

Computational and

计算机与定量语言 Quantitative Studies

M. A. K. Halliday

Edited by Jonathan J. Webster



北京大学出版社
PEKING UNIVERSITY PRESS

Volume 6 in the Collected Works of M. A. K. Halliday

Computational and Quantitative Studies

M. A. K. Halliday

Edited by Jonathan J. Webster



北京大学出版社
PEKING UNIVERSITY PRESS



continuum
LONDON • NEW YORK

著作权合同登记号 图字:01-2006-7148

图书在版编目(CIP)数据

计算机与定量语言=Computational and Quantitative Studies:英文/(英)
韩礼德(Halliday, M. A. K.)著. —北京:北京大学出版社, 2007. 11
(韩礼德文集)

ISBN 978-7-301-13033-9

I. 计… II. 韩… III. 机器翻译—文集—英文 IV. H085-53

中国版本图书馆 CIP 数据核字(2007)第 171690 号

First published 2005

This collection © M. A. K. Halliday 2005

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage or retrieval system, without prior permission in writing from the publishers.

The Work is published by arrangement with the Continuum International Publishing Group.

该作品由 Continuum 国际出版有限公司授权出版。

书 名: 计算机与定量语言

著作责任者: [英]韩礼德 著

责任编辑: 刘 强

标准书号: ISBN 978-7-301-13033-9/H · 1896

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn>

电 话: 邮购部 62752015 发行部 62750672 编辑部 62767347

出版部 62754962

电子邮箱: zpup@pup.pku.edu.cn

印 刷 者: 北京宏伟双华印刷有限公司

经 销 者: 新华书店

650 毫米×980 毫米 16 开本 20 印张 420 千字

2007 年 11 月第 1 版 2007 年 11 月第 1 次印刷

定 价: 36.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究 举报电话: 010-62752024

电子邮箱: fd@pup.pku.edu.cn

Collected Works of M. A. K. Halliday

Volume 1: *On Grammar*

Volume 2: *Linguistic Studies of Text and Discourse*

Volume 3: *On Language and Linguistics*

Volume 4: *The Language of Early Childhood*

Volume 5: *The Language of Science*

Volume 6: *Computational and Quantitative Studies*

Volume 7: *Studies in English Language*

Volume 8: *Studies in Chinese Language*

Volume 8: 汉语语言研究(汉译版)

Volume 9: *Language and Education*

Volume 10: *Language and Society*

总序

胡壮麟

由香港城市大学汉语、翻译和语言学系系主任 Jonathan J. Webster 教授主编的,并由英国 Continuum 公司自 2002 年陆续出版的《韩礼德文集》,总共 10 卷,已全部出齐。北京大学出版社获得 Continuum 公司的授权后,负责文集在中国境内的出版,并组织国内专家为读者撰写导读。这无疑是我国出版界和语言学界的一件大事。

就当代语言学研究来说,20 世纪下半叶一直表现为生成语言学和功能语言学的对峙,说得具体些,当乔姆斯基在 50 年代末一度以他的转换生成语法掀起一场革命,并成为国际上,特别是美国的语言学研究的主流时,能揭竿而起并与之抗衡的便是韩礼德的系统功能语法。^①如果说乔姆斯基的理论得益于后来成为他“革命”对象的美国结构主义,韩礼德则公开宣称他继承和发展了欧洲的弗斯学派、布拉格学派和哥本哈根学派。正是这两种力量的冲突、挑战和互补推动了这半个世纪精彩纷呈的语言学研究。

《韩礼德文集》在中国的出版还具有重要的意义,那就是韩礼德的成就除了受到欧洲语言学传统的影响外,也从中国语言学传统获得滋养。他早年师从罗常培先生和王力先生,在韩礼德的论著中不时绽放出这些大师的思想火花。^②《文集》的第 8 卷是最好的历史见证,在这个意义上,《文集》的出版是一次学术上的回归,为我国语言学研究如何实现全球化和本土化的结合提供了宝贵的经验。为此,北京大学出

① 胡壮麟、朱永生、张德禄、李战子,《系统功能语言学概论》,北京大学出版社,2005 年 9 月。

② 胡壮麟,《王力与韩礼德》,《北京大学学报·英语语言文学专刊》1991 年第 1 期,第 49—57 页。
收入张谷、王辑国编,《龙虫并雕,一代宗师——中外学者论王力》,广西教育出版社,第 200—216 页。

