EDITORIAL

Theme: Corpus Linguistics

ICAME - International Archive of Modern / Machine-readable English - is an international organisation of scholars working within English corpus linguistics. The ICAME conferences are annual events which typically attract some 60 people every year and which have turned out to be the natural meeting point for people interested in initiating cross-institutional and cross-national collaboration on corpus work.

The following section comprises papers from this year's conference which took place in Aarhus 18-22 May 1994. They reflect some of the present trends in corpus linguistics, for instance, a growing interest in spoken language corpora even though they are in many ways tougher to compile, store and analyse than written language corpora, and in parallel corpora, be they different variations within one language or parallel corpora of various sorts where English is compared to other languages. Some of the papers reflect new roads available to corpus linguistics because of new initiatives within linguistics as well as technological developments; one example of that is the COLT project (Haslerud and Steenström) which concerns spoken teenager language (special variation, spoken language) and which is meant to be stored on a CD-ROM with the sound track and various different transcriptions for further analysis.

We are grateful to the ICAME colleagues who have contributed their papers to this volume and are happy to present some of the new trends in English corpus linguistics to the Hermes readership.

Aarhus, October 1994
Karen M. Lauridsen & Ole Lauridsen

Text Corpora and FOLIO VIEWS

Three parallel text corpora of Danish, English, and French contract law, representing 6 text types.
Size: 1.000.000 running words.

Three parallel text corpora: Danish, German, and Spanish texts on genetic engeneering, representing 4 communication levels - layman to layman, layman to expert, expert to layman, expert to expert.
Size: 1.000.000 running words.

ENTREVIS.90: Interview with Spaniards published in two weekly magazines, Tiempo and Cambio16, Madrid, 1990.
Size: 1.000.000 running words.

HABLASAL.93: 12 secret tape recordings of conversation, Salamanca, Spain, 1993. Free conversation between Spaniards of different age, sex and education.
In preparation: a machine-readable text corpus of spoken Spanish.
Estimated size: 75.000 running words.

In all the text corpora it is possible to make lexical, morphological, syntactic, semantic, textlinguistic and pragmatic studies.

Available free of charge for non-commercial research purposes.

Ole Lauridsen & Kjær Jensen
The Aarhus School of Business
Fuglesangs allé 4
DK-8210 Århus V
E-mail: OL@lng.hha.dk
         KJ@lng.hha.dk