “keywords”, “keyness”, “n-gram” with the help of the freeware tool AntConc.
8.	1. Corpus search and retrieval software and UNICODE 2.Regular expressions: an introduction	The module gives an opportunity to analyse own UNICODE encoded corpora.
9.	Markup languages	The module presents different encoding and text exchange formats (TEI, ISO; SGML, HTML, XML; DTD; Unicode)
10.	Metadata: use and creation	The module surveys metadata standards and software tools for metadata creation.
11.	Annotation	The module gives an introduction to such topics as „annotation levels“, “segmentation” etc. Different annotation programs are presented.
12.	The use of linguistic annotation	The up-to-date projects, such as Weblicht or DOBES, are used in order to demonstrate the use of linguistic annotation.

Table 7: Course topics for course “Corpus linguistics” at Frankfurt University

3.2.3. Learning scenario

One learning unit consists of a combination of presence and online phases. The online phase is combined with the previous and the following face-to-face lesson in the classroom. The feedback to the E-Learning materials in the presence lecture plays a very important role in integrating of online methods. The presence lecture includes feedback in order to support the usage of online materials. This sequence is demonstrated in Figure 3:

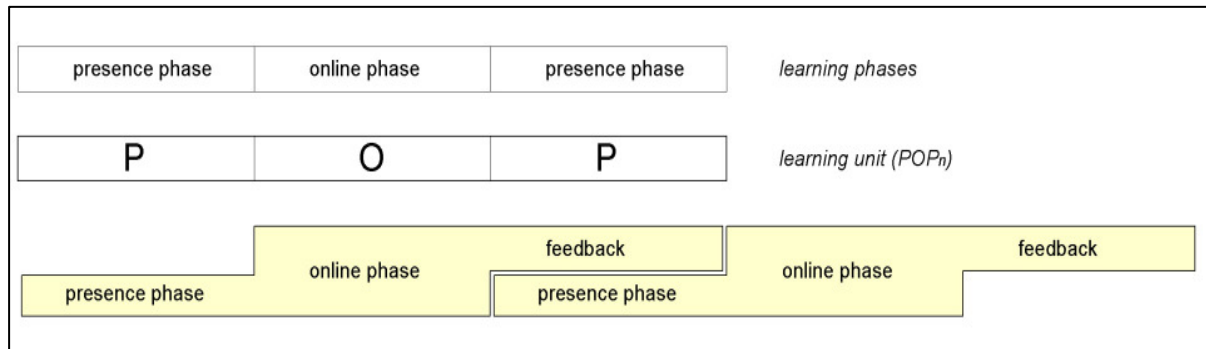


Figure 3: Learning scenarios: The POP model

Now the notion “module” should be more exactly defined. When each unit consist of the alternating presence and online phases, then one module consists of several dynamically growing learning units. The next figure shows the correlation between units and modules:

