

Incorporating Corpora

TRANSLATING EUROPE

Series Editors: Gunilla Andeman, *University of Surrey, UK*
Margaret Rogers, *University of Surrey, UK*

Other Books in the Series

In and Out of English: For Better for Worse
Gunilla Anderman and Margaret Rogers (eds)
Voices in Translation: Bridging Cultural Divides
Gunilla Anderman (ed.)

Other Books of Interest

Beyond Boundaries: Language and Identity in Contemporary Europe
Paul Gubbins and Mike Holt (eds)
Contemporary Translation Theories (2nd Edition)
Edwin Gentzler
Constructing Cultures: Essays on Literary Translation
Susan Bassnett and André Lefevere
Cultural Encounters in Translation from Arabic
Said Faïq (ed.)
Identity, Insecurity and Image: France and Language
Dennis Ager
Language Planning and Policy in Europe, Vol.1: Hungary, Finland and Sweden
Robert B. Kaplan and Richard B. Baldauf Jr. (eds)
Language Planning and Policy in Europe, Vol.2: The Czech Republic, The European Union and Northern Ireland
Richard B. Baldauf Jr. and Robert B. Kaplan (eds)
Literary Translation: A Practical Guide
Clifford E. Landers
Politeness in Europe
Leo Hickey and Miranda Stewart (eds)
The Pragmatics of Translation
Leo Hickey (ed.)
Translating Milan Kundera
Michelle Woods
Translation, Linguistics, Culture: A French-English Handbook
Nigel Armstrong
Translation, Power, Subversion
Román Alvarez and M. Carmen-Africa Vidal (eds)
Translation and Nation: A Cultural Politics of Englishness
Roger Ellis and Liz Oakley-Brown (eds)
Translation and Religion: Holy Untranslatable?
Lynne Long (ed.)
Urban Multilingualism in Europe
Guus Extra and Kutlay Yagmur (eds)
Words, Words, Words. The Translator and the Language Learner
Gunilla Anderman and Margaret Rogers

For more details of these or any other of our publications, please contact:
Multilingual Matters, Frankfurt Lodge, Clevedon Hall,
Victoria Road, Clevedon, BS21 7HH, England
<http://www.multilingual-matters.com>

TRANSLATING EUROPE

Series Editors: Gunilla Anderman and Margaret Rogers
University of Surrey

Incorporating Corpora

The Linguist and the Translator

Edited by

Gunilla Anderman and Margaret Rogers

MULTILINGUAL MATTERS LTD

Clevedon • Buffalo • Toronto

Library of Congress Cataloging in Publication Data

Incorporating Corpora: The Linguist and the Translator / Edited by Gunilla Anderman and Margaret Rogers.

Includes bibliographical references and index.

1. Translating and interpreting. 2. Corpora (Linguistics)

I. Anderman, Gunilla M. II. Rogers, Margaret

P306.2.I525 2007

418'.02-dc22

2007000090

British Library Cataloguing in Publication Data

A catalogue entry for this book is available from the British Library.

ISBN-13: 978-1-85359-986-6 (hbk)

ISBN-13: 978-1-85359-985-9 (pbk)

Multilingual Matters Ltd

UK: Frankfurt Lodge, Clevedon Hall, Victoria Road, Clevedon BS21 7HH.

USA: UTP, 2250 Military Road, Tonawanda, NY 14150, USA.

Canada: UTP, 5201 Dufferin Street, North York, Ontario M3H 5T8, Canada.

Copyright © 2008 Gunilla Anderman, Margaret Rogers and the authors of individual chapters.

All rights reserved. No part of this work may be reproduced in any form or by any means without permission in writing from the publisher.

The policy of Multilingual Matters/Channel View Publications is to use papers that are natural, renewable and recyclable products, made from wood grown in sustainable forests. In the manufacturing process of our books, and to further support our policy, preference is given to printers that have FSC and PEFC Chain of Custody certification. The FSC and/or PEFC logos will appear on those books where full certification has been granted to the printer concerned.

Typeset by Datapage International Ltd.

Printed and bound in Great Britain by MPG Books Ltd.

Contents

Acknowledgement	vii
Contributors	ix
Introduction	1
1 The Linguist and the Translator <i>Gunilla Anderman and Margaret Rogers</i>	5
2 Parallel and Comparable Corpora: What is Happening? <i>Tony McEnery and Richard Xiao</i>	18
3 Universal Tendencies in Translation <i>Anna Mauranen</i>	32
4 Norms and Nature in Translation Studies <i>Kirsten Malmkjær</i>	49
5 Being in Text and Text in Being: Notes on Representative Texts <i>Khurshid Ahmad</i>	60
6 Translating Discourse Particles: A Case of Complex Translation <i>Karin Aijmer</i>	95
7 The Translator and Polish–English Corpora <i>Tadeusz Piotrowski</i>	117
8 The Existential <i>There</i> -construction in Czech Translation <i>Jiří Rambousek and Jana Chamonikolasová</i>	133
9 Corpora in Translator Training and Practice: A Slovene Perspective <i>Špela Vintar</i>	153
10 NP Modification Structures in Parallel Corpora <i>Tamás Váradi</i>	168
11 A Study of the Mandative Subjunctive in French and its Translations in English: A Corpus-based Contrastive Analysis <i>Noëlle Serpollet</i>	187

- 12 Perfect Mismatches: 'Result' in English and Portuguese
Diana Santos 217
- 13 Corpora for Translators in Spain. The CDJ-GITRAD Corpus
and the GENTT Project
Anabel Borja 243

Acknowledgements

A number of people have helped us during the various stages of compiling this edited volume. A comprehensive overview of the use of electronic corpora in Translation Studies in a number of European areas required some diligence in tracking down language specialists in different countries. As a result, speed of production unfortunately had to be sacrificed to other considerations and we are most grateful to the contributors for the patience that they have shown. In addition, we owe a debt of gratitude to Multilingual Matters for allowing us the time to complete the volume in keeping with our original concept. We would also like to thank Akadémia Kiadó for allowing us to reproduce *NP Modification Structures in Parallel Corpora* in a revised version in Chapter 10. Last but certainly not least, our thanks, as always, go to Gillian James, this time more than ever; we are grateful not only for her attention to detail, persistence and patience, but also for her committed co-operation and initiative in bringing the volume together.

