

Lexis in Contrast
Corpus-based approaches

Edited by

Bengt Altenberg

University of Lund

Sylviane Granger

Université Catholique de Louvain

John Benjamins Publishing Company
Amsterdam/Philadelphia

Studies in Corpus Linguistics

Studies in Corpus Linguistics aims to provide insights into the way a corpus can be used, the type of findings that can be obtained, the possible applications of these findings as well as the theoretical changes that corpus work can bring into linguistics and language engineering. The main concern of SCL is to present findings based on, or related to, the cumulative effect of naturally occurring language and on the interpretation of frequency and distributional data.

General Editor

Elena Tognini-Bonelli

Consulting Editor

Wolfgang Teubert

Advisory Board

Michael Barlow, *Rice University, Houston*

Robert de Beaugrande, *UAE*

Douglas Biber, *North Arizona University*

Chris Butler, *University of Wales, Swansea*

Wallace Chafe, *University of California*

Stig Johansson, *Oslo University*

M. A. K. Halliday, *University of Sydney*

Graeme Kennedy, *Victoria University of Wellington*

John Laffling, *Herriot Watt University, Edinburgh*

Geoffrey Leech, *University of Lancaster*

John Sinclair, *University of Birmingham*

Piet van Sterkenburg, *Institute for Dutch Lexicology, Leiden*

Michael Stubbs, *University of Trier*

Jan Svartvik, *University of Lund*

H-Z. Yang, *Jiao Tong University, Shanghai*

Antonio Zampolli, *University of Pisa*

Volume 7

Lexis in Contrast: Corpus-based approaches

Edited by Bengt Altenberg and Sylviane Granger

Lexis in Contrast

Corpus-based approaches

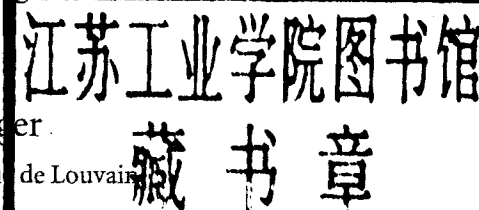
Edited by

Bengt Altenberg

University of Lund

Sylviane Granger

Université Catholique de Louvain



John Benjamins Publishing Company
Amsterdam/Philadelphia



™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences - Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Cover design: Françoise Berserik
Cover illustration from original painting Random Order
by Lorenzo Pezzatini, Florence, 1996.

Library of Congress Cataloging-in-Publication Data

Lexis in contrast : corpus-based approaches / edited by Bengt Altenberg, Sylviane Granger.

p. cm. (Studies in Corpus Linguistics, ISSN 1388-0373 ; v. 7)

Includes bibliographical references and index.

1. Lexicology--Data processing. 2. Contrastive linguistics--Data processing. 3. Lexicography--Data processing. 4. Translating and interpreting--Data processing. I. Altenberg, Bengt. II. Granger, Sylviane, 1951- III. Series.

P326.5.D38 LA495 2002

413'.028--dc21

2001037885

ISBN 90 272 2277 0 (Eur.) / 1 58811 090 7 (US) (Hb; alk. paper)

© 2002 - John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands

John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Preface	vii
List of contributors	ix
Introduction	
Recent trends in cross-linguistic lexical studies <i>Bengt Altenberg and Sylviane Granger</i>	3
Cross-Linguistic Equivalence	
Two types of translation equivalence <i>Raphael Salkie</i>	51
Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach <i>Elena Tognini Bonelli</i>	73
Causative constructions in English and Swedish: A corpus-based contrastive study <i>Bengt Altenberg</i>	97
Contrastive Lexical Semantics	
Polysemy and disambiguation cues across languages: The case of Swedish <i>få</i> and English <i>get</i> <i>Åke Viberg</i>	119
A cognitive approach to <i>Up/Down</i> metaphors in English and <i>Shang/Xia</i> metaphors in Chinese <i>Lan Chun</i>	151

From figures of speech to lexical units: An English-French contrastive approach to hypallage and metonymy <i>Michel Paillard</i>	175
Corpus-based Bilingual Lexicography	
The role of parallel corpora in translation and multilingual lexicography <i>Wolfgang Teubert</i>	189
Bilingual lexicography, overlapping polysemy, and corpus use <i>Victòria Alsina and Janet DeCesaris</i>	215
Computerised set expression dictionaries: Analysis and design <i>Sylviane Cardey and Peter Greenfield</i>	231
Making a workable glossary out of a specialised corpus: Term extraction and expert knowledge <i>Christine Chodkiewicz, Didier Bourigault and John Humbley</i>	249
Translation and Parallel Concordancing	
Translation alignment and lexical correspondences: A methodological reflection <i>Olivier Kraif</i>	271
The use of electronic corpora and lexical frequency data in solving translation problems <i>François Maniez</i>	291
Multiconcord: A computer tool for cross-linguistic research <i>Patrick Corness</i>	307
General index	327
Author index	333

Preface

Most of the articles in this volume represent a selection of papers presented at the 'Contrastive Linguistics and Translation Studies. Empirical Approaches' conference organised by Sylviane Granger at the Catholic University of Louvain in February 1999. All the contributions have been revised to fit the special theme of the volume. In addition, two contributions have been added to the original selection of papers: the introductory survey by Bengt Altenberg and Sylviane Granger and Wolfgang Teubert's article on the importance of translations in cross-linguistic lexical research.

The contributions reflect three striking tendencies that emerged during the conference. One is the rapidly growing interest in corpus-based approaches to the study of lexis, in particular the use of multilingual corpora, shared by researchers working in widely differing fields - contrastive linguistics, lexicology, lexicography, terminology, computational linguistics, machine translation and other branches of natural language processing.

