Maja Bärenfänger, Maik Stührenberg

E-Learning and Computational Linguistics An Introduction

0 Preface

E-Learning is more and more becoming a commodity in all areas of education ranging from curricular learning at schools and universities via further education and learning on the job to social and game-based learning in spare time activities. At the same time, research on E-Learning is scattered across different research disciplines, being often a matter of single initiatives and persons. In the light of this growing gap, the authors of this introduction want to showcase Computational Linguistics and Language Technology as possible key enabling technologies for current E-Learning.

1 Current trends in E-Learning

E-Learning, or Technology-Enhanced Learning is an interdisciplinary research area which involves research disciplines like Pedagogy, Pedagogical Psychology, Computer Science (particularly research on human-computer interaction (HCI), knowledge management, and software architectures for virtual learning management systems), Artificial Intelligence (especially research on intelligent tutoring systems (ITS)) and Computational Linguistics (in particular computer-assisted language learning, educational natural language processing and text technology for learning technology standards and applications). In the last 15 years, the research area "E-Learning" has undergone major changes. While the first decade, starting in the end of the 20th century, has mainly focused on learning infrastructure and learning content, more recently, the learning process and the learners have been taken more seriously. Today, a learner-centred paradigm dominates E-Learning. Current areas of interest are adaptivity and personalization (individual and ubiquitous learning settings, personal infrastructures supporting life long learning), learning in social communities, and learning & entertainment (game-based learning).

Taking up trends in the Web 2.0, attention has been given to aspects like virtual learning communities, web based collaboration, social media content and social learning infrastructures (e.g. edublogs and wikis). This insight also had an influence in research on and development of virtual learning environments: Personal learning environments or mashup personal learning environments are subjects of increasing interest, and are more and more seen as more adequate promoters of life long learning than traditional virtual learning environments. Moreover, applications and methods supporting learning at any time and any place are also seen as important factors to improve life long learning. Catch phrases in this context are "ubiquitous learning", "mobile learning" and "augmented reality". It can be expected that these research and development fields will become more and more important in the next years. Beside these trends, two more important aspects of current technology-enhanced learning have to be mentioned: E-assessment and game-based learning. E-Assessment has always played a major role in E-Learning, but has long been restricted to simple multiple-

choice or fill-in-the-gap tasks, because of a lack of more advanced methods to automatically evaluate input from open tasks or essays. In the last years, there has been profound progress in this area, especially through the application of educational natural language processing techniques. Game-based learning or serious games are the flip side of the coin. These areas focus on more informal and entertaining methods of training and self-assessment.

2 E-Learning and Computational Linguistics

Computational linguistics (CL) has a long tradition in research on E-Learning. For a long time, this research has been restricted to computer-assisted language learning (CALL). But in the last decade, attention has been shifted to other research areas concerning a large variety of aspects of E-Learning. In the following, we will give an overview of main areas and aspects. This overview is organized in three parts: 1. CL research on digital learning content and resources, 2. CL research on computer-mediated communication and collaboration, and 3. CL research on E-assessment.

1. **Digital learning content and resources**: Computational linguistics research on content is concerned with the following major tasks: a) the development and annotation of language resources for E-Learning (corpora, ontologies, and semantic resources), b) the analysis and evaluation of content (collaboratively constructed resources/user-generated content vs. learning objects). The first task is especially related to Text Technology. Text Technological knowledge is applied when it comes to structuring content in a modular way. Separating content and formatting by using methods such as XML allows for E-Learning content to be presented in various formats and screen sizes, establishing the foundation for mobile and/or ubiquitous learning. The second task deals with a variety of aspects that range from information extraction (e.g. term, keyword, glossary or definition extraction), information visualization (e.g. automatic structure discovery and visualization, concept visualization) to quality assessment (e.g. opinion mining, sentiment analysis).