Gunilla Anderman
Margaret Rogers
Guildford
October 2006

Contributors: A Short Profile

Gunilla Anderman († April 2007) received her PhD in theoretical linguistics from University College London. She was for many years Professor of Translation Studies at the University of Surrey where she taught at post- as well as undergraduate level. Her most recent publications included 'Linguistics and Translation' in *A Companion to Translation Studies*, edited by P. Kuhiwczak and K. Littau (2007) and *In So Many Words: Translating for the Screen* (2007) co-edited with J. Díaz Cintas (2007).

Khurshid Ahmad is Professor of Computer Science at Trinity College Dublin, Ireland and a visiting professor at the University of Surrey. He has published widely on a range of topics related to natural language, using methods and techniques of corpus linguistics and artificial intelligence to build multilingual terminology systems. His initial interest in corpora dates back to the late 1970s when he built corpora of Russian and German language texts for use in Computer-Assisted Language Learning. He has also published on the development of language in children using methods and techniques of neural computing. Currently, he is involved in building information-extraction systems for analysing Arabic and Chinese (news) corpora.

Karin Aijmer is Professor of English at Göteborg University, Sweden. Among her recent publications are *Conversational Routines in English* (1996), *Convention and Creativity* (1996) and *English Discourse Particles. Evidence from a Corpus* (2002). She is one of the editors of *English Corpus Linguistics. Studies in Honour of Jan Svartvik* (1991). More recently she has edited the volume *Pragmatic Markers in Contrast* (with Anne-Marie Simon-Vandenberg, 2006). She is the leader of the project 'Contrastive Studies in a Translation Perspective'. She is also responsible for the Swedish part in building a spoken learner corpus of advanced Swedish learners and has been involved in the collection of the Swedish component of the International Corpus of Learner English.

Anabel Borja is a sworn translator who has worked professionally for publishers, multinational companies, law firms and the courts. She is Senior Lecturer in legal translation at Universitat Jaume I, Castellón, Spain, where she co-directs the GITRAD research group. Her research is based on the comparative analysis and classification of legal texts

through the use of electronic corpora based on the concept of genre analysis. She is currently working on the creation of a Virtual Campus for Legal Translation and on an Expert Knowledge Management System Project for Legal Translation. Her publications include: *El texto jurídico inglés y su traducción al español* (2000) and the manual *Estrategias, materiales y recursos para la traducción jurídica* (2007).

Jana Chamonikolasová is Assistant Professor in the Department of English and American Studies in the Faculty of Arts, Masaryk University in Brno, Czech Republic. Her research fields include functional syntax (functional sentence perspective) and contrastive text analysis. She has published many scholarly papers and chapters in books on Czech–English contrasts, particularly in relation to prosody.

Kirsten Malmkjær is Professor of Translation Studies and Literary Translation at Middlesex University, UK, where she teaches on the BA in Translation Studies and the MA in the Theory and Practice of Translation. Her research specialisms – on which she has published and lectured widely – are translation theory, with a particular emphasis on matters of language and of philosophy, research methods, and the language of Hans Christian Andersen. Her most recent books include *Linguistics and the Language of Translation* (2006).

Anna Mauranen is Professor of English at the University of Helsinki, Finland. Her main research interests currently focus on English as a lingua franca and modelling spoken language. Her publications deal with corpus linguistics, speech corpora, translation studies, contrastive rhetoric and academic discourses. Her major publications include *Linear Unit Grammar* (with J. Sinclair, 2006), *Translation Universals – Do They Exist* (co-edited with P. Kujamäki, 2004), *Academic Writing, Intercultural and Textual Issues* (co-edited with E. Ventola, 1996) and *Cultural Differences in Academic Rhetoric* (1993). She is currently running a corpus-based research project on spoken English as a lingua franca, the ELFA corpus.

Tony McEnery is Professor of English Language and Linguistics, Lancaster University and Director of Research at the Arts and Humanities Research Council in the UK. He has published widely in the area of corpus linguistics where his major interests are the contrastive study of aspect, epistemic modality and corpus-aided discourse analysis. Previous books include *Corpus Linguistics* (with A. Wilson, 1996), *Corpus Annotation* (with R. Garside and G. Leech, 1997), *Aspect in Mandarin Chinese* (with R. Xiao, 2004), *Corpus-Based Language Studies* (with R. Xiao and Y. Tono, 2005) and *Swearing in English* (2005).

Tadeusz Piotrowski graduated from Wrocław University in 1980. Following a period working as a full-time translator, he began his research career working alternately at Opole University and Wrocław University, focusing on lexicography, corpus linguistics and translation. He has compiled and edited a dozen or so bilingual dictionaries, published three books on lexicography, and about 100 papers and reviews. He has also served as an associate editor for the *International Journal of Lexicography* (OUP). He is currently Professor of Linguistics at Opole University.

Margaret Rogers has a PhD in Applied Linguistics and is Director of the Centre for Translation Studies at the University of Surrey, where she teaches terminology, translation and text analysis on the undergraduate and postgraduate programmes in Translation Studies. She initiated the Terminology Network in the Institute of Translation and Interpreting, UK, and is a founder member of the Association for Terminology and Lexicography. Her publications focus on terminology in text, particularly in LSP texts in translation.

Jirí Rambousek teaches in the Department of English and American Studies in the Faculty of Arts at Masaryk University in Brno, Czech Republic. His research fields, in which he has published a number of journal papers and book chapters, include the translation of fiction – from Václav Havel's plays to children's literature – and the use of parallel corpora. He is also a practising translator.

Diana Santos's main interests are translation, corpora and evaluation. She has participated in the development of several corpus projects under the scope of Linguateca, such as AC/DC, Floresta and COMPARA. She holds a PhD from Instituto Superior Técnico, Lisbon, Portugal, 1996, on corpus-based semantic studies, leading to the publication of *Translation-based Corpus Studies: Contrasting English and Portuguese Tense and Aspect Systems* (2004). She is currently a research scientist at SINTEF, Oslo, leading the Linguateca project.

Noëlle Serpollet recently completed her PhD at Lancaster University with Professor Geoffrey Leech on the topic *Should and the Subjunctive: A Corpus-based Approach of Mandative Constructions in English and in French*. She is currently a lecturer in English linguistics and phonetics at the University of Orléans, France, and works on developing and analysing a large corpus of spoken French. Her research fields are corpus linguistics, contrastive analysis, translation theory and the history of the English language. Her theoretical framework is the French theory of enunciative and predicative operations (by A. Culioli).

