牛津应用语言学丛书

Corpus Concordance Collocation

语料库、检索与搭配

John Sinclair

上海外语教育出版社

牛津应用语言学丛书

语料库、检索与搭配

Corpus, Concordance, Collocation

John Sinclair 著

上海外语教育出版社

上海市版权局
著作权合同登记章
图字:09－1999－032号

牛津应用语言学丛书

Corpus, Concordance, Collocation

语料库、检索与搭配

John Sinclair 著

上海外语教育出版社出版发行
（上海外国语大学内）
深圳中华商务联合印刷有限公司印刷
新华书店上海发行所经销

开本 880×1187 1/32 6.25印张 242千字
1999年4月第1版 1999年12月第3次印刷
印数:1500册

ISBN 7－81046－574－0
H·585 定价:13.00元

出 版 前 言

这是一部论述如何建立语料库、进行词语检索和词语搭配分析的学术专著。作者J·辛克莱是伯明翰大学现代英语教授，主要从事话语研究和计算语言学研究，兼任英国议会、全国语言教育委员会等机构的顾问，科林斯伯明翰大学国际语言数据库的项目主持人。

近年来，计算机几经更新换代。随之，运用计算机进行的语言分析也有了很大的进展。这种分析展示了语言形式及其使用的一些不容置疑的特征，这是用传统描述方法无法得出的。作者描述了计算语言学的发展状况，详细阐明了有关语料库建立的具体过程，以及从语料库使用而得出的对语言的进一步认识。它不仅可以作为语料库语言学研究的入门读物，也可为语言学和应用语言学研究提供数据。

全书共分9章。前两章针对初步接触语料库语言学的读者的一些实际问题作出解答。语料库是存放语言素材或语料文本的“仓库”，是利用语料库对语言某个方面进行研究的基础。因此第一章谈语料库的建立。鉴于样本语料库只是一个有限的汇集，不敷使用，但实际使用中的语言的数据是无限的，因此建立一个监控语料库的设想便应运而生。监控语料库和语言本身一样，是处于不断发展之中的。它能过滤和筛选材料，并只保留必要的有关材料。伯明翰大学的第一个监控语料库已具雏形。第二章谈文本的基本处理，介绍用检索软件处理语料的基本原理和方法。第三章谈如何从语料库取得用法的证据。第四章谈词的意义和结构，对“yield”这个单词的不同意思的细微差别作了详细的分析。第五章讨论词和词组问题，把对动词短语的综述和含有“set”动词短语的分析研究结合起来，对词语选择与上下文情景的相互关系作了说明。第六章讨论词汇和语法汇合的情况，提出了从具体的实例推导出普遍规律的方法，介绍通过语料库建立新的语法体系

的初步成果。第七章提出对语料库例句的评价问题;作者认为通过文本处理得出的证据本身不一定可靠,需做评估甄别。第八章着重分析词语搭配问题。作者认为语言本身存在着有趣的词语搭配方式,一些单词由于某种原因互相结合,而另外一些单词则互相排斥;由此提出了"惯用原则"。第九章强调用通用语言来解释语言,即"以词释词"。作者总结了在研究中的得出的两个主要观点:(1)我们使用语料库来研究语言的方法比原来想像的重要的多。(2)如果词典里的定义是用一般英语写的,那么所有解释中出现的推论和细微的含意都可以用来进一步完善定义。最后,作者指出本书所介绍的内容只是对语料库语言学研究的一种尝试。80年代的语料库在今天看来也许太小了,日新月异的软件技术与实际需要相比也仍然显得过于原始。

辛克莱从80年代起就致力于建立和改进Cobuild语料库,这本书是他根据切身体验作出的总结和理论概括。从事计算机语言学研究,特别是语料库语言学研究的读者,将不难发现这本书的启发作用和指导作用。它同时也可作为语言学教师的参考书和攻读语言学和应用语言学专业的研究生的教学用书。

本社编辑部

To Angus McIntosh

Acknowledgements

This book is dedicated to Angus McIntosh, who taught me English Language many years ago. His interest in vocabulary was infectious, and his farsightedness guided me into corpus work and computing in 1960.

