Kai-Uwe Carstensen (Zurich)

Learning and teaching (in) Computational Linguistics

Editor's preface to *Linguistik online* 17, 5/03

When Elke Hentschel asked me to edit a volume on "Computational Linguistics" in *Linguistik online*, I was delighted, because this offered a rare opportunity for information flow "back" from this new discipline to linguistics proper. Milliseconds later, however, I realized that I would have a problem: As Computational Linguistics (CL) *is* a discipline in its own right, how could one possibly come up with a volume that would *not* look like a mere collection of articles, given the great variety of possible topics in CL (see, e.g., Carstensen et al. 2004, Mitkov 2003, Willée et al. 2002 for an overview of this variety)? I therefore decided to narrow down the thematic range of possible contributions.

Learning in its widest sense presents a suitably general subtopic for that purpose. It has recently gained importance in CL in the following respects:

1. *Learning in CL*: In the "old days" of CL, it was the computational linguist who generally provided the resources for CL systems (e.g. grammar rules). These turned out to be notoriously incomplete, however, which hindered the development of practical CL systems. With the so-called "statistical turn" in CL, the focus of interest has shifted to collecting large corpora of empirical data and inducing relevant structural information from them by means of suitable (machine learning) mechanisms.

2. *Learning with CL*: Computer Assisted Language Learning (CALL) has a long tradition but until recently has never gone beyond implementing very simple functionality. Over the years, a subfield of CL has been established that puts tools developed in CL to practical use within CALL, and which is also attempting to overcome the inherent problems in this domain (e.g. error diagnosis and feedback).

3. *Learning and teaching CL*: E-Learning (and more recently, Web-based learning or WBL) has opened new avenues for learning and teaching. A growing number of courses or course materials is going on-line, and the number of conferences and workshops discussing new concepts in WBL is increasing accordingly. It therefore may not come as a great surprise that E-learning/WBL issues are currently being discussed in CL, too (see e.g. Lemnitzer et al. to appear).

In somewhat simplified terms, these aspects can be summarized as "Learning in CL" and "Learning and teaching CL", which explains the condensed title of this volume.

The reader will notice that all the authors are or at least have been associated with Swiss universities/institutions, which is due to my restricted call for papers. My hope is that this volume is not only interesting in content but also gives a broadly representative picture of scientific "Learning and teaching (in) CL" research in Switzerland.

Kai-Uwe Carstensen and **Michael Hess** (Zurich) present a text-based, individual-oriented, problem-based approach (TIP) to teaching CL that combines two recent trends (problem-based and web-based learning) and a classical instructivist teaching style. The TIP approach is currently being introduced into the CL curriculum at the University of Zurich.

Pius ten Hacken (Swansea, formerly Basel) gives a historical argumentation on why Computer-Assisted Language Learning (CALL) has long been disregarded in CL and why this has changed during the past decade. He argues that a "revolution" in CL is the reason for growing perspectives for CALL in CL.

The role of e-learning in the computational training of translation students is the central topic of the contribution by **Susanne J. Jekat** and **Gary Massey** (Zurich/Winterthur). They show that the combination of new e-learning concepts and evaluation methods drawn from CL leads to a better teaching of translation skills.

Speech technology (speech recognition and synthesis) is one of the most relevant practical subfields of Computational Linguistics and Language Technology. In this field, learning speech models from given speech data is an important part of building functional systems, a frequent focus of interest being to maximize the *generality* of these models (e.g. for achieving speaker independence). The investigation presented by **Eric Keller** and **Brigitte Zellner Keller** (Lausanne) goes in quite the opposite direction: They are interested in rather *specific* prosody models that represent a certain range of linguistic variation (e.g. speaker or dialectal characteristics). Their research then addresses the question of how much speech data is needed to build up these models.

The article by **Manfred Klenner** and **Henriëtte Visser** (Zurich/Heidelberg) deals with anticipation-free error diagnosis in Intelligent CALL (ICALL) systems. By presenting their DiBEx system they show that explanatory feedback on faulty user input can be given by entering in a tutorial dialogue with the student that is driven by meta-reasoning based on grammatical knowledge. This procedure differs from approaches in which explicit error rules are used for providing this explanatory feedback.

Text mining - the discovery of interesting information in text collections - is a much-discussed topic in modern information technology. Often learning the similarity of documents, a prerequisite for the suitable organisation and visualisation of a text collection, requires an enormous computational effort. **Paola Merlo**, **James Henderson**, **Gerold Schneider** and **Eric Wehrli** (Geneva) present an approach in which the efficiency of the learning procedure is increased by using linguistically motivated ways of reducing the document representations.

The ambiguity of language has always been a fundamental problem confronting CL, and has been the main obstacle in developing efficient analysis mechanisms. As fast performance is vital for the practical use of CL technology, however, some research effort is directed towards improving disambiguation. In his article, **Gerold Schneider** (Geneva/Zurich) shows how syntax analysis (parsing) on the basis of classical dependency grammar benefits from learnt probabilities of syntactic relations (e.g. to resolve attachment problems).

Anne Vandeventer Faltin (Geneva) presents examples of state-of-the-art natural language processing tools in CALL systems. She argues that, although these tools are currently still imperfect, their practical use will at least encourage improvements to be made according to the needs of the language learners.

References

Carstensen, Kai-Uwe/Ebert, Christian/Endriss, Cornelia/Jekat, Susanne/Klabunde, Ralf/ Langer, Hagen (eds.) (2004): *Computerlinguistik und Sprachtechnologie. Eine Einführung*. 2nd rev. and ext. ed. Heidelberg/Berlin. (= *Spektrum Lehrbuch*).

Lemnitzer Lothar/Schröder, Bernhard (eds.) (to appear): *Computerlinguistik - neue Wege in der Lehre*. Bonn. (= Sonderheft der Zeitschrift *Sprache und Datenverarbeitung*).

Mitkov, Ruslan (ed.) (2003): *The Oxford Handbook of Computational Linguistics*. Oxford.

Willée, Gerd/Schröder, Bernhard/Schmitz, Hans-Christian (eds.) (2002): *Computerlinguistik. Was geht, was kommt?* Sankt Augustin. (= *Sprachwissenschaft, Computerlinguistik und Neue Medien* 4).