Tamás Váradi is a graduate of Eötvös Loránd University, Hungary, in English, Spanish and General Linguistics. His early research interests focus on second language acquisition, error analysis and communication strategies. Following a period in the first half of the 1990s as a guest researcher at Lancaster University and as Hungarian Lector at SSEES, University of London, he returned to Hungary, where he is currently Head of the Department of Language Technology at the Research Institute of Linguistics, Hungarian Academy of Sciences.

Špela Vintar is an Assistant Professor in the Department of Translation at the University of Ljubljana, Slovenia. Her research interests include computational terminology, extraction of multiword expressions, multilingual text mining, translation technologies and semantic natural language processing. She has been involved in a wide range of projects covering text mining, corpus building and translation, and has published extensively in scholarly journals and edited collections on related topics.

Richard Xiao is a researcher in the Linguistics Department, Lancaster University, UK. His major research interests include corpus linguistics, contrastive and translation studies, and aspect theory. He is a co-author of *Aspect in Mandarin Chinese* (with Tony McEnery, 2004) and *Corpus-Based Language Studies* (with Tony McEnery and Y. Tono, 2005). The results of his research have been presented in refereed journals and at international conferences.

Introduction

Broadly speaking, this volume falls into two parts. The first part discusses some aspects of Corpus Linguistics and its emerging role in Translation Studies. The second part includes a number of chapters dealing with corpora and translation in specific languages.

The first chapter – ‘The linguist and the translator’ by Gunilla Anderman and Margaret Rogers – aims to trace some developmental links between aspects of Linguistics in the Firthian tradition and the subsequent use of corpora to study translation and translation-related phenomena. This chapter sets the scene for much of what is to follow in that it historically roots these developments in the British linguistic tradition of the study of actual texts, including semantic issues. This approach to the study of languages experienced a resurgence, once ever larger quantities of text could be stored and processed using computers and software tools.

The contribution by Tony McEnery and Richard Xiao, ‘Parallel and comparable corpora: What is happening?’, provides information not only about the kind of corpus data that *are* available to the researcher and translator, but also about what is *not* yet available. In doing so, the authors review different types of multilingual corpora and explore the different purposes to which they may be put. A strategy for building multilingual corpora is outlined that meets both the needs of researchers and the needs of eventual consumers of corpus-based research. In presenting this programme of research, they draw on experience gained in a number of multilingual corpus-building projects in Lancaster over the past years, presenting the results of research past and present, while looking forward to what the future may hold.

A close association has been established in the Translation Studies literature between corpus-based studies of translations and what have become known as ‘translation universals’. In her comprehensive survey of ‘Universal tendencies in translation’, Anna Mauranen starts out by charting the provenance of current studies on universals in the earlier Translation Studies literature. As Mauranen points out, however, the topic has not been uncontroversial. She reminds us of the useful distinction – first established in typological studies of languages – between absolute universals and general tendencies, the latter currently presenting the more promising path. Before moving on to a discussion of each of what she cautiously calls the hypothesised translation universals, Mauranen

reviews some of the methods used to study these empirically. The universals that she succinctly surveys are: explicitation, simplification, conventionalisation, unique items in translation ('under-representation'), interference and untypical collocations.

In her closely argued paper 'Norms and nature in Translation Studies', Kirsten Malmkjær sets out to clarify how universals relate to norms. She argues, basing her case on sociology and theoretical linguistics, that norms are sociocultural whereas universals are cognitive. In considering those processes that are usually given as examples of universal tendencies, Malmkjær argues that only one is a candidate for the status of 'universal' according to her interpretation, namely the under-representation of linguistic features that are characteristically found in the target language. In conclusion, Malmkjær argues that corpus studies are better suited to the search for evidence of norms rather than universals.

In all corpus-based work, the issue of representativeness arises. It is this perspective that is considered in Khurshid Ahmad's paper on 'Being in text and text in being: Notes on representative texts'. Over the last 50 years, dictionary publishers and linguists have created a number of corpora, starting from 1-million-word corpora rising to billion-word corpora. The implicit claim of the corpus compilers is that the texts they have selected are in some sense representative – representative perhaps of a large number of language users or representative in the sense of a standard. In this contribution, the composition of four major corpora is discussed, indicating that there is a measure of objectivity in the selection of many text samples. Given that the pioneers of corpus linguistics were interested in the teaching of English as a second language, one can discern an emphasis on informative texts – texts used in science and technology, at the expense of literary texts, in the compilation of the corpora. However, Ahmad points out that there are instances where texts were published by a small group of publishers, based mainly in metropolitan areas, or where there is a gender imbalance between the authors of the texts. Ultimately, he argues, there is a degree of choice exercised by the compilers, but in this respect the behaviour of corpus linguists is not that different from that of much of the scientific community.

Words and expressions known as discourse particles do not affect the truth condition of an utterance and tend to modify the speech act rather than what is actually talked about. Functionally they express attitudes or emotions and contribute to the coherence of the utterance. In her contribution to this volume, 'Translating discourse particles: A case of complex translation', Karin Aijmer examines the use of English 'oh' and its translation into Swedish. Her findings show that there is no single lexical equivalent of 'oh' in Swedish. Instead, the many meanings that may be read into 'oh' must be translated in a number of different ways: if simply

rendered by the standard Swedish equivalent 'åh', the resulting translation is not a natural sounding construction in the target language.

While electronic texts are now available on the web in many languages of the world, in the case of some languages, their highly inflectional morphological structure compound the difficulties in compiling corpora. In 'The translator and Polish–English corpora', Tadeusz Piotrowski discusses the complexity of Polish morphology and the compilation of corpora including the problem of ambiguity created by the similarity of inflectional endings. From his account of the situation with respect to corpus compilation in Poland, it is also shown that corpora providing Polish language data are not as readily available as in some other languages discussed in this volume. Although data are not altogether missing in Poland, the information provided appears to be of greater interest to corpus linguists, while it is still only of limited use to practising translators.

In English, the so-called existential *there* has the status of a dummy subject, fulfilling a grammatical rather than semantic function. In contrast, Czech does not possess a lexical equivalent to the existential *there* in English; existence or occurrence is instead suggested by the intransitive character of the verb and the final position of the notional subject. For example, *Bylo ticho* ('was silence') corresponds to the English 'there was silence', while *V domě bylo ticho* ('In the house was silence') would also express 'There was silence in the house'. In 'The existential *There*-construction in Czech translation', Jiří Rambousek and Jana Chamonikolasová examine English existential sentences from the perspective of translation practice. Using a corpus of parallel texts, they analyse how Czech translators deal with English *there* constructions and the syntactic and semantic means they use to achieve functional equivalence in Czech.