The second tendency finds its expression in the wealth of methodological approaches represented at the conference, especially as regards the kinds of corpora used and the ways in which multilingual lexical information can be extracted from corpora and exploited for various purposes. This methodological diversity reflects to some extent the types of monolingual and multilingual corpora available at the time of the conference, but it is above all a healthy and promising sign of the vitality and desire for reorientation in a number of related fields where not only the object of research (lexis) but also the methodology (the use of corpora) are rapidly expanding and demanding increasing attention.

However, no matter what the purpose of the individual contributions may be, whether theoretical or practical, the driving force that unites them all is easily recognisable as the third — and perhaps most fundamental — tendency to have emerged from the conference: a common desire to give the cross-linguistic study of lexis a firm empirical foundation.

We have divided the articles into four main groups reflecting what we regard as some major concerns and aspects of the field: the exploration of

cross-linguistic equivalence, contrastive lexical semantics, corpus-based multilingual lexicography, and translation and parallel concordancing.

The conference brought together researchers from a wide range of countries and this is reflected in the diversity of the languages covered in the articles: English, Catalan, Chinese, Czech, Finnish, French, German, Italian, Lithuanian, Spanish and Swedish.

In preparing this volume we have benefited from the generous help of several people. Apart from the contributors themselves, we wish to thank an anonymous reviewer for many valuable comments and suggestions, Helen Swallow for her meticulous examination of the manuscript, and Kees Vaes and Elena Tognini-Bonelli for the confidence they have shown in entrusting us with the task of editing this volume.

Bengt Altenberg and Sylviane Granger
Lund and Louvain-la-Neuve, Autumn 2001

List of contributors

Victòria Alsina
Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona

Bengt Altenberg
Department of English, University of Lund

Didier Bourigault
Centre National de la Recherche Scientifique, Equipe de recherche en syntaxe et sémantique, Université Toulouse-le-Mirail

Sylviane Cardey
Centre de recherche en linguistique Lucien Tesnière, Université de Franche-Comté

Christine Chodkiewicz
Centre National de la Recherche Scientifique, Centre de terminologie et de néologie, Laboratoire de linguistique informatique, Université Paris 13

Janet DeCesaris
Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona

Peter Greenfield
Centre de recherche en linguistique Lucien Tesnière, Université de Franche-Comté

John Humbley
Centre de terminologie et de néologie, Laboratoire de linguistique informatique, Université Paris 13

Patrick Corness
School of International Studies and Law, Coventry University

Sylviane Granger
Centre for English Corpus Linguistics, Université Catholique de Louvain

Olivier Kraif

Laboratoire d'Ingénierie Linguistique et de Linguistique Appliquée, Université de Nice Sophia Antipolis

Lan Chun

Department of English, Beijing Foreign Studies University

François Maniez

Centre de Recherche en Terminologie et Traduction, Département des Langues Etrangères Appliquées, Université Lumière Lyon II

Michel Paillard

Département d'Études Anglophones, Université de Poitiers

Raphael Salkie

School of Languages, University of Brighton

Wolfgang Teubert

Department of English, University of Birmingham

Elena Tognini Bonelli

Università degli Studi di Lecce and The Tuscan Word Centre

Åke Viberg

Department of Linguistics, University of Lund

PART I

Introduction

Recent trends in cross-linguistic lexical studies

Bengt Altenberg and Sylviane Granger

1. Lexis and contrastive linguistics

1.1. Lexis: an expanding universe

The days are long gone when lexis was thought of as an unruly chaos, "a prison", to use Di Sciullo & Williams' (1987:3) words, "[which] contains only the lawless, and [where] the only thing the inmates have in common is their lawlessness". Following this period of neglect, during which lexis was most definitely the poor relation of grammar and syntax, there has been a radical restructuring of priorities, and the lexicon now features high on the agenda, in both theoretical and applied linguistics. As a result, there is a general trend towards lexically oriented approaches to language in which what was formerly regarded as syntactic phenomena has increasingly come to be viewed as projections of lexical properties. This development is noticeable in most branches of linguistics, formal as well as functional.¹ One influential strand of this development is the empiricist movement that is sometimes called 'British contextualism', most clearly represented by John Sinclair and his colleagues. Sinclair (1987a) attributes this dramatic turnabout to two concurring factors: Halliday's model of language and the advent of computers.

In 1966, in an article entitled 'Lexis as a Linguistic Level', Halliday called for recognition of a lexical level alongside the universally recognised grammatical level. From the start, however, he insisted that lexis was not to be viewed as totally separate from grammar: "If therefore one speaks of a lexical level, there is no question of asserting the 'independence' of such a level, whatever this might mean; what is implied is the internal consistency of the statements and

their referability to a stated model" (1966: 152). Alongside the grammatical and the lexical levels, there is also a lexico-grammatical level where lexical restrictions intersect with grammatical ones. The main argument offered by Halliday in support of a lexical level is the existence of collocations, i.e. combinatory restrictions which are neither grammatical nor semantic but which reflect "the habitual or customary places" of words (Firth 1957: 12). The acceptability of *strong tea* and *powerful car* and relative unacceptability of *powerful tea* and *strong car* demonstrate the existence of restrictions which depend on the syntagmatic relations into which words enter. Collocations are essentially based on probabilities, with words having a higher or lower likelihood of occurring together. But on the whole this probability is extremely low and, as a result, verification of Halliday's probabilistic approach relies on the existence of large corpora and computational techniques.²

Without the advent of computers the approach to lexis propounded by Halliday would never have had the tremendous impact it has already had and continues to have on the field of linguistics. Computers have made it possible to store ever larger collections of texts in electronic form and to analyse them using increasingly sophisticated, versatile and user-friendly software tools. But whereas grammar and semantics involve a high degree of abstraction, and are therefore relatively difficult to access using computer technology, lexis lends itself perfectly to the form-based research at which computers excel, whether those forms be letters, word spaces, punctuation marks or, indeed, words. Take frequency counts for example: an ideal field of enquiry in which to use computational techniques. For the first time ever, linguists have been able to rely on non-impressionistic large-scale frequency data. Although the reliability of frequency studies was questioned from a relatively early stage, this did not put an end to them but, instead, merely prompted corpus linguists to gather bigger and more tightly controlled corpora.