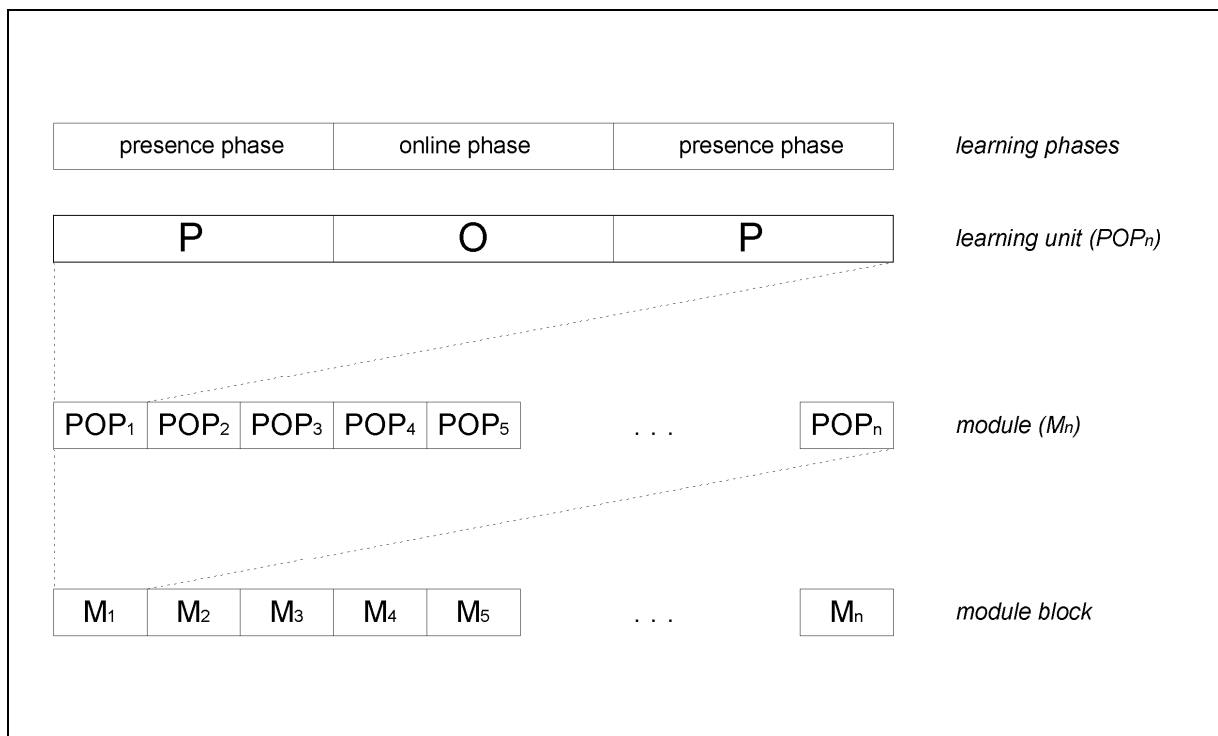


Figure 4: "Learning unit" and "module"

3.2.4. Module access approaches

Furthermore, different module access approaches according to student levels can be differentiated:

- Module Successive access. This access art suits for the beginners. Lecturer defines the order of modules in order to impart knowledge systematically.
- Module successive access combined with application-driven access. This level suits for the advanced students. After getting knowledge in modules, which are strictly ordered, students get an opportunity to choose further modules depending on their interests and specialisation.
- Application-driven access. This access should be used by special purpose students, who have already previous fundamental knowledge. Modules can be picked up according to their interests and specialisation.

student level	module access approach	
beginners	successive access	
advanced students	successive access + application-driven access	
special purpose students	application-driven access	

Table 8: Module access approaches

3.2.5. Online exercises (with screenshots)

The course includes different types of online exercises. For example:

- Multiple-select exercises with interactive elements;
- Multiple choice exercises;
- Input text exercises;
- Drag&drop exercises;

D-SPIN CLARIN

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Übung

Frage (nicht gewertet)

Um welche Textart handelt es sich beim Stein von Rosette?

Markieren Sie die richtigen Antworten

☐ Der Stein von Rosette stellt ein Parallelkorpora dar

☐ Der Stein von Rosette ist ein Beispiel für einen Übersetzungstext, und zwar eine Trilingue.

Versuche: [Frage überspringen](#)

Figure 5: Multiple-select exercise with interactive elements

Deutsche Sprachressourcen-Infrastruktur D-SPIN

Seite 32/35

Unicode: Übung

Verse: 16 *swa liuthjai liuhaj izwar in andwairþja manne, ei gasailþaina izwara goda waurstwa jah hauhjaina attan izwarana þana in himinam.*
sic luceat lux vestra coram hominibus / ut videant vestra bona opera / et glorificent Patrem vestrum qui in caelis est

Frage (bewertet)

Ermitteln Sie den Wert des gothischen Wortes "þana" in der UTF-8 Kodierung (in Dezimalschreibweise).

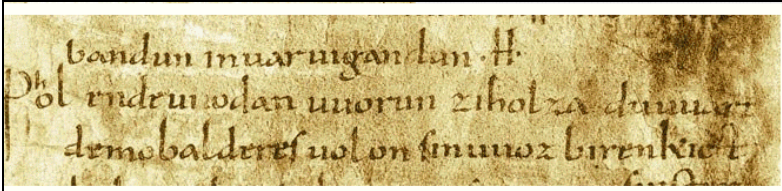
Fügen sie die Zahlenkette ein

Der Wert ist:

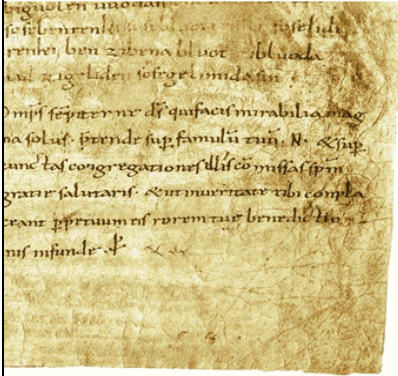
Attempts: 2

Nutzen Sie zur Hilfe die Unicode-Tabelle

Figure 6: Input text exercise



bandum inuarigan lan .H.
 Pol rnderu odan uuorun ziholza dunnert
 demobaldere suol on sinuuoze brenkiet



signeten inuocan
 sefelenrenle...
 renkei ben zibena bluoer...
 bal zig liden...
 Omis sepeten...
 na solat...
 tunc tot congregaciones...
 gratie salutaris...
 ceant ppetuum...
 mis in fine...