版社采纳了我们的建议,在出版 10 卷《文集》的同时,将第 8 卷全部译成中文,另行出版中文版,以供汉语界参考。

《文集》充分反映了韩礼德所走过的治学道路,其轨迹分见于各卷的主要内容。韩礼德首先研究现代汉语(第 8 卷),打好了音系学和方言调查的扎实基础。回英国后,进入对普通语言学的研究(第 1 卷和第 3 卷),继承、发展和建立科学的语言学研究的理论,把握前进的方向。为了在欧美学术界获得一席之地,韩礼德在此时期把英语作为研究分析对象(第 7 卷)。在研究方法上,他注意第一手材料的收集(第 4 卷),将语言学研究从句子层面提高到在具体语境中出现的语篇和话语(第 2 卷),因此他的研究是经得起实际的检验的。韩礼德特别注意语言学理论的价值在于它的应用,能否说明社会生活中的问题,并为社会服务,前者见之于第 10 卷的“语言与社会”,后者反映于第 9 卷的“语言与教育”。曾经有位学者向韩礼德提问,为什么转换生成语法在中国国内打不开局面,而系统功能语法却响应者众多?这两卷的内容有助于人们找到答案。20 世纪下半叶是现代科学技术,特别是电子技术,获得飞速发展的时代。韩礼德时年七八十岁,他能在自己的晚年,关注语言与科学技术的关系(第 5 卷和第 6 卷),这种活到老、学到老的精神令人钦佩不已。

最后,《文集》只概括了韩礼德 2002 年以前的主要论著和节选,因此有关韩礼德的学术思想和成就有待我们进一步挖掘和学习。其次,这几年韩本人一直是老骥伏枥,笔耕不辍,勤于思索。如 2006 年 3 月 26 日韩礼德教授在香港城市大学的“韩礼德语言研究智能应用中心”成立大会上,做了题目为“研究意义:建立一个适用语言学”的主旨报告。韩提出适用语言学(appliable linguistics)的长期目标是为了建立语言的意义发生系统,其工作机制是以社会理据来解释和描写语义发生,可见韩礼德已经认识到语言学研究最终要解决对“意义”的描写。对这个问题,结构主义和生成主义学派是不研究的,系统功能语言学在功能语义学方面只是刚刚起步,这将是语言学界在新世纪为之共同奋斗的目标。

参加《韩礼德文集》导读编写者均为我国著名学府的学者,在系统功能语言学研究方面享有盛誉,仅在此表示感谢。

卷别/导读作者

- 第1卷 论语法/黄国文(中山大学教授/博导)
- 第2卷 语篇和话语的语言学研究/朱永生(复旦大学教授/博导)
- 第3卷 论语言和语言学/方 琰(清华大学教授)
- 第4卷 婴幼儿的语言/李战子(南京国际关系学院教授/博导)
- 第5卷 科学语言/杨信彰(厦门大学教授/博导)
- 第6卷 计算机与定量语言/林允清(北京师范大学教授/博导)
/于 晖(解放军外国语学院副教授/博士)
- 第7卷 英语语言研究/何 伟(北京科技大学副教授/博士)
- 第8卷 汉语语言研究/彭宣维(北京师范大学教授/博导)
- 第8卷 汉语语言研究(汉译版)/彭宣维(北京师范大学教授/博导)
- 第9卷 语言与教育/张德禄(中国海洋大学教授/博导)
- 第10卷 语言与社会/任绍曾(浙江大学教授)

导 读

林允清 于 晖

语言是一种有趣且丰富的社会符号现象,而语言文字之博大精深更是令人叹为观止。面对纷繁芜杂的语言现象,韩礼德一直致力于构建一个尽可能全面描述语言的理论模式。系统功能语言学,顾名思义,以系统的概念为其出发点,这是它区别于其他语言学理论的基本特征。韩礼德继承了Firth“系统”的概念。每一个系统就是语言行为中处于纵聚合(paradigmatic)关系的一组选择(choice)的集合,即在特定情景中每个人可以选用的一组语言形式。系统也是借以对某个特定语言层面的语言现象进行描述的手段。语言就是由无数个这样的子系统所构成的一个巨大系统网络。

我们知道,同一系统之间的选择是析取的关系,即“或者A,或者B”,析取解释的是可能性(possibility)的问题。可能性决定了选择的质,概率(probability)则决定了选择的量的问题。也就是说,我们不仅要描述语言中出现的选择,也要说明选择出现的概率。同一系统中一组对立的选择出现的机率是均等的,还是不等的;如果不等,那么哪些选择的可能性更大些呢?这也是系统语法面临的一个问题。概率这一概念的引进丰富了系统语法的理论体系,系统不仅仅是一个可能性系统(a system of possibility),也是一个概率系统(a system of probability)。这就要求我们在进行语言的系统分析时,不仅要作定性的分析,也要进行定量的分析。随着现代计算机科学技术的发展和语料库的不断充实与完善,对自然语言进行定量的分析研究也有了可能性。