2. **Computer-mediated communication and collaboration:** Computational linguistics research in this area is primarily concerned with the development of tools that support collaborative work (especially tasks like editing, searching, evaluating – e.g. by information/structure/concept visualization, grammatical error detection and correction, glossary and definition detection, and quality assessment) and communication between learners (e.g. by summarizing discussion threads) or between learner and virtual learning environment (e.g. by providing an E-tutor/generating tutorial responses). The latter can be viewed as a special case of an adaptive dialogue system.

3. **E-assessment:** In the context of E-assessment, three types of computational linguistics research tasks can be distinguished: a) analysis of learner input (in particular grammatical error detection and correction, discourse and stylistic analysis, plagiarism detection), b) generation of feedback and generation of test question/tests, and c) monitoring learning process (e.g. visualization of learner's concept maps).

In the mentioned research fields, a variety of methods and techniques are used. These methods include, but are not limited to educational natural language processing techniques

(Educational NLP), or methods referring to text or data mining, latent semantic analysis (LSA), information retrieval and extraction.

As a result, E-Learning can be seen as a prime application of Computational Linguistics and Text Technology, both as a field in which knowledge and methods can be applied in an integrated manner but also as a research topic. In both ways, CL and its sub-disciplines can act as hinge for other disciplines involved such as Pedagogy or Computer Science. Still, many technologies are in their infancies and thus lack the robustness and usability needed for professional product development. At the same time, the users of E-Learning are typically open to new technologies and a playful use even of error prone beta versions of software.

One problem that has been observed from experts in the area is a difficulty in getting funding for interdisciplinary research in general and E-Learning-related projects in particular. Current subject-internal reviewing processes often seem to be an obstacle in launching new projects. Substantial funding and longer lasting programs would definitely help establishing Language Technology as an integral part of E-Learning software. The market opportunities for education technology, especially in game-based learning are huge. The authors of this paper hope to help creating more acceptance for the research to be done on this topic.

3 Important Conferences and Workshops

E-Learning as a highly interdisciplinary research and development field deals with a wide variety of aspects on learning and technology. Related conferences and workshops therefore often offer a broad spectrum of topics from different disciplines. Important broadly oriented E-Learning conferences are:

- CSEDU (International Conference on Computer Supported Education, since 2009)
- EC-TEL (European Conference on Technology Enhanced Learning, since 2006)
- ECEL (European Conference on e-Learning, since 2002)
- ONLINE EDUCA (International Conference on Technology Supported Learning & Training)
- ICL (International Conference on Interactive Computer aided Learning, since 1998)
- IEEE EDUCON conference (since 2010)

On the other hand, E-Learning is also discussed in more special contexts, e.g. in communities which focus on one specific aspect of E-Learning. Regular conferences and workshops that are especially relevant for computational linguists dealing with E-Learning are:

- ACL-HLT Workshop on Innovative Use of NLP for Building Educational Applications (since 2003)
- DeLFI (e-Learning-Fachtagung Informatik der Gesellschaft für Informatik, since 2003)

- LREC (International Conference on Language Resources and Evaluation, since 1998)
- RANLP (Recent Advances in Natural Language Processing, since 1995)

These conferences often offer workshops on "E-Learning and Computational Linguistics/Language Technology/Natural Language Processing". In the last four years, workshops had discussed topics like

- "The People's Web Meets NLP: Collaboratively Constructed Semantic Resources" (workshop in conjunction with ACL 2009 and workshop in conjunction with COLING 2010),
- "Supporting eLearning with Language Resources and Semantic Data" (workshop in conjunction with LREC 2010),
- "Natural Language Processing and Knowledge Representation for eLearning environments", "NLP for Educational Resources" (workshops in conjunction with RANLP 2007),
- "What can Natural Language Processing and Semantic Web technologies do for eLearning?" (workshop in conjunction with ACL 2007).

