

NEW DEVELOPMENTS

A COMPUTER CORPUS OF AUSTRALIAN ENGLISH*

Greville Corbett and Khurshid Ahmad

1. *Introduction*

A corpus of newspaper texts of Australian English, the Melbourne-Surrey Corpus, is now available through the International Computer Archive for Modern English (ICAME). In this paper we give a brief description of the corpus, details of availability and an account of its use to date.

2. *The Corpus*

The corpus consists of texts taken from *The Age*. We are grateful to Mr. Peter McLaughlin, Editorial Manager of *The Age*, for permission to make the corpus available for research purposes. The texts are all editorials which appeared from September 1, 1980 to January 30, 1981. This was an interesting period, including an election, and the editorials range over a wide variety of subjects. Within the dates given, the editorials were taken on a random 93 days, to give a total of 100,000 words of text. The original motivation for assembling a corpus with these specifications was to allow easy comparison with results obtained by Nixon (1972); he used a manual corpus of 100,000 words of editorials taken from *The Times* to investigate corporate concord — the agreements found with nouns like *committee* (*the committee has/have decided*). Plural agreement in such instances is markedly less common in Australian English than in British English.

The editorials for each date selected (there are typically two per day) are stored together in a separate file (thus there are 93 such files). There is a separate word count of each file, which is available from ICAME with the magnetic tape.¹ The material is all in upper case, which has some advantages for scanning. If upper/lower case information is required for particular items, this can be got from the copies of the originals, also lodged with ICAME. The computer texts follow the line breaks of the originals, so that matching to the originals takes a minimum of time. It is hoped that this corpus will be of value to those working on the varieties of English, and in particular will complement the work being done on spoken Australian English (Clark & Fraser 1982).

Availability of the corpus

As mentioned above, the corpus is available from ICAME. This organ-

ization was set up in 1977 to collect and distribute information on English language material in machine-readable form and on research based on such material, and also to compile at the University of Bergen an archive of corpora, copies of which are made available at cost. There are conditions on the use of material, which apply to the Melbourne-Surrey corpus just as to all the other corpora distributed through ICAME, the main one being that material must be used only for bona fide research of a non-profit nature. ICAME also produces a newsletter, which is distributed free of charge. Details of officers and addresses are given in Appendix 1.

Use of the corpus

The main use of the corpus to date has been in work which lies on the border between language teaching and linguistics. We have been involved in computer assisted language learning (CALL) since 1976, more recently with various collaborators, including Margaret Rogers (Surrey) and Roland Sussex (Melbourne). We wanted to give advanced learners a means to explore foreign languages. Advanced learners know that some rules are variable (corporate concord in English, mentioned above, is one such rule); however, for them to find sufficient examples to gain an insight into the regularities involved would take a great deal of time. For this reason a package called SEARCHSTRING was developed. It allows a student sitting at a terminal to produce a concordance interactively. This package does not, of course, offer all the facilities of concordance programs such as CLOC (Reed 1978) or the Oxford Concordance Program (Hockey & Marriott 1980); that is not its purpose. Rather it provides the student with examples in three lines of context, together with the source, and it does so rapidly. Once ten examples have been found, the student is given the option of continuing or of finishing the search. Besides the texts discussed here, the student can scan Russian or German texts. For more details of SEARCHSTRING see Ahmad, Corbett & Rogers (1985); some preliminary thoughts on computer corpora for teaching purposes were recorded in Ahmad, Corbett, Rogers & Sussex (1985:126-7). A sample run of SEARCHSTRING is given in Appendix 2.

Conclusion

We hope that this small addition to the material held by ICAME will be of value to researchers, particularly in Australia, and that some readers will be prompted to consider the potential of corpora for teaching linguistics and foreign languages as well as for research purposes.

NOTES

*Corbett wishes to thank the Council of the University of Melbourne for the award of a Research Fellowship (1980-81), and the Economic and Social Research Council (UK) for a Personal Research Grant (1985-86) number C00232218. Neither award was primarily for preparing the Melbourne-Surrey Corpus, which occupied a small proportion of the research time. However, without the first award the project would not have been undertaken, and but for the second it would not yet have been completed. The last stages of preparation were facilitated by a grant from the Centre for Translation and Language Studies, University of Surrey, which is also gratefully acknowledged. A similar paper is to appear in *ICAME Journal* 1987.

[1] Besides the number of words, the number of lines, the number of characters, and the maximum line length are given for each file, as well as for the corpus as a whole. We are grateful to John Hodder for use of his wordcounting program.

REFERENCES

- Ahmad, K., Corbett, G.G., & Rogers, M. 1985. Using Computers with advanced language learners: an example. *The Language Teacher* (Tokyo) 9, 3:4-7.
- Ahmad, K., Corbett, G.G., Rogers, M. & Sussex, R.D. 1985. *Computers, Language Learning and Language Teaching*. Cambridge University Press.
- Clark, J.E. & Fraser, H. (eds). 1982. *Australian Speech Archive* (Speech and Language Research Centre, Occasional Papers). Macquarie University.
- Hockey, S. & Marriott, I. 1980. *Oxford Concordance Program, Version 1.0, User's manual*. Oxford University Computing Service.
- Nixon, G. 1972. Corporate-concord phenomena in English. *Studia Neophilologica* 44: 120-6.
- Reed, A. 1978. *CLOC user guide* (BUG 13). Birmingham University Computer Centre.