语言是动态的,随社会系统中其他部分的动态发展而发展。因此,语言随历史的发展而发展变化;随着时间的推移,新的语言不断出现,也因此,在某一层面上看似‘相同’的语言形式在不同的社会群体中有着截然不同的形

式。语言是处于一种动态的历史发展过程。从系统功能语言学的角度来看,我们可以把语言看作是两极开放的连续体(cline)。在这个连续体上,每个选择之间的界限是模糊的。正是因为这种模糊的特性,才使得语言具备了无限的发展潜力。

本书是《韩礼德文集》丛书中的第6卷《计算机与定量语言》(*Computational and Quantitative Studies*)。本书分为三个部分:第一部分收集了韩礼德早期对计算语言学的分支领域——机器翻译的研究成果;第二部分收集了对概率语法和语料库的研究;第三部分探讨针对语言中的模糊现象,如何用计算机处理语言的意义等有关问题。

本卷第一部分是韩礼德早在上世纪五六十年代发表的关于语言学应用于机器翻译的理论基础的两篇文章。尽管距离文章发表过去了半个多世纪,在这期间计算机科学技术也有了质的飞跃,但其中提出的理论思想即使对今天机器翻译的研究仍有一定的借鉴意义。这是因为翻译首先要建立一个完善的语言学理论基础之上,一个能对源语言和目标语进行全面系统的描述和比较的语言学理论。其中,韩礼德提出机器翻译过程包含了三个阶段的工作(见第二章,“Linguistics and machine translation”)。首先要为每个层面的语言项挑选出最有可能的对等翻译;第二,注意周围语篇对这些可能的翻译选项所施加的限制作用,也就是语境对选择语言的影响,某些在词汇或语法层面合适的翻译如果从小句层面来看可能就并不妥当;第三,由于目标语和源语言的语言结构不同,在前两个阶段所选出的翻译选项在进入目标语时要受到目标语中词汇语法等关系的层层限制,才能得出最后的翻译结果。第一个阶段所得出的翻译结果就是我们通常所说的“逐字逐句”的翻译,这种译文通常是生硬蹩脚甚至是错误的,错误的原因往往是因为忽略了语境的限制作用。小句以及比小句更大的语言单位构成了小句以下的词及词组的语言环境,由于语境的作用,所有与语境不符的翻译都被筛选掉,所以语境在翻译过程起到了至关重要的作用。韩礼德提出的同义词库法(thesaurus method)也建立在把单词按照语境因素排列顺序的基础上(见第一章,“The linguistic basis of a mechanical thesaurus”)。最后要指出的是,机器翻译的前景更大程度上取决于语言学理论体系的完善而并不仅仅是计算机科技的发展。

本卷第二部分收录了韩礼德近一二十年来关于概率语法和语料库的研究成果。语料库的建设并非只是简单的语料收集过程,同时也是完善和丰富理论语言学的重要手段。通过对语料库中出现的各种语法结构进行分析统计,可以进一步提高人们对于语言结构的认识。那么在选择语料库时,是

以口语语料库还是以书面语语料库为主呢? 韩礼德认为日常发生的口语语料库对于不断构建和完善语言学理论起着重要的作用(见第八章, “The Spoken language corpus: a foundation for grammatical theory”)。为什么这种随意性的即时口语比书面语更为重要呢? 这不仅仅是因为口语在人类历史上远比书面语出现得要早, 更主要是因为是在口语中, 语言的潜势得以最充分地展现。人们平日的口语交谈是一种相对无意识的动态过程, 交谈者随着环境语境的变化不断及时调整自己交流的内容, 体现在语言上是丰富的语法语义结构。正是口语的这种即时性和创造性极大地丰富了人类的语言结构。只有通过分析口语语料库才能展现语言的全面图景, 才能构建日趋完善的语法理论。