In contrast to the above-mentioned workshops and conferences that are open for a wide variety of topics in the context of "E-Learning and Computational Linguistics/Language Technology/Natural Language Processing", the following conferences focus on a single special research and development field, namely computer-assisted language learning (CALL). Research and development on CALL has long been the only computational linguistics research area in E-Learning. Conferences in this field often have a long tradition and go back to the beginning of the 1990s. Important conferences on this topic are:

- EUROCALL (Conference of the European Association for Computer Assisted Language Learning, since 1993)
- WORLDCALL (Conference of the Worldwide Association for Teachers and Educators interested in Computer Assisted Language Learning, since 1998)
- CALICO (Conference of the Computer Assisted Language Instruction Consortium, since 1998)
- IALLT (Conference of the International Association for Language Learning Technology, since 2000)

These lists are not exhaustive and are only included to show both the broad spectrum of E-Learning as a research topic as such and the significance that E-Learning has gained in the CL community.

4 Special interest group "Language technology and text technological methods in E-Learning" (GSCL-SIG "E-Learning")

In 2007, "Language technology and text technological methods in E-Learning" was constituted as a special interest group of the German Society for Computational Linguistics and Language Technology (GSCL). The aim of this special interest group is to shape the Computational Linguistics and Text Technological perspective on E-Learning, particularly on aspects like personalization and adaptivity, and to elaborate on and discuss methods and applications that can be assigned to E-Learning specific tasks. The special interest group serves as a panel to support communication and cooperation between experts with different research profiles and competencies (hypermedia, natural language processing, learning technology standards, computer-assisted language learning, amongst others).

As a step towards an opening of the discussion to a wider audience, the GSCL-SIG organized a workshop on "Language Technology and Text Technological Methods for E-Learning" which took place on the 8th September 2010, in conjunction with KONVENS 2010. The workshop presented five plenary talks on various aspects of the workshop topic and two invited keynote talks on "Wikulu: Information Management in Wikis Enhanced by Language Technologies" (by Iryna Gurevych, TU Darmstadt) and "The Snowflake Effect in learning and research" (by Erik Duval, Katholieke Universiteit Leuven). The complete program of the workshop is available at "http://konvens2010.coli.uni-saarland.de /wsprogram_en.html".

5 Overview of papers in the first part of this volume

The first part of this special issue contains three papers that emerged from presentations at the workshop. The authors considerably extended and revised their original submissions so that the articles in this volume represent the current state of the art of the authors' work (see table 1 for an overview).[1]

The contribution "Using Latent-Semantic Analysis and Network Analysis for Monitoring Conceptual Development" by Fridolin Wild, Debra Haley and Katja Bülow focuses on language technology for monitoring learning progress. The authors report on a system called CONSPECT which helps online learners and tutors to monitor the learner's conceptual development by extracting and visualising a conceptual representation from a learner's blog post or the like. For the concept extraction, Latent Semantic Analysis (LSA) and Network Analysis (NA) are combined to a method called Meaningful Interaction Analysis (MIA). The system is evaluated in two verification experiments. In the first experiment, the output of CONSPECT (namely, concept clusters) is compared to human output (namely, human card sorts) in order to find out how closely humans agree with the system's concept clusters. The second experiment is used to evaluate if human text annotations are similar/different to text annotations made by CONSPECT. Wild/Haley & Bülow appraise the findings of both

[1] The papers in this part of the special issue were peer-reviewed in two rounds (extended abstract and full paper) by at least two members of the program committee of the workshop and by the editors of part one, Maja Bärenfänger and Maik Stührenberg. Through this review process, three out of seven submitted papers were selected for this publication.

experiments as a promising start for a computer-based system that monitors conceptual development.

In their article "Meaning versus Form in Computer-assisted Task-based Language Learning: A Case Study on the German Dative", Sabrina Wilske & Magdalena Wolska report on an empirical study which investigated the learning effects of different types of production settings (free vs. constrained production) and different types of feedback (implicit vs. explicit/metalinguistic) in a dialogue system. This system supports a goal-oriented communicative approach to language learning for German. In this approach, communication between learner and dialogue system is framed in a real world situation – in this case triggered by a directions giving task – that elicit the use of the dative case in prepositional phrases. Advantages of this approach are a clearly defined communicative outcome and a focus on meaning. The main focus of this article lies on the empirically grounded comparison of learning effects between free vs. constrained production and implicit vs. explicit feedback. The results of the empirical study indicate that a stronger focus on form (constrained production) has greater learning effects, especially with regard to accuracy gains in using the form.