The work would not have been possible without the co-operation of Collins Publishers and the University of Birmingham in the Cobuild project, now Cobuild Ltd. My evidence is largely cited from the Birmingham Collection of English Texts (now The Bank of English), donated by hundreds of copyright holders, who are listed in the front of Cobuild publications. It is based upon teamwork and years of discussion with colleagues in Cobuild, whose names, again, are featured on the individual publications.

To all of these I owe a profound debt of gratitude, and to my colleagues in the School of English, many of whom have taken a keen interest in the development of ideas about corpora and lexicology in the past decade.

One or two deserve specific mention for this book. Yang Hui-Zhong gave me a most useful framework for Chapter 2, and great encouragement in the early stages. Antoinette Renouf suggested the material for Chapter 4, and has worked closely with me on the design and management of corpora throughout the period covered by this book.

The description of *of* in Chapter 6 was built up over a number of presentations in the second half of 1988; at the TESOL Summer Institute of Flagstaff, Arizona, and at the BAAL Autumn meeting in Exeter, and with colleagues and students in Birmingham. I am grateful to Gwyneth Fox who helped me shape this chapter and the incorporation of grammar into the Cobuild approach. I am also grateful for the contributions of many generous people to this study.

In Chapter 7, I am particularly indebted to Jeremy Clear for his computing help, but his work pervades the whole book; also Tim Lane who implemented the computing for Chapter 8.

In the making of this book, I am indebted to Kay Baldwin for keying the manuscript, and to Geoff Barnbrook for arranging data transfers; to Elena Tognini Bonelli for reading the manuscript and making many helpful suggestions; and to David Wilson of OUP for his great care in controlling the process of publication.

Earlier versions of each chapter have been published separately, and I list the original sources below. All the original publishers have been contacted, and their co-operation in this publication is hereby acknowledged.

Chapter 1: *Language, Learning and Community*. C. N. Candlin and T .F. McNamara (eds.) 1989. National Centre for English Language Teaching and Research, Macquarie University, Sydney.

Chapter 2: *Computers in English Language Teaching and Research*. G. Leech and C. Candlin (eds.) 1986. Longman.

Chapter 3: *Dictionaries, Lexicography and Language Learning*. R. Ilson (ed.) 1985. ELT Documents No. 120. Pergamon Press/The British Council.

Chapter 4: *Linguistics in a Systemic Perspective*. J.D. Benson *et al.* (eds.) 1988. John Benjamins Publishing Company.

Chapter 5: *Looking Up*. J.M. Sinclair (ed.) 1988. Collins.

Chapter 6: *Learners' Dictionaries: State of the art*. M. L. Tickoo (ed.) 1989. SEAMEO Regional Language Centre, Singapore.

Chapter 7: *The English Reference Grammar*. G. Leitner (ed.) 1986. Niemeyer.

Chapter 8: *Language Topics*. Steele *et al.* (eds.) 1988. John Benjamins Publishing Company.

Chapter 9: *Linguistic Fiesta*. Yoshimura *et al.* 1990. Kurosio Publishers.

The author and series editors

John Sinclair has been Professor of Modern English Language at the University of Birmingham since 1965. He is Founding Editor-in-Chief of Cobuild. His current work centres on promoting the use of large corpora of natural languages, in particular the design of computer software for storage, access, and retrieval of data. Collocation, and the interdependence of lexis and grammar, are areas of personal research. He also maintains an interest in discourse analysis, and in the integration of text structure and corpus linguistics.

Ronald Carter is Professor of Modern English Language at the University of Nottingham, where he has taught since 1979. He is Chairman of the Poetics and Linguistics Association of Great Britain, a member of CNAA panels for Humanities, and a member of the Literature Advisory Committee of The British Council. Dr Carter has published widely in the areas of language and education, applied linguistics, and literary linguistics. He is Director of the Centre for English Language Education at the University of Nottingham and from 1989 to 1992 was National Co-ordinator for Language in the National Curriculum.