早在上世纪 50 年代, 韩礼德就对中国古典著作《元朝秘史》中的语言结构进行了定量分析, 其目的在于证明语言结构在语篇中出现的频率正体现了语法系统中概率的问题(‘frequency in text was the instantiation of probability in the grammar’, 见第四章, “Probabilistic grammar and the corpus”)。语料库的用途之一就在于为语言结构发生的频率提供较为可靠的数字依据, 从而可以推断出语言中各种语法结构的概率分布情况。在 20 世纪 90 年代以伯明翰大学为基地建设的 COBUILD 语料库项目中, 韩礼德与 Zoe James 合作完成了 COBUILD 项目的部分研究课题, 对英语中限定性小句的归一度和主要时态做量化研究(见第六章, “A quantitative study of polarity and primary tense in the English finite clause”), 进一步验证了系统中概率的问题, 并指出, 有些系统中的选择机会是均等的, 没有所谓标记性的结构, 而有些系统中的选择确实带有明显的倾向性的, 频率高的结构就是非标记性的。概率与系统的概念是密切相关的。因为概率的相对性, 概率的计算是相对于其他处于同一组纵聚合关系的选择而言, 这样一组选择就构成了系统。

对语言结构发生频率的计算是我们利用语料库的重要目的之一。然而, 我们利用语料库的目的不仅限于此(见第五章, “Language as system and language as instance: the corpus as a theoretical construct”)。比如, 通过对语料库的分析, 我们可以看出语言分布的频率与语篇体裁的类型亦有关系; 通过分析语料库, 我们可以考察从一个给定系统中进行的选择在多大程度上受到以往选择的影响, 这种概率有多大; 我们可以通过分析语料库来考察语言在人类历史的发展过程(phylogenesis), 人类个体对语言的习得进程(ontogenesis), 个体语篇的动态发展过程中(logogenesis), 在上述过程中语言结构从简到繁的概率分布; 对语料库的考察还可以帮助我们发现并存的

系统之间的关联程度如何。例如,从语法关系角度来看,小句复合体中存在着“并列关系”(parataxis)与“主从关系”(hypotaxis)两个系统;从逻辑—语义关系角度来看,小句复合体包含“投射”(projection)与“扩展”(expansion)两个系统。这两组系统的关系是并存的。如果对两组系统的选择的频率分布进行统计,也对系统间交叉结合所得的选择做概率的分析,我们就可以得出:同一系统中的一组对立选择在语篇中的分布情况与该语篇所属的语篇体裁类型有关系,因语篇体裁类型而异(见第三章,“Towards probabilistic interpretations”)。

由此我们可以看出,语料库语言学与理论语言学并不是两个对立的学科,语料库的研究是拓展语言学理论的一个重要途径和源泉。我们在运用语料库分析语言现象时也并非全无规律可言,而是要遵循一定的原则(见第七章,“Quantitative studies and probabilities in grammar”)。我们考察的系统是要普遍存在的,有大量生活的实例可供分析。我们要研究的对象也必须具备系统性,也就是至少具备两种以上的选择。比如,在英语中,及物动词的语态可以有两种选择,主动或被动。而不及物动词只能选择主动态,换言之,不及物动词的语态不具备系统性。所以,对不及物动词语态的系统研究就没有太多的理论意义。

第三部分收集的文章集中探讨如何把语言学理论应用到人工智能的技术中去。人类研究计算机智能化的历史并非是一帆风顺的,大致可以划分为几个阶段(见第十一章,“Computing meanings: some reflections on past experience and present prospects”)。上世纪 50 年代,人们才刚刚开始考虑运用计算机进行语言翻译的工作,但那时的研究工作是以数学的逻辑运算为主,语言并非主要的考虑对象。到了 20 世纪 60 年代中期,人们渐渐意识到语言的重要性,语言学也随之逐渐替代了逻辑成为机器翻译研究的重要理论基础。20 世纪 80 年代,随着新一代计算机的诞生,计算机的语言处理系统所能处理的自然语言的复杂程度也有所增加。20 世纪 90 年代中期至今,自然语言已渐渐成为研究计算机智能的中心课题。智能计算机应当用自然语言来进行计算,应当能够处理诸多复杂的语言现象。