The contribution by Špela Vintar, 'Corpora in translator training and practice: A Slovene perspective' starts by presenting an overview of resources available in Slovenia concluding that Slovene–English, in contrast to Slovene and other languages, is currently a language pair where it is possible to use parallel corpora effectively. This discussion is followed by an overview of mono- and multilingual corpora for Slovene; the role of corpora in translator training is described, as offered by the programme provided for the students in the Department of Translation at the University of Ljubljana. In view of the fact that the Slovene language community is one of the smallest in Europe, in the field of available bilingual language resources Slovene appears to be well catered for.

In the contribution from Hungary 'NP modification structures in parallel corpora', Tamás Váradi discusses the use of corpora in investigating structures involving NP modification in English and Hungarian. In spite of their radically different internal structure in each language, it is shown that maximally extended NPs lend themselves more easily to a

comparison than parts of speech such as, for example, adjectives. Among cases that show a departure from the *Adjective + Noun* pattern, the author focuses on Hungarian constructions equivalent to English *Noun + Prepositional Phrases*. The chapter concludes with a discussion of the criteria that may be used to distinguish genuine cases of explication from constructions that require expansion in order to meet the requirements of grammar.

In 'A study of the mandative subjunctive in French and its translations in English: A corpus-based contrastive analysis', Noëlle Serpollet investigates an aspect of grammar that has previously not been investigated bilingually. Indeed, the French–English language pair has so far received only limited contrastive attention using electronic corpora. Working first from French into English, the author analyses instances of the subjunctive in the Press category of the corpus and their translation into English. She then turns her attention to mandative constructions in English and examines how they are translated into French in the Learned Prose category. In conclusion, she compares her results with the findings obtained from an analysis of equivalent extracts from the 1-million-word corpus of British English LOB (1961) and its later counterpart, FLOB (1991).

The study by Diana Santos, 'Perfect mismatches: 'Result' in English and Portuguese', compares one aspect of the English verbal system, namely the meaning of 'result' closely associated with the present perfect, with the situation in Portuguese. The comparison is made bidirectionally using the COMPARA corpus, consisting of parallel literary texts (English–Portuguese and Portuguese–English). The structure of the corpus allows comparisons to be made not only between source and target texts, but also between translations and original texts. Santos concludes that while little attention is paid to result in Portuguese, compared to English, other distinctions and patterns are apparent. Her more general point is that a parallel corpus facilitates contrastive studies of a semantic kind, which help to identify complex differences even between related languages.

The final chapter, contributed by Anabel Borja, aims to provide an overview of the corpora available in Spain that might prove useful to translators and translation researchers working with Spanish as a source or target language. To this end, the author presents corpora containing texts in Spanish, including all the corpus resources identified as useful tools. In conclusion, a description is given of the CDJ-GITRAD corpus, a Multilingual Corpus of Legal Documents and the GENTT project, in which a multilingual encyclopaedia of specialised texts for translators is currently being compiled, specifically intended for medical, technical and legal translators working with Spanish.

Chapter 1

The Linguist and the Translator

GUNILLA ANDERMAN and MARGARET ROGERS

Introduction

In June 1956, J.R. Firth, holder of the first Chair of General Linguistics at the University of London, read a paper with the title 'Linguistics and Translation' to an audience at Birkbeck College, University of London. Firth concluded (1968a: 95):

The spread of world languages such as English, but not forgetting Russian, Chinese and Arabic, multiplies the need for translation from and into all these languages mutually and also into dozens of other languages which serve what has become more and more a common world civilization.

This observation presciently anticipated the need for translation that was to arise from the development of the European Economic Community (EEC), established in 1957, shortly after Firth's lecture, with six members and four languages, to the European Union of 2007 with its 27 member states and 23 official languages, soon to be more. It also pointed to the rapid spread of English as a global language, although Spanish and Hindi are now joining Chinese and English to make up the top four most frequently spoken languages in the world.¹

Linguistics and Translation: Early Pioneers

As a pioneer of the new discipline of linguistics in the UK, Firth's insight into the nature of language not only led him to predict an increased need for translation, it also made him an early advocate of the study of meaning in linguistics. At a time when American structuralist linguists were attempting to exclude meaning from linguistic analysis along with all psychological, or as Bloomfield called it 'mentalistic references', Firth clearly realised the importance of the task of incorporating linguistic meaning into the science of language. And as his definition of meaning as 'function in context' suggests, he was well aware of the importance of running text of the kind that computers are now able to process. In looking at words in their context, he was not, however, the first linguist to understand that – in isolation – separate lexical items are less likely to reveal to us their actual meaning.

Context provides an important link with earlier developments in foreign-language teaching, which in some ways foreshadowed what could

be termed the 'communicative turn' in language teaching in Western Europe in the latter part of the 20th century. Over half a century before Firth, Henry Sweet, a member of the Reform Movement, an ardent opponent of the exclusive concern with Latin and Greek among linguists and an advocate of the study of English as spoken, pointed out that detached sentences should not be substituted for connected texts as is often the case in the use of teaching methods that focus on grammar rather than on the text itself: 'it is only in connected texts that that the language itself can be given with each word in a natural and adequate context' (Sweet, 1899/1964: 163).

Awareness of the importance of not viewing words and constructions in isolation is also found in the work of Otto Jespersen, another member of the Reform Movement. In *A Modern English Grammar on Historical Principles*, Jespersen sets out to demonstrate the facts of English usage during different, historical periods. Supporting his discussion throughout are examples culled from the English canon and other sources. As his corpus, Jespersen used the texts in English available to him. His painstaking extraction of illustrative examples, in an endeavour 'to place grammatical phenomena in a true light' (Jespersen, 1961: VI), was at the time a gargantuan task, now routinely achieved in machine-readable corpus studies through, for instance, automatic grammatical 'tagging' of words.

While the importance of spoken language was vindicated by the establishment of the first Department of Phonetics at University College London in 1912 under the headship of Daniel Jones, who had studied under Henry Sweet, the notion of context was further developed by another London University scholar, the social anthropologist Bronislaw Malinowski. Following his first field study to record the life and work of the Trobriand islanders of New Guinea in the South-west Pacific between 1915 and 1918, Malinowski clearly saw the need for linguistics in developing a school of social anthropology in London, in particular in relation to the establishment of reliable ethnographic texts. For Malinowski, the notion of translation into English was crucial in his anthropological studies and was extended to include the definition of a term by ethnographic analysis, that is, by placing it within its context of situation and its context of culture, 'putting it within the set of kindred and cognate expressions, by contrasting it with its opposites, by grammatical analysis and above all by a number of well chosen examples [...] the only correct way of defining the linguistic and cultural character of a word' (1935, II,16, discussed in Firth, 1968b: 151).