Foreword

Describing English Language

The *Describing English Language* series provides much-needed descriptions of modern English. Analysis of extended naturally-occurring texts, spoken and written, and, in particular, computer processing of texts have revealed quite unsuspected patterns of language. Traditional descriptive frameworks are normally not able to account for or accommodate such phenomena, and new approaches are required. This series aims to meet the challenge of describing linguistic features as they are encountered in real contexts of use in extended stretches of discourse. Accordingly, and taking the revelations of recent research into account, each book in the series will make appropriate reference to corpora of naturally-occurring data.

The series will cover most areas of the continuum between theoretical and applied linguistics, converging around the mid-point suggested by the term *descriptive*. In this way, we believe the series can be of maximum potential usefulness.

One principal aim of the series is to exploit the relevance to teaching of an increased emphasis on the description of naturally-occurring stretches of language. To this end, the books are illustrated with frequent references to examples of language use. Contributors to the series will consider both the substantial changes taking place in our understanding of the English language and the inevitable effect of such changes upon syllabus specifications, design of materials, and choice of method.

John Sinclair, *University of Birmingham*
Ronald Carter, *University of Nottingham*

Corpus, Concordance, Collocation

In this book, John Sinclair explores the implications of his most recent research into the language of extended texts, with particular reference to the lexico-grammatical patterns such research reveals. Throughout his career, John Sinclair has dedicated himself to the analysis of corpora of extended stretches of English language data. This book represents a natural extension of his work on discourse developed during the 1970s.

It is, however, fascinating to go back a decade further to the 1960s and re-read some of the articles he wrote at that time on lexis, lexical patterns, and computer-based processing of language. In 'Beginning the Study of Lexis' (published in 1966 in Bazell, C.E., *et al.* (eds.) *In Memory of J.R. Firth*), in particular, he identified theoretical and descriptive issues and developed ideas which have only recently begun to be brought to fruition. With the increased power of modern computers, the issues raised in this book set a fascinating agenda for the next decade. It is important, however, not to overlook the consistency with which, over several decades, and often working against the grain of prevailing orientations in the field, John Sinclair has approached such an agenda. *Corpus, Concordance, Collocation* represents simultaneously a culmination and a new beginning.

Ronald Carter

牛津应用语言学丛书

Bachman, Lyle F., et al.	*Language Testing in Practice* 语言测试实践
Bachman, Lyle F.	*Fundamental Considerations in Language Testing* 语言测试要略
Brazil, David	*A Grammar of Speech* 口语语法
Brown, Gillian, et al.	*Language and Understanding* 语言与理解
Cook, Guy	*Discourse and Literature* 话语与文学
Cook, Guy, et al. (eds.)	*Principles & Practice in Applied Linguistics* 应用语言学的原理与实践
Ellis, Rod	*The Study of Second Language Acquisition* 第二语言习得研究
Ellis, Rod	*Understanding Second Language Acquisition* 第二语言习得概论
Howatt, A.P.R.	*A History of English Language Teaching* 英语教学史
Kramsch, Claire	*Context and Culture in Language Teaching* 语言教学的环境与文化
Seliger, H. W., et al.	*Second Language Research Methods* 第二语言研究方法
Sinclair, John	*Corpus, Concordance, Collocation* 语料库、检索与搭配
Skehan, Peter	*A Cognitive Approach to Language Learning* 语言学习认知法
Spolsky, Bernard	*Measured Words* 客观语言测试
Stern, H.H.	*Fundamental Concepts of Language Teaching* 语言教学的基本概念
Stern, H.H. (Allen, P., et al. eds.)	*Issues and Options in Language Teaching* 语言教学的问题与可选策略
Widdowson, H.G.	*Practical Stylistics* 实用文体学
Widdowson, H.G.	*Aspects of Language Teaching* 语言教学面面观
Widdowson, H.G.	*Teaching Language as Communication* 语言教学交际法

Contents

2 Basic text processing

3 The evidence of usage

4 Sense and structure in lexis

5 Words and phrases

6 The meeting of lexis and grammar

7 Evaluating instances

8 Collocation

9 Words about words