如何实现计算机的智能化,真正实现人机对话,使计算机能够接收处理人类自然语言,这也是语言学家长期以来面临的一个难题。如果要把自然语言作为计算机智能中的元语言,就要克服由于自然语言固有的特点而引发的种种难题,如,人们谈话中不同的语气,肯定的、试探性的、夸张的、含糊的等等,语言的歧义,不明指称,语言的重复,口误或笔误,有意识的答非所问,讽刺等等诸多问题(见第十章,“Fuzzy grammatics: a systemic functional

approach to fuzziness in natural language”)。这些现象给计算机在自然语言的处理过程设置了重重障碍。

要想克服这些困难,我们需要构建一个完善的语法理论体系来描述这些语言现象,这个理论模式要能对语言从系统、语境、功能、模糊性和发展的几个角度来分别描述(见第九章,“On language in relation to fuzzy logic and intelligent computing”)。首先,我们要从系统的角度来描述语言。从系统的观点来看,所谓句子的描述,就是要明确衍生出各个句子成分的系统以及系统中其他对立的选择项。通过对比其他可能的选择与实际做出的选择,来解释该选择要表达的意义或功能。也可以说,选择就是意义。系统代表了意义潜势。系统语言学对语言的描述实质上是解释语言成分的纵聚合关系。其次,从语境的角度来描述语言。语言在本质上是一套社会符号系统,是人们用于日常交流、传播知识、文化等的重要工具。每一种语言的繁衍、发展与演变都植根于一定的文化环境中。语境可以用语场、语旨、语式三个变量来描述。这三个变量揭示了语境与语言之间系统有机的联系。第三,要研究语言的功能。功能主义的目的就是探讨语言与语言功能之间的关系。如果语言是在实现其功能的过程中不断演变的,那么其社会功能也会影响到语言本身的特征。同样,如果语言是一种有目的性的社会过程,那么这一社会目的就会影响到语言在不同层面的实现特点。第四,我们要接受语言中的模糊现象。语言是模糊的,有很多不确定的因素在内。第五,我们要从发展的视角来看待语言。不仅要探察在人类历史的长河中语言的进化过程,也要研究人类从婴儿到成人这个过程中语言的发展规律。

系统功能语言学以其理论优势,为我们提供了一个相对完善的理论体系,为自然语言在计算机技术中的应用奠定了基础。这是因为系统功能语法是全方位的,不仅仅局限在某一个层面;系统语法提供给我们描述语言的手段是多样的;系统语法注重语言的历史发展、现状和未来发展趋势;系统语法并不是一个独立的学科,而是被包容于范围更广的社会符号学的研究领域,并和其他人文学科密切相关;系统语法本身具有包容性,并不规定所谓正统的理论教条,为理论的丰富提供了足够的空间;系统理论强调语言的不确定性,不同层面之间的界限是模糊的。只有当我们能够正确地描述语言时,我们才有可能把人类的自然语言作为计算机智能的元语言,才能够有一天真正实现人机对话,实现计算机的高度智能化。

PREFACE

The works presented in this volume on language in relation to computing should not be read simply as a retrospective on Professor Halliday's work in this area over the last fifty years. Rather, this volume is better thought of as a prospective from a scholar whose foresight has grown out of decades of experience working with language, first as a language teacher, then as a grammarian, never seeking to tame its 'ferociously complex' (Chapter Eleven) nature, only wanting to 'engage with it seriously and in appropriate theoretical terms' (Chapter Nine). For Halliday, this has meant modelling language paradigmatically, contextually, functionally, fuzzily and developmentally.

In the late 1950s, while working at the Cambridge Language Research Unit on a project concerned with machine translation, Professor Halliday realized the need to 'formalize paradigmatic relations, those of the system', but as he recalls, 'I did not know how to do it – and I totally failed to persuade anyone else of this!' (Chapter Seven). Since then, however, as the chapters in this volume clearly evidence, he has succeeded in advancing our knowledge of how one can indeed 'model language paradigmatically: that is, as choice, with the system network as organizing concept' (Chapter Nine).

With the advent of the spoken-language corpus and advances in computing technology, it has become possible to conduct quantitative studies of naturally occurring, spontaneous speech, taking us beyond what can be recovered by introspection to discover language as it is actually used in the context of situation (Chapter Eight). Speech in situ opens a window to the soul of language – 'as semiotic

system-&-process'. The ranks and levels, properties and functions of an underlying, invisible system are revealed as one interrogates the corpus to find out who is doing what with language, where, when, why and how. Corresponding to the representation of context in terms of field (what is going on), tenor (who is involved), and mode (in what medium) are metafunctions associated with processes of construing human experience (ideational), enacting social relationships (interpersonal), and creating discourse (textual). It is not surprising given the complexity of human experience and dynamics of social relationships that language is 'inherently fuzzy', and 'perhaps the single most complex phenomenon in nature' (Chapter Eleven). Language is not dysfunctional because it is fuzzy. Rather, precisely because of its dynamic character, its potential 'is indefinitely large – it is always capable of growing further, without any limits that we are aware of' (Chapter Nine). Language never ceases to amaze.