Credited question

Auf welcher Ebene würden Sie die Tokenisierung von Handschriften durchführen?

Enter text string

Attempts: 2

Figure 7: Input text exercise

Deutsche Sprachressourcen-Infrastruktur D-SPIN
Seite 8/35

Drag&Drop Übung: Mehrwortlexeme

Bitte finden Sie die Mehrwortlexeme und ziehen sie mit der Maus in das grau markierte Feld hinein.

Er prahlte in der , dass er

in nur Stunden nach angekommen sei .

Mehrwortlexeme

"Frankfurter Allgemeinen"

4,5

New York

Lösung abgeben

Figure 8: Drag&drop exercise

3.2.6. E-Learning modules

	Topic	SCORM format	ZIP format	Web access (Internet Explorer is recommended)
1	Development of corpora	S_Modul_geschichte_m01.zip [4385.0 KB]	Modul_geschichte_m01.zip [10105.5 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_geschichte_m01/1/start.html
2	Types of corpora	S_Modul_Korporaarten_m02.zip [1588.8 KB]	Modul_Korporaarten_m02.zip [547228.8 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_Korporaarten_m02/1/start.html
3	Historical and diachronic corpora	S_Modul_Historische_und_diachrone_Korpora_Teil_1_m03.zip [2168.6 KB]	Modul_Historische_und_diachrone_Korpora_Teil_1_m03.zip [5755.5 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_Historische_und_diachrone_Korpora_Teil_1_m03/1/start.html
4	Historical and diachronic corpora: TITUS	S_Modul_Historische_und_diachrone_Korpora_Teil_2_m04.zip [736.9 KB] S_Modul_Titus_Teil_1_m04.zip [8420.4 KB]	Modul_Historische_und_diachrone_Korpora_Teil_2_m04.zip	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_TITUS_Teil_1_m04/1/start.html
5	Introduction to query systems	S_Modul_Titus_Teil_2_m05.zip [2576.7 KB]	Modul_Korpusabfrage_m05.zip [2353.7 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_TITUS_Teil_2_m05/1/start.html
6	Concordances and concordancers: application	S_Modul_Konkordanzprogramme_Teil_1_m06.zip [920.6 KB]	Modul_Konkordanzprogramme_Teil_1_m06.zip [13811.5 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_Konkordanzprogramme_Teil_1_m06/1/start.html
7	Corpus search and retrieval software: Search in own corpus.	S_Modul_Konkordanzprogramme_Teil_2_m07.zip [1810.5 KB]	Modul_Konkordanzprogramme_Teil_2_m07.zip [24918.1 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_Konkordanzprogramme_Teil_2_m07/1/start.html

8	Corpus search and retrieval software and UNICODE Regular expressions: an introduction	S_Modul_Konkordanzprogramme_Teil_3_m08.zip [1333.4 KB]	Modul_Konkordanzprogramme_Teil_3_m08.zip [24547.1 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_Reg_Ausdr_m08/1/start.html
9	Markup languages	S_Modul_Textauszeichnung_m09.zip [2559.3 KB]	Modul_Textauszeichnung_m09.zip [5976.9 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_Textauszeichnung_m09/1/start.html
10	Metadata: use and creation	S_Modul_Metadaten_m10.zip [1398.7 KB]	Modul_Metadaten_m10.zip [3080.1 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_Metadaten_m10/1/start.html
11	Annotation	S_Modul_Annotation_m11.zip [4602.0 KB]	Modul_Annotation_m11.zip [8992.5 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_Annotation_m11/1/start.html
12	The use of linguistic annotation	S_Modul_Annotation_Nutzung_m12.zip [1088.9 KB]	Modul_Annotation_Nutzung_m12.zip [1692.9 KB]	http://titus.uni-frankfurt.de/d-spin/e-learning/Modul_Annotation_Nutzung_m12/1/start.html

Table 9: E-Learning modules, Frankfurt University fall/winter 2009/2010

3.2.7. Course access information

There are two possibilities to access the course material: via web link (Table 9) or via learning platform.

