

0.1 Forschungsprogramm – Research Programme

The research programme of the EGK is designed in a way that will maximally exploit the exceptionally wide range of topics and methods in language technology and cognitive systems in which Edinburgh and Saarbrücken are centres of excellence. It is organized along two major axes. On the one hand, it spans the range from *language technology* to *cognitive systems*; on the other hand, it places strong emphasis both on *symbolic processing* and on *data-intensive, corpus-based methods*, as well as *hybrids* between these two extremes. Methods and perspectives are tried on several language-oriented applications and, in addition, explored on a more fundamental, computational level. This unique combination will not only present the student with a wide variety of dissertation topics, but also encourage his appreciation of a broad variety of paradigms.

The first axis, which is mirrored in the title of the EGK, is between research on cognitive systems, with the goal of understanding the way humans deal with language, and engineering approaches designed for some clearly-defined applications. Similar tensions have long existed in AI. Nowadays, there is increasing interplay between these two paradigms, as researchers from both sides recognize the value of the other. In language-oriented research, we see engineers trying to analyze and employ human strategies in their systems (e.g. through analysis of corpora of naturally occurring texts and highly controlled psycholinguistic experiments) and cognitive scientists testing their models in actual implementations. In Artificial Intelligence, there is an increasing convergence between computational frameworks and cognitive models in areas such as knowledge representation, inference, and probabilistic modeling.

Moreover, recent developments have been marked by a similar competition between symbolic, knowledge-based methods and data-intensive, stochastic, corpus-based approaches to language processing. However, we see great potential in exploring not only both approaches in isolation more deeply, but also in combining them to obtain hybrid systems that inherit the advantages from both sides. Both approaches are present in the research programme in their pure forms: symbolic processing in the forms of theorem proving and constraint programming, and data-intensive methods in the form of corpus analysis. More interestingly, both approaches arise naturally, in varying combinations, throughout the programme.

As concrete focal points for these high-level goals, we identify five thematic areas which highlight the potential for such interdisciplinary collaborative research.

- **Data-Intensive Models.** Statistical and corpus-based approaches to language processing and speech processing, the use of corpora in developing realistic semantic processing models, data-intensive models of human linguistic performance.
- **Inference.** Inference for natural language, automated reasoning, and constraint programming – and how methods from these fields can be combined to build usable systems.
- **Knowledge representation, lexicon, and ontology.** Applications of knowledge representation techniques to lexical semantics and ontological modeling.
- **Speech and Language Analysis.** Combining theoretical (symbolic) and statistical models in speech recognition, wide-coverage deep linguistic processing, models of human language processing, and human processing of computer generated text and speech.

- **Dialog and Generation.** This research is concerned with how language can be exploited to improve human-interaction with information systems. Topics include the development of human-computer dialog systems, the integration of multi-modal information, strategies for information presentation, human language production, prosody for speech synthesis and mapping concepts to speech.

Some of these areas, such as inference and data-intensive models, are more concerned with a deeper understanding of a single aspect of the general research area and its applications to language processing. Others, such as dialog and generation, or speech and language analysis, are areas where *all* available approaches can be tried. That is, they can serve as a laboratory for testing the relative efficacy of the various approaches – and for seeking areas of synergy.