Malinowski's method, according to Firth, included a number of different stages of translation. First, he discusses what he refers to as an interlinear word-for-word translation, sometimes described as a 'literal' or 'verbal translation', 'each expression and formative affix being rendered by its English equivalent' (Firth, 1968b: 149). As a second step this was followed by a free translation in what Firth describes as 'running

English'. Thirdly, the interlinear and free translations were collated, leading to the fourth stage, namely, the compilation of a detailed commentary, or 'the contextual specification of meaning', in which the free translation was related to the verbal translation including a discussion of 'equivalents' (Firth, 1968b: 149). The notion of 'context' that Firth embraces is that of Malinowski's 'context of situation' in its widest sense: 'It is clear that one cannot deal with any form of language and its use without assuming institutions and customs' (Firth, 1968b: 156). Also anticipating the ambivalence often expressed by 21st-century translators towards dictionaries, Firth (1968b: 156) expresses his reservations about the traditional use of dictionaries, citing Malinowski: 'I should agree that "the figment of a dictionary is as dangerous theoretically as it is useful practically" and, further, that the form in which most dictionaries are cast, whether unilingual or bilingual, is approaching obsolescence [...]'. In the current age of the fast-moving knowledge society, contextual solutions to terminological and phraseological problems are crucial to today's translators, who frequently turn to on-line documentation or even customised electronic corpora as an alternative to traditional dictionaries as well as electronic term bases or term banks, which rarely fully contextualise meaning and use.

Malinowski's influence is still discernible today in the computer-based processing of texts. His concept of a 'coefficient of weirdness', referring to strange language which becomes less strange in its context of use, and linking the user of language and the things s/he is trying to influence or connect with, has been adopted and adapted to shape statistical procedures for identifying specialist terms semi-automatically, originally for translation purposes. The basis of this is not any magical properties which specialist terms may exhibit, but rather their distributional characteristics compared with the distribution of content words in general-language texts, which are lexically less dense (cf. Ahmad & Rogers, 2001).

As one of the first 20th-century linguists to show a concern with the importance of translation, in his 1956 paper Firth (1968a: 86) recognises four different types of translation. The first type he calls 'creative translation', intended primarily as literature in the language into which it is rendered by the translator. The second type of translation to which Firth refers is 'official translation', the kind of language transfer used in documents and treaties in so-called 'controlled' or 'restricted' languages, and most closely related to what today we would call specialist translation, in which terminological studies of special domains play an important part.

The third type of translation selected by Firth for special attention is translation as used by linguists engaged in the description of a particular language, and the fourth, to which we return below, is 'mechanical translation'. As an example of the third type, we can cite Firth's description of an Indian novelist writing in English making repeated references to

Urdu and its pronominal system and terms of personal address, for which there are no equivalents in English. In this case, a contrastive analysis involving the principles of translation may help to illustrate the lack of equivalence between the different pronominal systems in the two languages. Firth also discusses the problem of carrying grammatical structures across what he calls the 'bridge' of translation, reflecting the contrastive method employed by the Prague linguists who arrived at important insights into the differences in information structure between European languages. This important text-based characteristic can, in turn, be related to different formal characteristics of the languages in contrast.² While the non-finite construction *your having done that will spoil your chances* is prevalent in English, in translation into other European languages, Firth (1968a: 92) points out that it is likely to require a separate clause with a verb in the finite form. In cases of this type, that is, of structural difference, he feels that linguistics, in providing this information, would be able to make a contribution to translation. In fact, one of the earliest English-language attempts to map out a systematic approach to the study of translation – Catford's (1965: 1) study – explicitly acknowledges his debt to Halliday, and in turn, Firth: 'The general linguistic theory made use of in this book is essentially that developed at the University of Edinburgh, in particular by M.A.K. Halliday and influenced to a large extent by the work of the late J.R. Firth'. While Catford has been in later years much criticised for what has generally been described as an approach which reduces translation to a linguistic decoding/encoding exercise, interestingly, he espouses a contextual view of language 'as related to the human social situation in which it operates' (Catford, 1965: 1), echoing Firth's concern with 'context of situation' and the functional orientation of Halliday's model of grammar. His notion of 'shifts' in particular can be usefully applied to the description of translation solutions from a formal point of view, a perspective which may still be pedagogically useful and still has relevance for certain aspects of professional translation.

For example, we could still imagine at least two further contexts of application for Firth's observation on clause correspondence. Firstly, it would be a construction to avoid in the drafting of documents for translation into many other languages, such as in the European Union, although from a stylistic point of view it is the absence of variation that constitutes one of the reasons for the blandness of such international documents, contributing to what has come to be known as 'Euro-English' (cf. Wagner, 2005). Secondly, for similar reasons of translatability, the type of English-specific clause contraction described above would be a good candidate for elimination in the pre-editing stage of machine translation, which brings us to Firth's fourth and last type of translation, namely 'mechanical translation'. As an example of this type, he cites the work pioneered by Dr Andrew Booth who, by the end of 1952, had produced an electronic stored programme computer in full

operation at the Birkbeck College Computation Library, University of London. In this context, Firth also makes tentative reference to the ease with which set phrases and clichés may be handled by machine translation. The path to the present-day use of computers in translation – particularly in Computer-Assisted Translation with its stored ‘translation memory’, that is, a database of pre-translated phrases and even sentences – is not difficult to detect, as repetitive formulations lend themselves most easily to this kind of treatment. While the increased power of modern-day computers speeds processing and facilitates the storage of large amounts of data in memory, the use of computer programs to ‘understand’ text through rule-based systems is still very limited (cf. Quah, 2006 for a summary); this goes some way to explaining why automatic translation systems, for instance, are highly constrained in their use, in relation either to subject field and genre, or to purposes for which the output is fit (for example, information only). Statistically or lexically based systems may prove more fruitful, as foreshadowed in the early work on corpora by Sinclair and Halliday suggesting that ‘grammar’ is highly localised and lexically based (cf. ‘lexicogrammar’).

Throughout the discussion in ‘Linguistics and Translation’, the thinking that was to inspire Michael Halliday and John Sinclair is not difficult to detect. In words that were to be echoed by Halliday in discussions of his approach to a theory of grammar, Firth (1968a: 90, *emphasis added*) writes: ‘the whole of our linguistic behaviour is best understood if it is seen as a network of relations between people, things and events, showing *structures* and *systems* just as we notice in all our experience’. Equally discernible in Firth’s discussion is the semantic notion of ‘collocability’ (as discussed in Sinclair, 1966: 417).