Link for the learning material "Corpus linguistics" (Goethe Universität Frankfurt am Main):

<http://moodle.studiumdigitale.uni-frankfurt.de/moodle/>

The course can be accessed by the course students and D-SPIN-Partners. D-SPIN Partners can access the course via:

[Fachbereich 9], [Korpuslinguistik], [Externe Nutzer], [Als Gast anmelden]

Access code, which should be treated as confidential, is: corpora

The E-Learning materials were prepared in German.

3.2.8. Résumé

The created course at the Goethe University in Frankfurt am Main demonstrated that the usage of eLearning methods provides to be up-to-date also in the field of comparative and historical linguistics. Besides the theoretical knowledge in language grammar students need the practical experience in how to find the relevant information from the digitalized resources.

With the help of online demonstrations and exercises the students practiced in their work with different tools, which helped to find the origin or statistical appearance of a linguistic phenomenon searched for.

Two learning phases were distinguished during the course: the online and the presence phases. These phases were organised in accordance with POP-Model. The feedback to E-Learning materials in the presence lecture played a very important role in integrating of online methods.

Additionally three different module access approaches according to student levels were differentiated: successive access for beginners, successive access combined with application-driven access for advanced students and application-driven access for special purpose students.

The automatic access from module to module should be possible with the next release of Moodle 2.0.

4. Creating e-learning materials

The goals of D-SPIN WP6 include both giving courses for LRT-users and developing e-learning materials. This section focuses on the latter. Since our main goal was to develop content and templates for courses, we had to avoid putting too much effort into technical issues and needed to rely on solutions that are readily available.

The following sections illustrate our way of creating e-learning materials to support our training courses.

4.1. *Planning an authoring pipeline*

The project goals clearly state that the training and learning materials should be easy to maintain and standard-compliant. Furthermore, it became obvious that they must be usable in various learning management systems, since different such platforms are used at different training sites. It was agreed that dependencies on particular learning platforms need to be avoided.

Two target platforms were identified, in order to test and ensure that the authoring pipeline would work well for both platforms: ILIAS and Moodle. Both platforms allow to import learning materials as SCORM-packages (Sharable Content Object Reference Model)¹⁰.

Since relying on export capabilities of learning platforms was not recommended, we opted for a conservative design of our authoring pipeline:

1. Create materials with off-line authoring tools;
2. Save and export modules as SCORM packages;
3. Import modules in any learning management system that supports SCORM, such as ILIAS and Moodle.

4.2. *Choosing an authoring tool*

During the preparation phase for the course, several off-line authoring tools were considered. A critical requirement was the capability to produce SCORM-packages, to create content that can be used in different e-learning platforms. Furthermore, the authoring tool should support the creation of exercises and allow for a modular organization of course materials. Additionally, we looked for a tool with an intuitive look-and-feel and the possibility to work efficiently, for instance through well-developed support for copy-paste-modify editing and the use of templates throughout the authoring process. We considered the following tools: eXe eXeLearning¹¹, WBTEexpress Moodle Edition¹², eLAIX¹³, and CourseLab, and decided to use CourseLab¹⁴.

¹⁰ An Introduction to SCORM 2004 3rd Edition:
http://projects.aadlcolab.org/scourse/2004_3rdEd/viewer/index.html

¹¹ eXe eXeLearning: <http://exelearning.org/wiki>

¹² WBTEexpress Moodle Edition : <http://de.wbteexpress.com/wbte-moodle-edition.html>

¹³ eLAIX: <http://www.boldt-media.de/download.html>

¹⁴ CourseLab - free e-Learning authoring tool: <http://courselab.com/>

The authoring tool CourseLab fulfilled our requirements to a high degree. It provides a WYSIWYG-interface, which is well usable for authors that are familiar with Microsoft PowerPoint. The tool exports learning modules in a custom XML-based, SCORM-compliant format, applying JavaScript and XML to display the course contents in browsers. Exported modules can also be viewed and browsed off-line on the local file system, which allows for local testing. Note that CourseLab is closed-source and only available as windows-native freeware, downloadable from courselab.com. Registration is required.