These two factors have contributed to bringing the study of words to the forefront of linguistic research, along with a change of name from vocabulary to lexis. But it is not only the name which has changed. It has become an altogether different phenomenon, in three ways in particular.

First and foremost, lexis and grammar are now seen as interdependent. This idea, first introduced by Halliday, was further developed by Sinclair, who criticised the traditional decoupling of lexis and grammar and claimed that it was "more fruitful to start by supposing that lexical and syntactic choices correlate, than that they vary independently of each other" (1991: 104). This interrelation of grammar and lexis is one of the key features in the new corpus-based

Longman Grammar of Spoken and Written English (Biber et al. 1999), which gives pride of place to lexico-grammatical associations — both grammatical associations of lexical words and lexical associations of grammatical structures.

Closely linked with this development is the fact that lexis has now been firmly placed on the syntagmatic axis. While paradigmatic relations for a long time dominated lexical studies, the pendulum now seems to have swung in the opposite direction so that it is now on the analysis of co-occurrence relations that attention is focused. This new emphasis on the company words keep, to use Firth's expression, has led to the discovery of a wide range of word combinations or multi-word units, which vary in fixedness and idiomaticity.

The third major change which has taken place in perceptions about lexis is that it is now recognised as displaying a much higher degree of stylistic differentiation than had previously been thought. In the case of English, the analysis of corpora has led to the discovery of a wide range of dialectal differences related to regional provenance (American English, Indian English), age (teenager English), sex (female lexis), time (Middle English lexis), social class, as well as diatypic differences in terms of field, mode and tenor (spoken lexis, ESP lexis, informal lexis).

Lexis has undergone a dramatic transformation and come out less autonomous, more open to other layers of language, notably grammar, composed of both single words and multi-word units and entering into a complex network of paradigmatic and syntagmatic relations.

1.2. The revival of Contrastive Linguistics

Like lexicology, contrastive linguistics now also occupies a dominant position in linguistics, but it has reached this position via a rather different route. Whereas in the case of lexis, its time had come, contrastive linguistics had already had its glory days back in the 1960s, before falling into disfavour, principally because of its association with structuralism. What we are now witnessing is thus more of a revival, and a dramatic one at that.

When Contrastive Analysis (CA) emerged as a scholarly discipline in the decades after World War II, it was regarded mainly as an applied branch of linguistics serving practical pedagogical purposes in foreign and second language teaching. In accordance with the linguistic climate of the time (structuralism, early generative grammar), phonology and grammar held centre stage, while lexis played a subordinate role.³ The high hopes it had raised — that similarities and differences between languages could predict, or at least explain, prob-

lems in foreign and second language learning and make language teaching more efficient — were largely thwarted. For a time CA became a suspect field of study, especially in the United States (on the history and deficiencies of CA, see Ringbom 1994, Sajavaara 1996, Chesterman 1998). However, in Europe CA continued to thrive and large contrastive projects were established in the 1970s, comparing English and other European languages. There, in particular, the view persisted that CA still had much to offer, not only to language pedagogy, but also to translation theory, the description of particular languages, language typology and the study of language universals (on various early approaches, see Di Pietro 1971, James 1980 and Krzeszowski 1990).

Now CA — or contrastive linguistics (CL), as it is increasingly called — is again an active and expanding field which generates lively theoretical and methodological discussion. A large number of research projects, conferences and journals are devoted to cross-linguistic work of various kinds, especially in Europe. And lexis, moreover, is very much the focus of attention.

Broadly speaking, there are three main reasons for this, although they are closely interrelated and difficult to separate. Internationalisation and the gradual integration of Europe have created an increasing demand for multilingual and cross-cultural competence, for translation, interpreting and foreign language teaching. The importance of accurate and efficient communication across language boundaries has become a concern not only of linguists and teachers but of governments, commercial institutions and international organisations. As a result, there has been a rapidly increasing awareness of the need for large-scale cross-linguistic research.

At the same time, there have been important developments within linguistics. A growing interest in real-life communication has shifted the focus away from the earlier preoccupation with abstract language (sub)systems and the reliance on the native speaker's intuition as the main source of linguistic knowledge in the direction of natural discourse and empirical data as evidence for linguistic observations. The earlier tendency, fostered by structuralism and early generative grammar, to regard language as consisting of autonomous systems (with phonology and grammar in the centre) has given way to a more complex and dynamic view of language which allows greater interaction between the systems and fuzzier boundaries between them. As mentioned, lexis has acquired a more central position in several respects: the concept of the lexical item has expanded and the interdependence between lexical choice and contextual factors has led to a growing tendency to enrich the lexicon with information of a grammatical, semantic and pragmatic nature (see e.g. Atkins

et al. 1994). These tendencies have had a profound influence on lexical CL.

A third important reason for the revival of contrastive studies is the computer revolution and the possibility of analysing natural language on the basis of large text corpora. This has opened up new possibilities of research on the basis of bilingual or multilingual corpora and experiments in natural language processing, e.g. in the fields of machine translation, information retrieval and computational lexicography. Corpora provide empirical data for linguistic theories and practical applications or serve as testing grounds for linguistic and computational models. The information gained from corpora is both richer and more reliable than that derived from introspection.

