Preface

The British National Corpus is a collection of over 4000 samples of modern British English, both spoken and written, stored in electronic form and selected so as to reflect the widest possible variety of users and uses of the language. Totalling over 100 million words, the corpus is currently being used by lexicographers to create dictionaries, by computer scientists to make machines 'understand' and produce natural language, by linguists to describe the English language, and by language teachers and students to teach and learn it — to name but a few of its applications.

Institutions all over Europe have purchased the BNC and installed it on computers for use in their research. However it is not necessary to possess a copy of the corpus in order to make use of it: it can also be consulted via the Internet, using either the World Wide Web or the SARA software system, which was developed specially for this purpose.

The BNC Handbook provides a comprehensive guide to the SARA software distributed with the corpus and used on the network service. It illustrates some of the ways in which it is possible to find out about contemporary English usage from the BNC, and aims to encourage use of the corpus by a wide and varied public. We have tried as far as possible to avoid jargon and unnecessary technicalities; the book assumes nothing more than an interest in language and linguistic problems on the part of its readers.

The handbook has three major parts. It begins with an introduction to the topic of corpus linguistics, intended to bring the substantial amount of corpus-based work already done in a variety of research areas to the non-specialist reader's attention. It also provides an outline description of the BNC itself. The bulk of the book however is concerned with the use of the SARA search program. This part consists of a series of detailed task descriptions which (it is hoped) will serve to teach the reader how to use SARA effectively, and at the same time stimulate his or her interest in using the BNC. There are ten tasks, each of which introduces a new group of features of the software and of the corpus, of roughly increasing complexity. At the end of each task there are suggestions for further related work. The last part of the handbook gives a summary overview of the SARA program's commands and capabilities, intended for reference purposes, details of the main coding schemes used in the corpus, and a select bibliography.

The BNC was created by a consortium led by Oxford University Press, together with major dictionary publishers Longman and Chambers, and research centres at the Universities of Lancaster and Oxford, and at the British Library. Its creation was jointly funded by the Department of Trade and Industry and
the Science and Research Council (now EPSRC), under the Joint Framework for Information Technology programme, with substantial investment from the commercial partners in the consortium. Reference information on the BNC and its creation is available in the *BNC Users’ Reference Guide* (Burnard 1995), which is distributed with the corpus, and also from the BNC project’s web site at http://info.ox.ac.uk/bnc.

This handbook was prepared with the assistance of a major research grant from the British Academy, whose support is gratefully acknowledged. Thanks are also due to Tony Dodd, author of the SARA search software, to Phil Champ, for assistance with the SARA help file, to Jean-Daniel Fekète and Sebastian Rahtz for help in formatting, and to our respective host institutions at the Universities of Oxford and Bologna. Our greatest debt however is to Lilette, for putting up with us during the writing of it.