THE FINITE STRING NEWSLETTER

SITE REPORT
LANGUAGE IN THE COMPUTER AGE
A NOTE ON CURRENT ACTIVITIES
IN COMPUTATIONAL LINGUISTICS IN GERMANY

Dafydd Gibbon
Umiversity of Bielefeld

A recent dramatic increase in activities in computational linguistics in the German Federal Republic prompts this note. In particular, considerable attention has been focused on the major event in this field during the last few years, the three-week SiC Summer School ("Sprache im Computerzeitalter" — "Language in the Computer Age") organized in Munich in September 1986 by the German Linguistics Society (Deutsche Gesellschaft fuer Sprachwissenschaft, DGfS), and attended by 260 linguists, including students, faculty members, and representatives of industrial research and development departments.

A mark of the importance attributed to the field, and of the impact of this Summer School, is the inauguration of a "Sektion Computerlinguistik" in the DGfS during its Annual Meeting in March 1987, with the aim of providing an official German partner in computational linguistics for internationally oriented research, under the auspices of the official representative body for linguistics in Germany. This initiative was supported by 37 eminent theoretical and computational linguists on the staffs of German universities and industrial R&D departments. The Society considered the time to be ripe for such a step, which would provide official representation in a linguistic context for researchers in computational linguistic questions but working in other fields, such as applied linguistics (by the "Gesellschaft fuer Angewandte Linguistik"), phonetic speech signal processing, or language data processing (in the "Gesellschaft fuer linguistische Datenverarbeitung").

A small amount of background information may help to put these recent developments into perspective. A distinction will be made for this purpose between four related fields, three of which were in focus at the SiC Summer School:

1. Computational linguistics (as a branch of theoretical linguistics);
2. Natural language processing in artificial intelligence contexts;
3. Linguistic data processing (as a source of tools for text documentation and analysis in the humanities);
4. Signal processing (analysis and synthesis) in phonetics, with recently developed connections to problems in phonological and prosodic parsing and generation.

The early predominance of linguistic data processing ("linguistische Datenverarbeitung") has been relativized in Germany, as elsewhere, by the increase in importance of the other three fields during the past decade, which was inevitably reflected in the SiC courses.

An interesting feature of the German scene, perhaps explicable in terms of the Whorfian hypothesis, and at any rate reflected in the title of the school, is the loan translation "Computerlinguistik" for "computational linguistics." The more abstract, theoretical sense of "computational" tends to play second fiddle; "computability," as a central notion, is "Berechenbarkeit," though no-one would (I hope) dream of referring to "Berechnungslinguistik." However, the term tends to

cloud the boundaries between the areas mentioned above even more than usual.

Historically, the current upsurge of interest in computational linguistics is, then, partly associated with the factor of rising general interest in concrete applications of computers, particularly microcomputers, with the idea that experience in language-oriented computer applications may be a useful kind of qualification for language and linguistics graduates in an age when traditional occupations such as state school language teaching are on the decline.

Another pragmatic factor is the current demand for research and development in natural language oriented branches of AI, in particular for German natural language access systems for databases and expert systems and in the context of machine translation, activities which have been increasing in importance over the past ten years in German universities, industry and in funding policy.

A factor which has perhaps contributed less to current interest than it has elsewhere is theoretical linguistics, with an interest in computational properties of natural languages, and in developments and controversies which have motivated the development of computational tools for the development and testing of linguistic descriptions. During the past decade, linguistics in Germany has been strongly descriptive, with emphasis on empirical and interpretative research in fields such as sociolinguistics, discourse analysis, and language acquisition, with less emphasis on the formal properties of language structures and processes. The inauguration of the "Sektion Computerlinguistik" in the DGfS has re-affirmed the importance of such issues, however.

A more detailed understanding of current interests may be gleaned from the SiC Summer School programme. Five types of events were offered: thematic courses, intensive practical programming courses in LISP and PROLOG (elementary and advanced), a lecture cycle by course teachers, invited and guest lectures, and panel discussions. Munich in September also offered a limitless range of additional activities, of course.

The courses were held by Hans Altmann and Joachim Jacobs (Munich), Rainer Baeuerle (Tuebingen), Robin Cooper (Madison & Edinburgh), Konrad Ehlich (Dortmund) and Jochen Rehbein (Hamburg), Hans-Juergen Eikmeyer (Bielefeld), Elisabet Engdahl (Madison & Lund), Guenter Goerz (Erlangen), Helmar Gust (Osnabrueck), Christopher Habel (Hamburg), Hans Haugeneder and Manfred Gehrke (Siemens, Munich), Roland Hauser (Munich), Wolfgang Hoeppner (Hamburg, Koblenz), Werner Kasllmeyer (Mannheim), Martin Kay (Xerox Palo Alto), James Kilbury (Trier), Juergen Krause (Regensburg), Wolfgang Kreitmair (Tuebingen), Michael Keoenig (Berlin), Sebastian Loebner (Dusseldorf), Katharina Morik (Berlin), Peter Sells (CSLI Stanford), Hans Guenther Tillman & Lieslotte Schiefer (Munich), Hans Uszkoreit (IBM Stuttgart and Stanford), Dietmar Zaefferer (Munich), and Annie Zaenen (CSLI Stanford).

There was a wide range of topics, from more specifically descriptive and theoretical linguistics topics like "Syntax and Intonation," "Quantification," "Anaphora," and "Discourse Analysis" to contemporary linguistic theories (in particular GPSG, LFG, GB) and computational linguistic questions of parsing, unification, simulation, linguistic formalisms, with linguistically relevant AI topics such as knowledge representation or question-answer systems, and finally signal processing in experimental phonetics.

In addition to lectures by SiC course teachers, invited lectures were given on computational linguistic and AI NL topics by Lauri Karttunen (Stanford), Henk Zeevat (Edinburgh), Jun-Ichi Tsujii (Kyoto), Petr Sgall (Prague), Egbert Lehmann (Siemens Munich), Claus Rainer Rollinger (IBM Stuttgart), Jaap Hoepelman (Fraunhofer Gesellschaft Stuttgart), and Thomas Christaller (Gesellschaft fuer Mathematik und Datenverarbeitung St. Augustin).

The range of courses was rather wide; nevertheless, there remained sufficient coherence for students to be able to construct individual programmes, consisting as a rule of four of the courses (two hours, every other day) with complementary and interconnected topics. The present writer's selection was Kay on "Unification Grammar," Kilbury on "Parsing in PROLOG," Uszkoreit on GPSG and, further afield, Tillmann and Schiefer on "Signal Processing in Experimental Phonetics." Reactions to the courses were varied, of course, and tended to be polar and "Relevanz" oriented; not a few participants were dismayed at finding so much theoretical linguistics and so few recipes for instant NL systems.

But the impact of the SiC Summer School on German linguistics has already been considerable and positive, not least in re-encouraging work among German linguists in fundamental theoretical questions about language and computability, and fostering an awareness of the importance of "useful formalisms." The effect of the Summer School will undoubtedly continue to be felt in the coming years.

Conference On Theoretical Aspects Of Reasoning About Knowledge

6-9 March 1988, Asilomar Conference Center, Monterey, California

SPONSORED BY: The IBM Corporation and the American Association for Artificial Intelligence.

While traditionally research in this area was mainly done by philosophers and linguists, reasoning about knowledge has been shown recently to be of great