These new developments have brought about a revival of interest in CL. CL now permeates a number of fields inside and outside linguistics and its impact has been especially strong in areas concerned with natural language processing, such as machine translation and computational lexicography. Indeed, the analysis of individual languages has even been described as forming a part of CL (Weigand 1998b: vii). The new tendencies have also given rise to increased cooperation between experts from a number of fields: linguistics, lexicography, translation, computer science, psychology and cognitive science. Even if the problems of describing and relating many languages are as formidable as ever, great advances have been made in identifying and addressing the issues and there is new hope and great vitality in the field.

2. Multilingual corpora

2.1 Types of corpora

As we have seen, one factor that has influenced the contrastive study of lexis more than any other is the computer revolution and the development of multilingual corpora. Several types of multilingual corpora need to be distinguished. Unfortunately, the terminology used to describe the different types is inconsistent and confusing (for some different typologies, see Baker 1995, 1999 and Hartmann 1996). We shall here use the typology and terms set out in Figure 1 (cf. Johansson 1998: 4–7).

Depending on the number of languages involved, one distinction that can be made is that between bilingual and multilingual corpora. To simplify matters, we shall use 'multilingual' as a general inclusive term and only be more specific when necessary. A more important distinction is that between comparable corpora and translation corpora. Comparable corpora consist of original

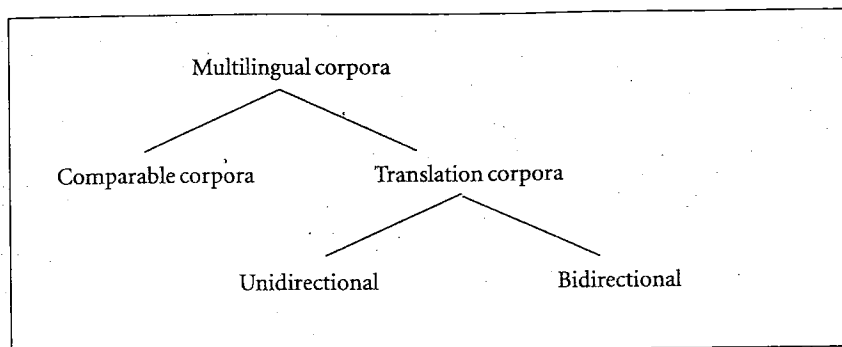


Figure 1. Types of multilingual corpora

texts in each language, matched as far as possible in terms of text type, subject matter and communicative function. Corpora of this kind can either be restricted to some specific domain (e.g. genetic engineering, contract law, job interviews) or be large 'balanced' corpora representing a wide range of text types. Translation corpora consist of original texts in one language and their translations into one or several other languages. If the translations go in one direction only (from language A to language B) they are unidirectional; if they go in both directions (from language A to language B and from language B to language A) they are bidirectional. The term 'parallel corpus' is sometimes used as an umbrella term for both comparable and translation corpora, but it seems more appropriate for aligned translation corpora, where a unit (paragraph, sentence or phrase) in the original text is linked to the corresponding unit in the translation (see Section 2.2).⁴

Each of these types has its advantages and disadvantages (see Aijmer et al. 1996, Teubert 1996, Johansson 1998). Comparable corpora represent natural language use within the genres they contain and are unaffected by various translation effects (see below). Domain-specific corpora are especially useful for terminological studies. If 'comparability' is taken in a broad sense, very large 'balanced' corpora representing a wide range of genres and text types can serve as comparable corpora. Since corpus size and large quantities of data are important factors in contrastive lexical research, they are especially useful in collocation studies and as control corpora for results derived from translation corpora.

The problem with comparable corpora is, somewhat paradoxically, the comparability of the data. It is difficult, and in some cases impossible, to know what to compare, i.e. to relate expressions with comparable meaning and function in the languages compared. Moreover, unlike translation corpora, compa-

rable corpora cannot reveal sets of cross-linguistic equivalents in cases where one or both languages provide a choice of alternatives (unless these have been identified in advance). Another problem with comparable corpora is their functional and stylistic comparability. If the source texts of the corpora are not selected according to the same principles, any comparison is bound to be uncertain. For these reasons, the use of comparable corpora is either limited to restricted domains or to very large balanced corpora where such factors as topic, register, and communicative function can be controlled.

Translation corpora have the advantage of keeping meaning and function constant across the compared languages.⁵ They also make it possible to discover cross-linguistic variants, i.e. alternative ways of rendering a particular meaning or function in the target language. By reversing this process, i.e. starting from the range of variants discovered in language B and observing how these are rendered in language A, it is possible to discover paradigms of cross-linguistic correspondences (see Section 5.2).

The disadvantage of using translation corpora is that translations tend to retain traces of the source language ('translationese' — see e.g. Gellerstam 1986, 1996) or display other general characteristics of translated texts (see Baker 1993, Schmied and Schäffler 1996). The results based on translation corpora therefore have to be verified on the basis of original text corpora. Another disadvantage of translation corpora is that they rarely provide a full or balanced representation of the languages compared. By definition they are restricted to genres and text types that are translated, which tends to confine them to certain written text types. Moreover, what is translated tends to vary from one language to another: for reasons of cultural dominance certain text types may be translated in one direction but not in the other. As a result, translation corpora are seldom large and well balanced, a fact which limits their usefulness for certain types of cross-linguistic studies.