While the focus of study may be on 'everyday discourse as contextualized in real life' – what some might call the 'commonplace', the system underlying that discourse cannot be adequately described only in commonsense terms, without reference to designed systematic knowledge. As Professor Halliday puts it, 'One would not expect to design a robotic system on the basis of Year Six mathematics!' (Chapter Nine). Whether the endeavour is machine translation, corpus linguistics or intelligent computing, what is needed is a theory of language which will provide a fuller, more realistic picture of language as a system rich in meaning potential. Systemic-functional grammar accomplishes just that.

Professor Halliday foresees the day when computing – 'the most future-oriented of all present-day human activities' – becomes computing with meaning, 'when language comes to function not just as the object of computing but as its instrument' (Chapter Eleven). Before this can happen, however, two 'noticeable disjunctions' must be addressed: (i) the disjunction between 'language and knowledge', which needs to be reconceptualized as a single unified semiotic system with two related levels of representation, the lexico-grammatical and the semantic; and (ii) the disjunction between 'the instance and the system of which it is an instance' (Chapter Eleven), recognized as simply different perspectives on the same phenomenon.

The lasting significance of the chapters in this volume comes from the fact that their insights are grounded in a theory which neither idealizes nor trivializes the object of study, but diligently strives to paint a picture of language which captures its true character.

ACKNOWLEDGEMENTS

We are grateful to the original publishers for permission to reprint the articles and chapters in this volume. Original publication details are provided below, and also at the beginning of each chapter.

'The linguistic basis of a mechanical thesaurus, and its application to English preposition classification' from *Mechanical Translation*, 3:3, December 1956, pp. 81–8. © Massachusetts Institute of Technology. Reprinted by permission.

'Linguistics and machine translation' from *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, Band 15, 1962, Heft 1/2. Reprinted by permission of Akademie Verlag, Berlin.

'Towards probabilistic interpretations' from Eija Ventola (ed.), *Trends in Linguistics: Functional and Systemic Linguistics: Approaches and Uses*, Berlin and New York: Mouton de Gruyter, 1991.

'Corpus studies and probabilistic grammar' from Karin Aijmer and Bengt Altenberg (eds), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, London and New York: Longman, 1991. Reprinted by permission of Longman Group Limited.

'Language as system and language as instance: the corpus as a theoretical concept' from Jan Svartvik (ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82*, Berlin: Mouton de Gruyter, 1991.

'A quantitative study of polarity and primary tense in the English finite clause' from John M. Sinclair, Michael Hoey and Gwyneth Fox (eds), *Techniques of Description: Spoken and Written Discourse*, London: Routledge, 1993. Reprinted by permission of Routledge.

'Quantitative studies and probabilities in grammar' from Michael

Hoey (ed.), *Data, Description and Discourse: Papers on the English Language in Honour of John M. Sinclair on his Sixtieth Birthday*. London: HarperCollins, 1993. © 1993 Michael Hoey.

'The spoken language corpus' from Karin Aijmer and Bengt Altenberg (eds), *Proceedings of ICAME 2002: The Theory and Use of Corpora*, Göteborg 22–26 May 2002, Amsterdam: Editions Rodopi. Reprinted by permission of Editions Rodopi.

'On language in relation to fuzzy logic and intelligent computing' © 1995 IEEE. Reprinted, with permission, from *Proceedings of 1995 IEEE International Conference on Fuzzy Systems. The International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium*. IEEE: Piscataway NJ, 1995.

'Fuzzy grammatics: a systemic functional approach to fuzziness in natural language' © 1995 IEEE. Reprinted, with permission, from *Proceedings of 1995 IEEE International Conference on Fuzzy Systems. The International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium*. IEEE: Piscataway NJ, 1995.

'Systems of the English clause: a trial grammar for the PENMAN text generation project' reprinted by permission of Information Sciences Institute, University of Southern California.

CONTENTS

Preface	xvii
Acknowledgements	xix
 PART ONE MACHINE TRANSLATION: THE EARLY YEARS	 1
Editor's Introduction	3
1 The Linguistic Basis of a Mechanical Thesaurus	6
2 Linguistics and Machine Translation	20
 PART TWO PROBABILISTIC GRAMMAR AND THE CORPUS	 37
Editor's Introduction	39
3 Towards Probabilistic Interpretations	42
4 Corpus Studies and Probabilistic Grammar	63
5 Language as System and Language as Instance: The Corpus as a Theoretical Construct	76