Halliday and Sinclair: The ‘Neo-Firthians’

The debt of Halliday and Sinclair, sometimes called ‘neo-Firthians’, is clearly expressed in the edited collection *In Memory of J. R. Firth* (Bazell *et al.*, 1966). Both acknowledge the importance of his notion of context of situation as inherited from Malinowski, as well as his concept of collocation, first used in the essay ‘Modes of Meaning’ (cf. Firth, 1968c: 204). His description of the term, often quoted by Halliday, is well known: ‘You shall know a word by the company it keeps! One of the meanings of “ass” is its habitual collocation with such other words as “you silly –”, “he is a silly –”, “don’t be such an –”’ (Firth, 1968c: 179). As Halliday (1966: 152) observes: ‘lexis seems to require the recognition merely of linear co-occurrence together with some measure of significant proximity, either a scale or at least a cut-off point. It is this syntagmatic relation which is referred to as “collocation”’. But in addition to collocational ‘span’ (the environment of a lexical item), a lexical description also appears to require

categories such as 'simple' and 'compound', as well as possibly 'phrasal' lexical items. Looking to the future and bearing in mind the potential of computers to formalise intuition about the behaviour of lexical items, Halliday (1966: 160, *emphasis in original*) further points out that '[a] thesaurus of English based on formal criteria, giving collocationally-defined lexical sets with citations to indicate the defining environments, would be a valuable complement to Roget's brilliant work of intuitive semantic classification in which lexical items are arranged "according to the *ideas* which they express"'. In the absence of such a work, even 'a table of the most frequent collocates of specific items [...] would be of considerable value for those applications of linguistics', Halliday (1966: 160) suggests, 'in which the interest lies not only in what the native speaker knows about his language but also in what he does with it'. These studies, he continues presciently, may include investigations of register and style, of child language, the language of aphasics and other target groups which, as we now know, became established fields of linguistic exploration during the latter part of the 20th century (Halliday, 1966: 160).

Alongside Michael Halliday, John Sinclair was one of the first scholars in the UK to explore the potential of computer text-processing for language description, focusing on lexical properties and, in particular, refining the notion of collocation to take account of how an item predicts the occurrence of others and is predicted by others (Sinclair, 1966). In those early days Sinclair predicted that it might not be possible for many years to establish how words relate to each other to create meaning in the context of the 'total frequency of the two items' (Sinclair, 1966: 428). In conclusion he pronounces the practical problems to be immense although well compensated for by the prospect of opening up new ways of describing language through the theory of lexis.

At the time that Halliday and Sinclair were developing their theoretical framework for the study of lexis, in their own different ways incorporating semantics as derived from Firth's 'function in context', a number of linguists were focussing their attention on grammar. The 1960s had seen the appearance of Chomsky's *Aspects of the Theory of Syntax* (1965), and the subject of the nature of the rules involved in the generation of grammatical constructions was increasingly attracting interest among linguists. Some 40 years later, Sinclair (2007) was to claim that attempts to develop rule-based descriptions of any natural language had failed. According to Sinclair's terse observation on computational linguistics as carried out in what became known as the natural language processing (NLP) community, it was both 'headless' (that is, without a theory, as the initially promising Chomskyan formalisms proved difficult to operationalise) and 'legless' (that is, without data in the form of naturally occurring language) (Sinclair, 2007: 24). Sinclair's judgement is based on a simple evaluation measure: how well does the grammar (actually, partial grammar) account

for the behaviour of language as observed in 'open text', even relatively highly constrained text such as that found in so-called sublanguages or domain-specific varieties (as opposed to constructed examples)? For Sinclair (2007: 36), a portfolio of what he calls 'local grammars' – early work was on the grammar of definitions – rather than a general grammar of any particular language seems 'closer to the way language is used' and the most realistic prospect for the foreseeable future of the machine 'understanding' text.

Sinclair's method is corpus-based, teasing out ever more complex patterns reflecting meaning–form relations, usually starting with a particular token or word form. The seeds of this approach are already discernible in Sinclair's 1966 paper, where he is critical of the notion of grammar as a set of systems – for example, active versus passive, a binary choice – which are treated independently. Instead, he turns to lexis 'which describes the tendencies of items to collocate with each other'; such tendencies, he goes on, 'cannot be got by grammatical analysis, since [they] cannot be expressed in terms of small sets of choices' (Sinclair, 1966: 411). If we project from this monolingual position to the process of translation, from a linguistic point of view it could be argued that successful translation consists in finding authentic or 'attested' examples (cf. Firth's 'attested language') of corresponding *tendencies* in the target language, problematising the notion of 'equivalence' more than ever. Under this view, the value of dictionaries, term banks and term bases organised around headwords or key concepts rather than larger 'chunks' of language becomes questionable.

In the context of the present, the widespread use of Translation Memory (TM), which could be said to 'learn' (in the sense of matching patterns with the support of a human user – it's not that smart yet) to translate chunks of text above the word level, can be seen as a vindication of this approach within a narrow range of texts, in so far as TM is used to translate certain domain-specific genres of text of some length. It is also interesting to note in this context that corpus-based machine translation systems, including the statistical approach based on probabilities of occurrence in aligned parallel texts (the original exemplar being Canadian Hansard), are now serious contenders to rule-based systems (cf. Quah, 2006: 76–84). What was for Firth called 'collocation' has been extended as a notion by Sinclair and Halliday in the computer age, in a way which is resonant in some respects of the alternative term 'automation' proposed by the Prague linguists in the 1930s (cf. McEnery & Wilson, 2001: 24).

The Emergence of Corpora for Linguistic Research

In the early days, the question also being asked was the extent to which corpus-based studies might be used in order to shed light on phenomena of grammar. Prominent among the grammatical constructions discussed was

the passive in English and the syntactic rules required in order to generate it. In 1966, Jan Svartvik published his PhD thesis as *Voice in the English Verb*, based on material taken from the files of the *Survey of English Usage* representing co-existing varieties of spoken and written educated English. Svartvik's research, carried out under the auspices of the director of the project, Randolph Quirk at University College, London, set out to investigate 'to what extent "corpus-passives" differ from "rule-generated passives" and from actives' (Svartvik, 1966: 6).