It is obvious from this comparison of the advantages and disadvantages of the two main types of multilingual corpora that they should be seen as complementary sources of cross-linguistic data. The possibility of combining comparable and translation corpora, thus taking advantage of the specific merits of both types, has also been recognised in various contrastive projects, e.g. in the composition of the English-Norwegian Parallel Corpus (see Johansson 1998) and the English-Swedish Parallel Corpus (see Altenberg and Aijmer 2000) and in the cross-linguistic methodology advocated by Teubert (1996).

The cross-linguistic insights gained from translation corpora obviously increase considerably if more than two languages can be compared. One inter-

esting example of a multilingual bidirectional translation corpus involving a number of languages is the Oslo Multilingual Corpus.⁶ The basis of this corpus is the English-Norwegian Parallel Corpus (ENPC) (Johansson 1998), which is closely linked to similar English-Swedish and English-Finnish translation corpora. By extending the ENPC to include translations between English, German, Dutch and Portuguese, it will be possible to compare six languages using English original texts as a starting point.

2.2 Text alignment and search tools

To be maximally useful translation corpora must be aligned in such a way that a unit in the original text is linked to the corresponding unit in the translated text. The linked units can then be displayed together and compared, and parallel concordancers and other multilingual search tools can be applied to the aligned texts.

Translation corpora can be aligned paragraph by paragraph or, more commonly, sentence by sentence, but experiments are also being made to align translation corpora at phrase and word level.⁷ Automatic sentence-level alignment, which was first developed for the French and English versions of the Canadian Hansard (see e.g. Brown et al. 1991, Gale and Church 1991), is normally based on statistical matching of features that link corresponding sentences in the source and target texts, such as sentence length (in terms of words or characters), typographical features (e.g. initial capitals, punctuation marks) and cognate words, but there are also programs that make use of a combination of statistical feature matching and a bilingual lexicon of unambiguous equivalents in the languages involved (see Hofland 1996, Hofland and Johansson 1998).⁸ The main obstacle to automatic sentence alignment is represented by cases where a sentence in the original text has been divided into two (or more) sentences in the translation or, conversely, where two (or more) sentences in the original text have been combined into one in the translation. Sentence-level alignment programs generally achieve a high degree of accuracy, but the result has to be checked and corrected manually. Multilingual alignment, i.e. alignment of a source text and its translations into several languages, has also been carried out with good results (see e.g. Hofland and Johansson 1998:98f.).

Efforts have also been made to align parallel texts at word or phrase level (see e.g. Church and Gale 1991, Kay and Röscheisen 1993, Merkel 1999:113ff.). This is a much more difficult task than sentence alignment, since a given word in the source text may be rendered by many translation equivalents and structural

paraphrases, and sometimes none at all. Word alignment programs must therefore rely heavily on bilingual lexicons, contextual pattern matching and sophisticated statistical techniques. Since perfect word alignment is difficult to achieve, most text alignment programs used today are sentence-based. A survey of various alignment techniques and an examination of two major problems confronting word alignment, viz. the lack of isomorphism of lexical units across languages and the semantic discrepancy between source and target expressions that is often found in translation corpora, is presented by Kraif in this volume.

Text alignment is a prerequisite for parallel concordancers and other multilingual tools. These vary in approach and degree of sophistication. Here we shall distinguish two main types: (1) parallel concordancers and search tools ('browsers') which operate on previously aligned corpora and which identify and present a search word (or expression) in its context together with the corresponding aligned unit in the other language, and (2) word-based concordancers pairing lines of text on the basis of computed word correspondences in the compared languages.

In the first type the user selects a search item in L1 or L2 as input and either (a) leaves the equivalents in the other language open, or (b) pre-selects one or several potential equivalents in the other language. In the former case the program presents all the aligned sentence pairs containing the search item in one of the languages and it is up to the user to identify any relevant equivalents in the aligned output. This is illustrated in the following example, which shows a small sample of a search for *drug(s)* (in bold) in the sentence-aligned English-French Canadian Hansard corpus using the web-based TransSearch interface.⁹

Police have to comfort and question the victims of murderers, rapists, armed bandits, **drug** dealers

Les policiers doivent réconforter et interroger les victimes de meurtriers, de violeurs, de bandits armés et de trafiquants de drogue

It means that cheaper generic **drugs** will not be available to them.

Cela veut dire qu'ils ne pourront plus obtenir de médicaments génériques bon marché.

Each time they stop a car they never know whether the driver is armed, on drugs, a hood or an upstanding member of the community.

Chaque fois qu'il arrête une voiture, il ne sait jamais si le conducteur est armé,

drogué, ou s'il s'agit d'un truand ou d'un membre respecté de la collectivité.

Many young people feel either rejected or marginalized in society which creates additional problems of crime and drug and alcohol abuse.

Dans notre société, bien des jeunes se sentent rejetés ou marginalisés, ce qui occasionne d'autres problèmes de criminalité, de toxicomanie et d'alcoolisme.

If a pre-selected equivalent of the search item is specified, the program only presents the aligned sentence pairs that contain the search item and the pre-selected equivalent in the other language. This is illustrated in the following example, which shows a small sample in KWIC format from a TransSearch bilingual query for *drug(s)* translated either as *médicament(s)* or *drogue(s)*.

drug(s)/médicament(s)

...withdrawal of Bill C-91 which gives brand name	drugs a 20-year market monopoly
... du projet de loi C-91 qui donne aux fabricants de	médicaments brevetés un monopole de 20 ans...
...to the Canadian people access to information as to	drug safety and efficacy.
...à l'information sur l'innocuité et l'efficacité des	médicaments

drug(s)/drogue(s)

A lot of the	drugs that come into this country....
Bon nombre de	drogues introduites dans notre pays...
...who were lured into prostitution, hooked on	drugs and exploited...
...dans la prostitution, rendues dépendantes de la	drogue et exploitées...
...organized crime hides the profits of the	drug trade, international smuggling....
...camoufler les profits du commerce de la	drogue, de la contrebande internationale....
...moving to coastal communities if the	drug trade continues the way it has.
...dans les localités côtières, au train où va le trafic de	drogues.

