

THE CURRENT STATUS OF THE PENMAN LANGUAGE GENERATION SYSTEM

Eduard H. Hovy

Information Sciences Institute
of the University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90292-6695
Tel: 213-822-1511
Email: HOVY@ISI.EDU

Penman is one of the largest English language generation programs in the world. Developed mainly at ISI/USC, it is the result of over 15 person-years' work, and forms the core of an investigation of the computational aspects of the theories of Systemic Functional Linguistics.

In the past year, the Penman project has undergone a number of changes. The program itself has been restructured into a software package and has been distributed to over 15 sites worldwide (mostly to academic institutions). This involved the creation of a number of auxiliary software tools and the writing of over 600 pages of documentation.

In addition, the project has entered into two collaborations that show great promise. The KOMET project at the IPSI Institute in West Germany (funded by the German Federal Government) is a new project with goals very similar to our own. In addition to funding some of our work, they are building a systemic grammar of German and are collaborating on our parsing research. The Department of Linguistics at the University of Sydney, Australia, is the home of Systemic Linguistics and is coordinating the worldwide growth of various aspects of a grammar of English, which will be provided to us periodically to be incorporated into Penman.

Current research is being pursued in three principal areas: parsing, text planning, and text variation control during generation. Dr. Bob Kasper is implementing the ability to perform inference over disjunctions in the knowledge representation language Loom, preparatory to representing Penman's grammar in Loom and then linking his parser up to perform simultaneous semantic and syntactic parsing using subsumption classification as the central operation. Dr. Eduard Hovy is continuing work on the planning of coherent multisentential and longer texts using plans based on a generalization of rhetorical relations and on schemas, aiming for the automatic production of such texts as multipage reports. Dr. John Bateman is working with others (in Germany and with Dr. Cécile Paris from ISI) on developing methods of representing interpersonal and situational features in ways which facilitate the control of Penman to produce textual variations from single input, tailored to hearers' level of expertise.