The use of electronic corpora for linguistic research had its roots, according to McEnery and Wilson (2001: 21–22), in the mid-1950s in the pioneering and ambitious work by Alphonse Juiland on French, Spanish, Romanian and Chinese. More well known work was carried out in English from the early 1960s at Brown University in the USA (Francis & Kucera, 1982), shadowed by the Lancaster-Oslo-Bergen (LOB) corpus of British English. But it was only in the 1970s that Randolph Quirk's 1960s work at University College London on the *Survey of English Usage* was computerised by Jan Svartvik (McEnery & Wilson, 2001: 22). The 'corpus' in Corpus Linguistics is nowadays always machine-readable. It is worth reflecting for a moment, however, that the digitisation of text was not always so self-evident for the computing community who, of necessity, were called upon to support and help drive forward the introduction of computers into linguistic analysis and its applications. In the 1960s and 1970s, and to an extent the 1980s, the main application of computer power was number crunching, using physically large mainframe computers which operated in batch mode with data input mechanisms which relied on punch cards or tape. The use of 'non-standard character sets' (even in languages with Latin characters such as French, German and Spanish) was problematic, as was a perception among some computing colleagues that upper and lower case differences were of no import. So while the difficulties of monolingual text storage and processing were significant, extending this to multilingual material represented an even bigger culture change.

Hurdles were gradually overcome, as interest boomed in the 1980s, particularly in English (cf. for example Johansson 1982), along with technological developments, and by the mid-1990s, Corpus Linguistics had become 'mainstream' (Thomas & Short, 1996: ix) (for further information on the development of corpora for different purposes and of various kinds, cf. Aarts & Meijs, 1990; Aijmer & Altenberg, 1991; Butler, 1992; Garside *et al.*, 1987; Kennedy, 1998).

In addition to being used for empirical research into the patterns of actual language use, facilitated by the possibility to store and process ever larger amounts of digital text and to approach the nevertheless ill defined notion of 'representativeness' (cf. Ahmad, this volume), corpora were also used for the first time to support the compilation of grammars (for example Quirk, Greenbaum, Leech and Svartvik's *A Comprehensive Grammar of the*

English Language (1985)) and dictionaries (for example, Sinclair's 1987 *Collins Cobuild English Language Dictionary*). Around this time in the mid-1980s, suggestions were also emerging for the use of corpora in language learning (for example, Ahmad *et al.*, 1985: 126–127). The general theme characterising all these developments was the study of language through observation rather than introspection or experimentation, promising richer and more sensitive descriptions of syntactic and lexical patterns. But of translation and electronic corpora there was little yet to be seen: the focus of discussions concerning 'computers and translation' was still largely machine translation (for example Lewis, 1992), most corpora remained monolingual and the development of text-alignment software was in the domain of research laboratories (for an overview of the development of text-alignment, cf. Véronis, 2000).

Corpus-based Studies and Translation

However, signs of interest did start to emerge in the 1980s, particularly in the Scandinavian countries, as Translation Studies started to develop into a discipline in its own right. Svartvik has since suggested that the obvious difficulties posed by introspection for non-native speakers of English engaged in English studies may have contributed to the popularity of the then rather unfashionable corpus approach in the early days, that is, in the 1960s (Svartvik, 2005). Twenty years later, linguists interested in corpus-based studies began to turn to the potential role of corpora in the study of translated texts, initially literary texts such as novels. At the Department of Swedish at Gothenburg University a selection of novels translated into Swedish from English together with a corresponding amount of text from novels originally written in Swedish, all published between 1976 and 1977, has long provided the basis for studies into the features characterising texts in translation (cf. Gellerstam, 1986 for a more detailed discussion). The increased volume of text which could be studied using machine-based search and retrieval methods gave rise to insights that would have been difficult to 'observe' using traditional paper-based methods. Gellerstam was able to show that the distribution of words in translated texts differs from that in original texts, casting new light on the hoary chestnut of 'translationese'. In 1989 further corpus-based information about aspects of translation between English and Swedish was made available through the publication of Hans Lindquist's study of English adverbials in translation (Lindquist, 1989), presented as a PhD dissertation at the University of Lund under the supervision of Jan Svartvik. Lindquist's text corpus of five British and five American novels was used to provide a 'corpus' (a lexicographical understanding of the term) of 2000 adverbials, at first stored on cards, and then partly

computerised (Lindquist, 1989: 31), reflecting well the ongoing developments at that time in the move from paper to computer.

Originally firmly rooted in the 'paper tradition' with its provenance in the late 1970s is Gideon Toury's descriptive approach to works of literary translation in which so-called 'norms', which Baker (2001: 163) interprets as 'regularities of translation behaviour within a specific sociocultural situation', are studied. What Toury's approach shares with the neo-Firthian legacy is not the close analysis of linguistic behaviour in text, but the emphasis on the description of 'attested' texts – here, literary translations – rather than an idealised, in his view prescriptive, notion. While it has been suggested that Baker's contribution to the use of corpora in Translation Studies lies in the application of Corpus Linguistics to Descriptive Translation Studies (cf. for instance, Kenny, 2001a: 50), the resulting work on so-called 'universals of translation'³ has a strong linguistic flavour, for example, a tendency to explicitation, disambiguation, simplification, grammatical conventionality and 'a tendency to over-represent typical features of the target language' (Mauranen, this volume), echoing earlier work such as that of Gellerstam.

As we have seen, Corpus Linguistics was originally centred on monolingual corpora (although incorporating both written and spoken texts). Baker rightly points out (1995: 225–226) that additional criteria of corpus design beyond those of, for instance, general versus restricted domain, synchronic versus diachronic, genre, geographical variant, were needed for translation research, including range of translators and respective genre in each language. While these criteria are potentially important, the former is hard to fulfil in practice as even the authorship of original texts, let alone translations, is often hard to trace for many 'pragmatic texts' (*Gebrauchstexte*); such considerations are important when research into language for special purposes (LSP) translation (*Fachübersetzen*) is becoming more prominent as a research area in Translation Studies, including corpus-based studies.

It is also worth noting that the typologies of corpora for translation and translation-related research vary (compare, for instance, Baker, 1995, Bernadini *et al.*, 2003 and Teubert, 1996 for variations on the main themes in the context of translator education). The 'value' of translated texts in what is usually called 'parallel corpora' – source texts and their translations – may also vary depending on the purpose of the study. In Descriptive Translation Studies the point is not to evaluate but to describe and explain translation behaviour, while in bilingual lexicography or terminography (cf. Ahmad & Rogers, 2001; Teubert, 1996) the aim is to establish lexical equivalents that will function in new texts in the target language alongside authentic texts in that language. For such purposes, translated texts may provide skewed data for the very reason that

interests descriptivists: that translations differ, sometimes in subtle ways, from comparable original texts.