These bilingual concordances yield a wealth of information, notably on the most frequent multiword units (*drug abuse/dealers/cartels/smugglers/barons/trafficking/trade*) and their equivalents in the other language. In the case of *drogue* and *drug* they are the ideal starting point from which to uncover the rules governing the choice between the singular and plural form in the two languages.¹⁰

The sophistication of sentence-based concordancers or browsers varies, but most programs allow the user to choose which of the languages he wishes to regard as the source language (L1) and which as the target language (L2), to use 'wildcards', and to restrict the search by means of various contextual conditions

or word-class tags (if the corpus is tagged for word-class). Some examples of various types of (paragraph-based or sentence-based) multilingual browsers are *ParaConc* (Barlow 1995), *Multiconcord* (Wools 1998), the *Translation Corpus Explorer* (Ebeling 1998) and the *Pedant Bilingual Concordance* (Ridings 1998). A detailed demonstration of how *Microsoft Word* can be used to align source texts and translations and be combined effectively with a mark-up program and the parallel concordancer *Multiconcord* is given by Corness in this volume.

Word-based concordance programs are closely related to word alignment and are consequently more problematic. This type makes use of a statistical matching technique which creates an index indicating which words in L1 tend to correspond to which words in L2. It takes just one search word as input and uses the pre-computed index of word correspondences to align concordance lines in L1 with their translations in L2 (see e.g. Church and Gale 1991). Obviously, this is a complicated statistical task and the outcome depends on the efficiency of the index and on the 'closeness' of the translation. Parallel concordance programs of this type are still in an experimental stage, and the most robust and immediately useful multilingual search tools available today are therefore concordancers and browsers of the first type.

Even if fully automatic and accurate word alignment and word-based concordancing programs may be a utopian goal, there is no doubt that multilingual research tools, however constructed, are extremely useful instruments for anyone concerned with lexical CL, for theoretical as well as practical purposes. By allowing the user to compare an L1 keyword in its context with its counterpart in another language they make it possible to arrive at empirically founded, richer and much more delicate descriptions of translation equivalents. This is also amply demonstrated in the studies in the present volume, many of which depend, implicitly or explicitly, on various kinds of alignment and parallel concordance techniques.

2.3 Some uses of multilingual corpora

Multilingual text corpora can be used for a variety of purposes in contrastive lexical studies. Their main uses can be summarised as follows (cf. Johansson 1998):

- they offer a firm empirical basis for cross-linguistic lexical studies, providing richer and more reliable information about the degree of correspondence between lexical items in different languages than comparisons based on introspection;
- they give new insights into the lexis of the languages compared — insights

- that are likely to be missed in studies of monolingual corpora;
- they can be used for a range of comparative purposes and increase our knowledge of language-specific, typological and cultural differences, as well as of universal features;
- they can be used to study lexical systems as well as the contextual use of lexical items, and thus provide information about paradigmatic as well as syntagmatic lexical relations;
- they can serve to disambiguate polysemous items, reveal the degree of mutual correspondence of lexical items in different languages, and uncover cross-linguistic sets of translation equivalents in the languages compared;
- they are of theoretical as well as practical importance: theoretically, they provide input data for lexical models and serve as testbeds for lexical theories and hypotheses; practically, they are essential for applications in a number of fields, such as multilingual lexicography and terminology, natural language processing, machine-assisted translation, translator training, information retrieval, and language teaching;
- they illuminate lexical differences between original texts and translations and can be used for studies of individual translation problems and strategies, as well as of language-related and universal translation effects.

In the following sections we shall give a brief survey of some of these uses and indicate some major tendencies in corpus-based contrastive studies of lexis in the last decade. The emphasis will be on theoretical and methodological approaches to the study of lexis, but we shall also touch briefly on some developments in multilingual lexicography (Section 6) and machine-assisted translation (Section 7).

3. Theoretical and methodological issues

3.1 Some contrastive approaches

Traditionally, CL has been described as involving three methodological steps: description, juxtaposition and comparison (see e.g. Krzeszowski 1990:35). The description includes selection of the items to be compared and a preliminary characterisation of these in terms of some language-independent theoretical model. The juxtaposition involves a search for, and identification of, cross-linguistic equivalents. In the comparison proper the degree and type of correspondence between the compared items are specified.

Modern lexical CL often follows this procedure, but a characteristic feature of recent corpus-based contrastive work is the great variety of approaches employed. This is largely due to the expansion of the field and the new research possibilities that multilingual corpora and search tools offer. The methodology chosen and the delicacy of the analysis depend to a large extent on the purpose of the analysis, e.g. whether it is primarily 'theoretical' (focusing on a contrastive description of the languages involved) or 'practical' (intended to serve the needs of a particular application). This in turn may determine the role that the corpus is allowed to play in the analysis. One distinction that is sometimes made in corpus linguistics, and which is also applicable to CL, is that between 'corpus-based' and 'corpus-driven' approaches (see e.g. Francis 1993 and Tognini Bonelli 2001 and in this volume). The former may involve any work — theory-driven or data-driven — that makes use of a corpus for language description, but it is also used in a restricted sense to refer to studies which start from a model postulating a cross-linguistic difference or similarity on theoretical grounds and use a multilingual corpus to confirm, refute or enrich the theory. The latter approach, on the other hand, may start from an implicit or loosely formulated assumption but uses the corpus primarily to discover types and degrees of cross-linguistic correspondence and to arrive at theoretical statements. In practice, however, the distinction may be slight. The difference lies rather in the importance attached to the initial assumptions and the role that the data play in the analysis. Here we shall use the term 'corpus-based' as an umbrella term covering both types of corpus-informed studies.