Some additional work was necessary, to add printing capabilities (a demand that was correctly anticipated by a colleague and usability expert) and to improve cross-browser compatibility. Cafer Travaci (University of Giessen) provided a patch for adding printing capabilities, that can be applied to the modules created with CourseLab.

The choice of the authoring tool is problematic to some degree, since it introduces a dependency in that stage of the authoring pipeline. In the worst case, learning materials will have to be rebuilt using another authoring process. Fortunately, however, since the exported modules can be viewed without CourseLab, both access to the modules and their contents will not be lost.

5. Summary and outlook

Among the training activities in D-SPIN, two university courses for potential LRT-users were developed and conducted at the Universities of Giessen and Frankfurt. While introductory topics and foundations were similar in both courses, the courses differ in their design. This can be related to the different institutional settings, hence different groups of participants, and consequently, to individual approaches taken by the instructors.

To support these courses, e-learning materials were developed at both universities. However, e-learning played different roles in the two courses. In Frankfurt, the course was based on self created e-learning materials, which were specifically tailored for this course. In contrast, the Giessen course utilized existing text books and additional reading materials as well as student presentations. Except for course management and communications, e-learning played a rather supplementary role here.

Regarding further D-SPIN work on training and education, the focus will lie on the D-SPIN Summer School 2010. The learning materials from the courses at Frankfurt and Gießen will be used to give an “interactive tutorial on corpus linguistics” at this event.

Furthermore, the already developed courses will be offered again in summer 2010 and winter 2010/11, respectively. In that context, the created e-learning materials will be reused and complemented. Since the courses are described in terms of term schedules, session topics, reading lists, presentations and activities, these descriptions might serve as templates for further courses. We consider these course “templates” a central part of our training materials.

6. Reports

Ahlborn S. Modulorientiertes Lernverfahren. (Module oriented teaching experience). Sitzung der Institutsvertreter „Empirische Runde“. 16. November 2009. Goethe-Universität Frankfurt am Main.

Ahlborn S. Mehrzweckorientierter eLearning-Kurs mit Einsatz rekursiver Übungen (Multipurpose eLearning-Course with recursive exercises). 4. eLearning-Netzwerktag. 14. Dezember 2009. Goethe-Universität Frankfurt am Main. //http://www.studiumdigitale.uni-frankfurt.de/events/nwt2009/Folien/Ahlborn_Netzwerktag_2009.pdf

Binder, F., and Dan Cristea (2009) “Training and dissemination in CLARIN – the potential to bridge gaps” Article in CLARIN Newsletter #6 June 2009. <http://www.clarin.eu/newsletter/>

7. References

- Binder, F. and Cristea, D. (2009) "Training and dissemination in CLARIN – the potential to bridge gaps" Article in CLARIN Newsletter #6 June 2009.
<http://www.clarin.eu/newsletter/>
- Carstensen, K.-U.; Ebert, Ch.; Endriss, C.; Jekat, S.; Klabunde, R.; Langer, H. (Eds.) (2004) *Computerlinguistik und Sprachtechnologie. Eine Einführung*. (2nd edition) München: Elsevier.
- CLARA "Common Language Resources and their Applications" <http://clara.uib.no/>
- CourseLab - free e-Learning authoring tool: <http://courselab.com/>
- Gries, S. Th. (2008) *Statistik für Sprachwissenschaftler*. Göttingen : Vandenhoeck & Ruprecht.
- Horton, W. (2006) *Evaluating e-learning*. Alexandria, VA: American Society for Training & Development.
- ILIAS Learning Management. <http://www.ilias.de>
- Kirkpatrick, D.L. and Kirkpatrick, J.D. (2006) *Evaluating training programs: The four levels*. (3rd edition) San Francisco: Berrett-Koehler, 2006
- Lemnitzer, L. and Zinsmeister, H. (2006) *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.
- Mukherjee, J. (2002) *Korpuslinguistik im Englischunterricht*. Frankfurt am Main: Peter Lang.
- Moodle: open-source community-based tools for learning. <http://moodle.org/>
- Rehm, G.; Witt, A.; Lemnitzer, L. (Eds.) (2007): *Data structures for linguistic resources and applications*. Proceedings of the Biennial GLDV Conference 2007. Tübingen: Narr.
- SCORM: An Introduction to SCORM 2004 3rd Edition:
http://projects.aadlcolab.org/scourse/2004_3rdEd/viewer/index.html