To conclude with the new century, publications on corpora and translation started to appear with regularity, some with a specific research brief (for example Kenny, 2001b using a German–English parallel corpus of literary texts), some with a pedagogical bent (for example Zanettin *et al.*, 2003, covering both literary and specialist texts for a variety of training purposes), some with a broad sweep (for example Olohan, 2004) and others with a linguistic focus (for example Aijmer & Hasselgård, 2004).

Over 30 years after the publication of his paper in honour of Firth, in considering the risky and rather unfashionable question of what might constitute a theory of good translation, Halliday (2001: 13–14) still calls on Firth's broad sense of meaning as operating 'at all linguistic strata', including both expression and content, in order to 'be able to explain why [the text] means what it is understood to mean'. He sees this as a step towards understanding equivalence, which he predictably problematises as a systemic issue in which 'equivalence at different strata carries differential values', suggesting that 'contextual equivalence [is valued] perhaps most highly of all' (Halliday, 2001: 15). This view is nuanced in the context of what Halliday (2001: 15) calls 'the task', which may lead to differential assignments of value in parts of the system.

The early semantic seeds produced by the husbandry of Malinowski and sown by Firth were nurtured by linguists such as Halliday and Sinclair and can be seen to have taken firm root in the work of what we now know as 'Corpus Linguistics'. The relationship of Corpus Linguistics to translation is one that is still developing between system and text; the current volume is a contribution to this development.

Notes

1. David Graddol: personal communication, June 2006.
2. Cf. Firbas (1999) in relation to Firth's 'creative translation', and Rogers (2006) with respect to 'official translation'.
3. Cf. Mauranen (this volume) for an overview, and Malmkjær (this volume) for a critical review of norms and universals.

References

- Aarts, J. and Meijs, W. (eds) (1990) *Theory and Practice in Corpus Linguistics*. Amsterdam: Rodopi.
- Ahmad, K. (this volume) Being in text and text in being: Notes on representative texts. *Incorporating Corpora*. Clevedon: Multilingual Matters.
- Ahmad, K., Corbett, G., Rogers, M. and Sussex, R. (1985) *Computers, Language Learning and Language Teaching*. Cambridge: CUP.
- Ahmad, K. and Rogers, M. (2001) Corpus linguistics and terminology extraction. In S.-E. Wright and G. Budin (eds) *Handbook of Terminology Management* (Vol. 2, pp. 725–760). Amsterdam/Philadelphia: John Benjamins.

- Aijmer, K. and Altenberg, B. (eds) (1991) *English Corpus Linguistics*. London and New York: Longman.
- Aijmer, K. and Hasselgård, H. (eds) (2004) *Translation and Corpora. Selected papers from the Göteborg-Oslo Symposium 18–19 October 2003*. Göteborg: Acta Universitatis Gothenburgensis.
- Baker, M. (1995) Corpora in translation studies: An overview and some suggestions for future research. *Target* 7 (22), 223–243.
- Baker, M. (2001) Norms. In M. Baker (ed.) *Routledge Encyclopedia of Translation Studies* (pp. 163–165). London: Routledge (assisted by K. Malmkjær).
- Bazell, C.E., Catford, J.C., Halliday, M.A.K. and Robins, R.H. (eds) (1966) *In Memory of J.R. Firth*. London and Harlow: Longmans, Green and Co.
- Bernadini, S., Stewart, D. and Zanettin, F. (2003) Corpora in translator education. An Introduction. In F. Zanettin, S. Bernadini and D. Stewart (eds) *Corpora in Translator Education* (pp. 1–13). Manchester: St. Jerome.
- Butler, C.S. (ed.) (1992) *Computers and Written Texts*. Oxford and Cambridge, MA: Blackwell.
- Catford, J.C. (1965) *A Linguistic Theory of Translation. An Essay in Applied Linguistics*. London: Oxford University Press.
- Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Collins Cobuild English Language Dictionary (1987) London and Glasgow: Collins.
- Firbas, J. (1999) Translating the introductory paragraph of Boris Pasternak's *Docot Zhivago*: A case study in Functional Sentence Perspective. In G. Anderman and M. Rogers (eds) *Word, Text, Translation. Liber Amicorum for Peter Newmark* (pp. 129–141). Clevedon: Multilingual Matters.
- Firth, J.R. (1968a) Linguistics and translation. In F.R. Palmer (ed.) *Selected Papers of J.R. Firth, 1952–59* (pp. 84–95). London and Harlow: Longmans, Green and Co.
- Firth, J.R. (1968b) Ethnographic analysis and language with reference to Malinowski's views. In F.R. Palmer (ed.) *Selected Papers of J.R. Firth, 1952–59* (pp. 137–167). London and Harlow: Longmans, Green and Co.
- Firth, J.R. (1968c) A synopsis of linguistic theory. In F.R. Palmer (ed.) *Selected Papers of J.R. Firth, 1952–59* (pp. 168–205). London and Harlow: Longmans, Green and Co.
- Francis, W.N. and Kučera, H. (1982) *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin (with the assistance of A. Mackie).
- Garside, R., Leech, G. and Sampson, G. (eds) (1987) *The Computational Analysis of English: A Corpus-based Approach*. London: Longman.
- Gellerstam, M. (1986) Translationese in Swedish novels translated from English. In L. Wollin and H. Lindquist (eds) *Translation Studies in Scandinavia. Proceedings of the Scandinavian Symposium on Translation Theory (SSOTT)* (pp. 88–95). Lund: Lund Studies in English 75.
- Halliday, M.A.K. (1966) Lexis as a linguistic level. In C.E. Bazell, J.C. Catford, M.A.K. Halliday and R.H. Robins (eds) *In Memory of J.R. Firth* (pp. 148–162). London and Harlow: Longmans, Green and Co.
- Halliday, M.A.K. (2001) Towards a theory of good translation. In E. Steiner and C. Yallop (eds) *Exploring Translation and Multilingual text Production: Beyond Content* (pp. 13–18). Berlin and New York: Mouton de Gruyter.
- Johansson, S. (ed.) (1982) *Computer Corpora in English Language Research*. Bergen: Norwegian Computing Centre for the Humanities.
- Jespersen, O. (1961) *A Modern English Grammar on Historical Principles* (Vols I–VII). London: Allen and Unwin; Copenhagen: Ejnar Munksgaard.