In the following sections we shall briefly examine some of the theoretical and methodological issues involved and how these have been approached in some recent corpus-based contrastive studies of lexis.

3.2 *Tertium comparationis* and translation equivalence

Any cross-linguistic comparison presupposes that the compared items are in some sense similar or comparable. That is, to be able to say that certain categories in two languages are similar or different it is necessary that they have some common ground, or *tertium comparationis*. For lexis it is obvious that the compared items should express 'the same thing', i.e. have the same (or at least similar) meaning and pragmatic function (see James 1980: 90f.). However, what exactly this 'thing' is is not always obvious, and the problem of identifying a *tertium comparationis* in CL has been discussed a great deal in the past (see e.g. James 1980: 169ff., Krzeszowski 1990, and Chesterman 1998: 27ff.).

Krzeszowski (1990:23f.) has distinguished seven types of equivalence: statistical equivalence, translation equivalence, system equivalence, semantico-syntactic equivalence, rule equivalence, substantive equivalence and pragmatic equivalence. However, although there is something to say for this taxonomic approach, it seems that the only way we can be sure that we are comparing like with like is to rely on translation equivalence (see James 1980: 178). Chesterman (1998:37ff.) develops this in the following way. Any notion of equivalence is a matter of judgement. Similarly, cross-linguistic equivalence is not absolute, but a matter of judgement or, more precisely, translation competence. "On this view, estimations of any kind of equivalence that involves meaning must be based on translation competence, precisely because such estimations require the ability to move *between* utterances in different languages. Translation competence, after all, involves the ability to *relate* two things" (ibid.:39).

The fact that equivalence is a relative concept also has another consequence. It is not realistic to proceed from a *tertium comparationis* that is based on 'identity of meaning'. For one thing, this would be putting the cart before the horse and we would run the risk of methodological circularity: the result of the contrastive analysis would be no more than the initial assumption (cf. Krzeszowski 1990:20). For another, the area we want to explore is often fuzzy and impossible to define satisfactorily (e.g. epistemic modality or pragmatic particles). In such cases we cannot start from a *tertium comparationis* that is founded on equivalence in a strict sense (identity of meaning). Instead, what we have to do — and what we generally do — is to start from a perceived or assumed similarity between cross-linguistic items (cf. James 1980: 168f.). Viewed in this way, CL becomes a way of refining initial assumptions of similarity. Chesterman (1998:58) expresses this as follows:

In this methodology, the *tertium comparationis* is thus what we aim to arrive at, after a rigorous analysis; it crystallizes whatever is (to some extent) common to X and Y. It is thus an explicit specification of the initial comparability criterion, but it is not identical with it — hence there is no circularity here. Using an economic metaphor, we could say that the *tertium comparationis* thus arrived at adds value to the initial perception of comparability, in that the analysis has added explicitness, precision, perhaps formalization; it may also have provided added information, added insights, added perception.

The crucial role that translation equivalence plays in CL has important methodological consequences. We have already described the differences between comparable corpora and translation corpora (Section 2.1). When

items are compared across comparable corpora, it is difficult to know if we are comparing like with like. Any judgement about cross-linguistic equivalence (or similarity) must be based on the researcher's 'translation competence'. This is true at both ends of the analysis: initially, when items are selected for comparison, and finally, when the results of the comparison are evaluated. When we use translation corpora the situation is different. Although we normally start with an initial assumption about cross-linguistic similarity — the very basis for comparing anything at all — we can place more reliance on the translations found in the corpus. The corpus can be said to lend an element of empirical inter-subjectivity to the concept of equivalence, especially if the corpus represents a variety of translators.

However, despite the usefulness of translation corpora, to what extent can we trust the translations we find in them? Can we treat all the translations that turn up as cross-linguistic equivalents? There does not seem to be a simple answer to this question. In one sense, every translation is worth considering as a potential translation equivalent as it reflects the translator's 'competence'. However, translations are rarely literal renderings of the original. Translators transfer texts from one language (and culture) to another and the translation therefore tends to deviate in various ways from the original. We have already mentioned possible translation effects — traces of the source language or universal translation strategies — and they may involve additions, omissions and various kinds of 'free' renderings that are either uncalled for or motivated by cultural and communicative considerations.¹¹

How, then, can we determine which translations should be regarded as 'equivalents' in a stricter sense? One solution has been to resort to the procedure of 'back-translation' (see Ivir 1983, 1987), i.e. to restrict the comparison to forms in L2 that can be translated back into the original forms in L1. This is likely to eliminate irrelevant differences that are due to the translator's idiosyncrasies or motivated by particular communicative or textual strategies.

Another solution is to rely on recurrent translation patterns, i.e. to resort to a quantitative notion of translation equivalence (cf. Krzeszowski 1990:27). If several translators have used the same translation, this obviously increases its relevance. However, this too implies a risk: by restricting the comparison to recurrent translations we may throw away valuable evidence and miss the cross-linguistic insights that 'unexpected' translations often provide.

