

# BNC: Third Quarter Progress Report

Lou Burnard

15 Sept 1991

- *Task group and other related meetings* OUCS staff attended meetings of Task Groups A and D during this quarter.
- *Computer facilities* A semi-automatic security file back up system is now in operation on the the project's workstations, in addition to that provided by the OUCS VAX cluster, with which they are closely linked.
- *Software* A public domain SGML parser was obtained and installed, and has proved very useful in validating texts. Other new software installed includes *gcc* (a public domain C compiler), PC-emulation software and a windows interface for Ingres.
- *Database* A pilot implementation of the database design specified in TGCW05 was carried out and tested successfully.
- *Text Accession* No new texts have been acquired during this quarter, as effort has been directed primarily into experiments with the material so far received and in optimising its processing. Informal discussions have been held with OUP and the OUCS KDEM service about ways of speeding up throughput.
- *Text Encoding* Following dicussion of TGCW01, a prototype CDIF dtd was implemented, (based largely on the TEI dtd). A program for converting the OUP 'Pilot corpus format' into the prototype CDIF was written and tested with good results. Discussions continue with other participants as to the breakdown of responsibility for encoding various parts of CDIF.
- *Text Enrichment* Considerable progress was made towards an encoding of the linguistic annotation carried out at Lancaster in a way compatible with both existing methods and the more complex scheme proposed by the TEI.
- *Documentation* A revised version of the Document Register approved at the last Project Committee Meeting was produced. Aside from minutes and internal notes, OUCS produced working papers on problems relating to the encoding of spoken texts (TGAW11 and TGAW12) and on lexical annotation (TGDW02, TGDW03 and TGDW04), as well as two samples of CDIF texts (TGCW11 and TGCW12).