

Profile of a linguist

Ulrich Heid is a computational linguist - an assistant professor at Stuttgart University's Institute for natural language processing (*Institut für maschinelle Sprachverarbeitung, Computerlinguistik - IMS-CL*) - who likes working on practical problems. He is in fact currently responsible for bilateral industry/university projects using lexicon and corpus technologies in the areas of terminology and lexicography.

During the last two to three years his main activity has been in the development of corpus extraction tools and methods, which has many applications. In particular he has been carrying out terminology extraction and standardisation for Mercedes Benz. Other ongoing projects include the European Commission ESPRIT project DISC (Dialogue Systems and Components - best practice in development and evaluation), dictionary verification with corpora for the German language publisher Langenscheidt, and lexicon work done as part of the EAGLES initiative, on the description of multiword units.

Ulrich Heid was born and bred in Stuttgart, and did his first degree in French linguistics and history. Although at that time he still considered himself a linguist in the traditional mode, in the Romance languages department he came under the influence of Professor Christian Rohrer, who was in the process of establishing the IMS-CL.

After some work in the area of Romance linguistics and German-French contrastive grammar and lexicon, in 1988 Ulrich Heid became heavily involved in work on lexicon structure for machine translation and on dictionary construction in the framework of the Polygloss project and German national research (*Begleitforschung*) accompanying EUROTRA (Eurotra was the European Commission's major project to create an automatic translation system between the official languages of the European Union).

In 1990 he was asked to take over the planning and co-ordination of the Eurotra-7 Study, for the European Commission, on the re-use of lexical resources. This was the first time he had undertaken project co-ordination, and it led to other involvement in European projects in computational lexicography. He taught computational lexicography at the University of Lille in the framework of an ERASMUS project for a European diploma in lexicography.

From 1993 onwards he became more and more involved in the rapidly growing field of corpus

lexicography. He carried out a number of lexical projects in Stuttgart, involving corpus exploration tools and lexicon building from corpora, and began to do similar work for Langenscheidt. He collected corpora from German newspaper material and is currently using it to build up syntactical and collocational dictionaries. For some seven months at the end of last year and the beginning of this year he had an opportunity, thanks to a grant from an association seeking to further German-American cooperation in information technology (*Verein zur Förderung der deutsch-amerikanischen Zusammenarbeit auf dem Gebiet der Informatik und ihrer Anwendung e.V.*), to go as a visiting researcher to the International Computer Science Institute at Berkeley in California and to work in a project on corpus-based lexicons in Charles Fillmore's working group on Frame Semantics.

He also collaborated with Gerhard Freibott, of Krupps Industries, on a terminological database; more recently he has carried out case studies for companies in the areas of terminology extraction from corpora, partly in collaboration with the well-known terminology consulting firm of Fry and Bonthron.

Ulrich Heid sees the various fields - translation, terminology, lexicography, processing of corpora, computational linguistics - all becoming more and more closely involved with one another. Terminology, he thinks, will increasingly be extracted from text, while on the other hand terms will be looked at in their linguistic context. Certainly terminology has now become a key instrument for translation, whether human translation or computer-assisted translation. He foresees greater attention being paid to example-based translation and phrase-matching approaches, and the development of tools for syntactic information.

He intends to pursue his particular interest in corpus-based term extraction and corpus-based dictionary building. He will continue to try to figure out to what extent services for glossary building and terminology extraction can provide really useful tools for language processing, both in academic settings and in practical translation and documentation work. Certainly such work should help to provide NLP dictionaries for German which are required for further progress in machine translation. New tools are likely to become available for customising grammars and lexicons, and this is also something with which he would definitely like to be involved. ■



**Dr
Ulrich
Heid**