PCA42

BNC: Progress Report : 1993, first quarter

Lou Burnard

16 April 1993

- *Computer facilities.* There was a total of about one week's downtime this quarter, due to a long-awaited operating system upgrade and a totally unexpected hardware fault.

- *Personnel* No changes.

- *Text Accession.* A total of 36,577,855 words of written text have now been received from OUP, of which so far 3,304,537 were returned as insufficiently British, lacking in permissions or inadequately marked-up. A total of 1,045,197 words of spoken text has been received and is being extensively tested; it is feared that 40-50% of this may have to be returned for further work. Throughput this quarter has been poor: we received 8.5 millions from OUP, returned a total of 2.8 millions, and passed on 14.5 millions to Lancaster, about two-thirds more than last quarter. We received 13.5 million words in tagged format back from Lancaster.

  Revisions to the data capture and transfer formats used by OUP keyboarders were discussed and agreed. (see TGC W45). These should improve throughput cponsiderably, once implemented. A set of rules for establishing the "Britishness" of texts were discussed and (eventually) approved at the Technical Committee Meeting of 25 Feb (see TGA W22).

- *Text Encoding and Enrichment.*

  Other work this quarter has delayed completion of work on the CDIF header, although some progress has been made with respect to the spoken texts. Some consideration has been given to the task of encoding output from Claws in CDIF; no technical problems are anticipated.

- *Documentation and Presentations.* Aside from minutes and internal notes, OUCS staff produced working papers on *Britishness text for written corpus texts* (TGAW22); *Report on Longman Spoken Corpus Data* (TGCW42); *Proposals for amendment to OUP Transfer format* (TGCW45); *BSC — Batch (or BNC) Spell Checker* (TGCW47).

  Gavin Burnage took part in a poster session at the SERC/JFIT workshop in Keele. Dominic Dunlop gave a presentation to the Oxford Unix

Users Group. The project received visitors from Coventry University (Frances Deepwell) and from Lisbon (Paula Marques-Neto).