	Wild/Haley/Bülow	**Wilske/Wolska**	**Miyakoda/Kaneko/ Ishikawa**
Research Area	E-assessment/(self-) awareness of the learner's progress, particularly by computer-based analyses of states in a learner's conceptual development	Computer-mediated communication, in particular task-based instructional dialogues with a computer-based language learning system	Digital learning content, namely multimodal user-generated items for vocabulary learning
Learning Setting	Self-directed learning of factual knowledge in a personal learning environment	Instructed learning of grammatical structures (namely: the dative case in prepositional phrases) in a computer-based dialogue system	Self-directed vocabulary-learning in a personal and ubiquitious learning environment for mobile devices
Methods	Latent Semantic Analysis and Network Analysis	NLP methods (e.g. parsing with mal-rules, template-based dialogue generation)	Text technological methods (in particular: mobile web applications)
Domain	Factual knowledge learning	Language learning	Language learning

Table 1: Overview of the papers

The contribution "Effective Learning Material for Mobile Devices: Visual Data vs. Aural Data vs. Text Data" (Haruko Miyakoda, Kei-ichi Kaneko & Masatoshi Ishikawa) introduces

a vocabulary learning system for mobile devices which supports learners in creating their own multilingual and multimodal learning material, distributing and sharing it with other learners as a podcast via the iTunes application, and evaluating their own material or material made by others by score or comment. Like Wilske/Wolska, the authors also investigate the dependency between learning effect and learning setting – their focus lies on the effect of different types of vocabulary material (text, visual, aural, and combinations of these) on memory retention rate/memorization. For this purpose, two experiments with 100 learners of German as a foreign language are carried out that show that – contrary to former claims – visual data does not provide much aid in vocabulary memorization.

The three papers are as diverse as the research field that has been outlined in Section 2, adding pieces to the E-Learning mosaic: They touch different research areas, are related to different learning settings, and use different methods to achieve their respective research aim:

All contributions did carry out empirical studies and tests – either to evaluate the implemented system in comparison to human performances, or to evaluate which one of several alternatives – e.g. which type of learning content (visual, audio, text) or which type of feedback (focus on form vs. focus on meaning) – induces better learning effects. While the contribution by Wild/Haley/Bülow primarily focuses on technical and methodical aspects, Wilske/Wolska and Miyakoda/Kaneko/Ishikawa especially concentrate on empirical learner studies and the dependency between learning effects and learning setting.

In general these three papers should convince the reader of the broad distribution of Computational Linguistics approaches in the field of E-Learning.

6 Acknowledgments

This volume would not have been possible without the help of many people. First of all, we would like to thank the editor-in-chief of the JLCL, Lothar Lemnitzer, for his support and patience in the edition process. Our special thanks go to Aljoscha Burchardt for his valuable contribution to this introductory article: The preface and some thoughts that lead to the statements in Section 2 go back to him. We also would like to thank the program committee for their reviews of the submitted papers – not also for this special issue, but also for the preceding workshop: Delphine Bernhard, Lothar Lemnitzer, Henning Lobin, Cerstin Mahlow, Michael Piotrowski, Csilla Puskás, Tonio Wandmacher, Eline Westerhout, Fridolin Wild. The final responsibility for accepting or rejecting submissions or for requesting a second or third revision lies with Maja Bärenfänger and Maik Stührenberg.

We would also like to thank the Organizing Committee of KONVENS 2011 and the department of Computational Linguistics & Phonetics at Saarland University for making this workshop become possible. And of course, we want to express our thanks to all speakers who presented their research on our workshop, and to the authors for compiling paper versions of their talks. Last but not least, we would like to thank all participants of the workshop on "Language Technology and Text Technological Methods for E-Learning", that not only made the workshop very successful and interesting but also laid the foundation for this special issue.