Appendix 1: Details about ICAME Journal

For information on material available
from ICAME and order forms:

Jostein Hauge
Director
Norwegian Computing Centre
for the Humanities
University of Bergen
Harald Hårfagres gate 31
P.O. Box 53
N-5014 Bergen
NORWAY

For ICAME

Stig Johansson
Editor, ICAME Journal
Department of English
University of Oslo
P.O. Box 1003
Blindern
N-0315 Oslo 3
NORWAY

Appendix 2: Demonstration run of SEARCHSTRING

OK, RUN SEARCHSTRING

[typed by student]

This program is designed to enable you to search through a large amount of text to find examples of particular words in context. Material is available in the following languages:

- (1) English
 - (2) Russian
 - (3) German
- type 1 for English
or type 2 for Russian
or type 3 for German

1

[typed by student]

Do you want to:

- (1) see the results of the search on the screen as the program runs?

or

- (2) wait to look at the results until the search is over?

Type 1 or 2

1

[typed by student]

Do you want a print-out of the results of the search? Type YES or NO

NO

[typed by student]

Input character string

GOVERNMENT

[typed by student]

The string you asked for is:

GOVERNMENT

Providing this is the string you want, press RETURN. However, if you have made a typing mistake, type ERROR and press return, in order to correct it.

[student presses return]

Do you want the search to include inflected forms?

Type YES or NO

NO

[typed by student]

Scanning ENGLISH corpus for GOVERNMENT

```
-----
- Variety      : AUSTRALIAN -
- Text type    : NEWSPAPER -
- Filename     : AGG1      -
- Subject area : POLITICS  -
-----
```

Current file contains no examples of "GOVERNMENT"

>>>>>> Press RETURN to scan the next file <<<

Scanning ENGLISH corpus for GOVERNMENT

```
-----
- Variety      : AUSTRALIAN -
- Text type    : NEWSPAPER -
- Subject area : POLITICS  -
- Filename     : AGE2      -
-----
```

WASHINGTON PRESSED FOR THIS FACILITY OR WHETHER
THE FRASER GOVERNMENT LOBBIED FOR THE REQUEST
DESPITE SOME FEARS THAT THE PRESENCE OF

OF SHARED INTERESTS. BUT, AS IN THE 1950S AND 60S,
THE PRESENT COALITION GOVERNMENT SEEMS
OVER-EAGER TO LOCK THIS COUNTRY INTO THE ROLE OF A

GREATER INDEPENDENCE AND SELF-RELIANCE IN ITS
DEFENCE PREPARATIONS, AND THE GOVERNMENT SHOULD
EXPLAIN HOW ITS LATEST DECISIONS FIT IN WITH THIS

Scanning ENGLISH corpus for GOVERNMENT

```
-----
- Variety      : AUSTRALIAN -
- Text type    : NEWSPAPER -
- Subject area : POLITICS  -
- Filename     : AGE3      -
-----
```

OF PLACE OF TRUST EITHER CIVIL OR MILITARY" IN VICTORIA TODAY. THE STATE GOVERNMENT FEARS IT DOES. IF SO, AS THE PREMIER, MR HAMER, HAS

PROHIBITION, IF IT STILL BINDS THE STATE, ARE QUITE INTOLERABLE. THE STATE GOVERNMENT HAS HAD ITS SUSPICIONS FOR SOME TIME, AND LAST YEAR ASKED

SUSPICIONS FOR SOME TIME, AND LAST YEAR ASKED THE FEDERAL GOVERNMENT TO LEGISLATE TO ALLOW THE STATES TO REPEAL OBSOLETE IMPERIAL LAWS

IS THAT THE STATUTE OF WESTMINSTER OF 1931, WHICH ENABLES THE COMMONWEALTH GOVERNMENT TO REPEAL REPUGNANT BRITISH LEGISLATION, DOES NOT

AFFAIRS. DESPITE ITS APPREHENSION, THE VICTORIAN GOVERNMENT HAS ONLY RECENTLY DECIDED IT SHOULD ACT OVER THE ACT OF SETTLEMENT. THE CHAIRMAN

OR HER OTHER REALMS.

THE STATE GOVERNMENT MUST URGENTLY ANSWER THE QUESTION WHETHER THE ACT OF

IMMEDIATE ACCESS TO THE REGISTER, OTHERS, INCLUDING GOVERNMENT BACKBENCHERS AND OPPOSITION MPS, WILL HAVE TO OBTAIN MINISTERIAL

>>>>>> Press RETURN to continue <<<<<<<<

10 examples of the string 'GOVERNMENT' have now been found. Do you wish to continue with the search? Type YES (to continue) or NO (to stop)

[at this point the student can continue, select a new string, or finish using the package]

Department of Linguistic & International Studies,
University of Surrey,
Guildford, Surrey.

[Received November 1986]