A variant of this approach which combines Ivir's idea of back-translation and a quantitative notion of equivalence is to calculate what has been called the 'mutual correspondence' (or translatability) of two items in a bidirectional

translation corpus (see Altenberg 1999). If an item *x* in language A is always translated by *y* in language B and, conversely, item *y* in language B is always translated by *x* in language A, they will have a mutual correspondence of 100%. If they are never translated by each other their mutual correspondence will be 0%. In other words, the higher the mutual correspondence value is, the greater the equivalence between the compared items is likely to be. Although the mutual correspondence of categories in different languages seldom reaches 100% in a translation corpus (even 80% seems to be a comparatively high value), a statistical measure of translation equivalence can be a valuable diagnostic of the degree of correspondence between items or categories in different languages (see e.g. Altenberg 1999 and Ebeling 1999: 257ff.). However, it does not tell us where to draw the line between equivalence and non-equivalence. Ultimately, the notion of equivalence is a matter of judgement, reflecting either the researcher's or the translator's bilingual competence.¹² Both involve a judgement of translation equivalence.

3.3 Language system vs. language use

In the past, contrastive analysis was chiefly concerned with comparisons of abstract systems across languages. However, corpora reflect language use, and translation equivalence is always equivalence-in-context (Chesterman 1998: 31). This broadens the scope of contrastive analysis. The aim is to account for both language systems and language use, i.e. the task is not only to identify translation equivalents and 'systematic' correspondences between categories in different languages, but to specify to what extent and in what respect they express 'the same thing' and where similarities and differences should be located in a model of linguistic description.

The extended scope of corpus-based CL creates theoretical as well as methodological problems. As has been pointed out by Salkie (1997) in a comparison of English *but* and French *mais*, translation equivalents in two languages seldom have the same distribution and seldom have 100% correspondence in multilingual corpora. This raises a number of important questions. For example, how regular does an observed difference have to be in order to count as systematic (rather than random or unpredictable)? Where should the difference be located — in the language system (*langue*) or in language use (*parole*)? To what extent can linguistic (sub)systems be isolated from each other, and in what ways do they interact? (See Salkie in this volume for further discussion of this question.)

The fact that translation equivalents seldom have 100% correspondence in translation corpora has been demonstrated in a number of studies. In Altenberg's (1999) comparison of adverbial connectors in English and Swedish not even cognate or functionally similar items like *instead*: *i stället* and *on the other hand*: *å andra sidan* reach a mutual correspondence of 80%. The correspondence of cognate or functionally similar verb pairs across languages tends to be surprisingly low. For example, Altenberg's comparison of the prototypical causative verbs *make* in English and *få* in Swedish (this volume) reveals a mutual correspondence of only 52%. Similarly, Viberg's (1996a: 161) comparison of the cognate verb pairs *go/gå* and *give/ge* in English and Swedish shows that they are only translated by each other in about a third of the cases, and the mutual correspondence of the primary 'possession' verbs *get* and *få* in the same languages is shown to be as low as 15% (Viberg, this volume).

It is obvious that a low degree of mutual correspondence between functionally related items has several explanations. In the case of Viberg's verb pairs the reason is the diverging polysemy and the different meaning extensions that verbs tend to develop in different languages (see Section 4.1). In the case of the English and Swedish connectors examined by Altenberg, some of the differences are clearly system-related. For example, connectors with zero correspondence reveal the existence of lexical gaps in either language: the Swedish explanatory connector *nåmligen* has no exact counterpart in English and the English transitional connector *now* has no counterpart in Swedish. Items with intermediate correspondence values often illustrate differences in the stylistic or functional status of the connectors in the two languages. This is typically revealed by an asymmetrical translation tendency. For example, English *therefore* is more often translated into Swedish *därför* than the other way round, because *därför* is a more common and stylistically more neutral resultive connector in Swedish than *therefore* is in English.

However, there is also evidence of system interchange. This is clearly revealed in Altenberg's comparison of causative English *make* and Swedish *få* in the present volume. In both languages the 'periphrastic' causative verb construction with *make* and *få* can be replaced by alternative constructions, such as a synthetic causative verb or a structurally reorganised causative construction. Epistemic modality is another area where different subsystems tend to interact. For example, as shown by Aijmer (1999) in her comparison of epistemic possibility in English and Swedish, when there is a gap in the Swedish system of modal auxiliaries, it can be filled by a modal adverb. Similarly, when English *may* and Swedish *kan* are not good equivalents, the translators tend to

choose a corresponding adverb or a combination of modal elements.

A similar tendency is revealed in Johansson's (1997) multilingual comparison of the generic pronoun *man* in German and Norwegian and its counterparts in English. Many languages have a generic pronoun (e.g. *man* in German and the Scandinavian languages, *one* in English, and *on* in French), but their frequency and stylistic status vary from language to language. Consequently, translations between such languages tend to display different tendencies depending on the direction of the translation. When a generic pronoun is translated from a language where it is comparatively infrequent (such as English) into a language where it is relatively frequent (such as the Scandinavian languages and, in particular, German and French), it is generally rendered by a generic pronoun in the target language. However, translations in the opposite direction show a different tendency. The generic pronoun in the source language is less often translated by a generic pronoun in the target language. Instead, it tends to be rendered by a range of syntactically restructured impersonal expressions, such as non-finite clauses, agent-less passives, imperatives and nominalisations. These cross-linguistic differences suggest that the *tertium comparationis* needs to be defined at the intersection of several structural systems. Further examples of system interaction will be given in Section 4.2.

The shift from one construction in the source language to another in the target language is often accompanied by a change of viewpoint. For example, in changing an original active clause with generic *man* as subject into either a construction with a specific personal pronoun (e.g. *I*, *he* or *she*) or an impersonal passive or non-finite construction, the translator can in some sense be said to view the situation expressed in the source language from a different perspective. A shift in perspective of a different kind is examined by Salkie (this volume) under the term 'modulation' and used as a way of explaining the various 'unexpected' translations of the German adverb *kaum* into English and of the English verb *contain* into French.

We see, then, that translation corpora confront the researcher with a wealth of different translation 'types' reflecting various degrees of cross-linguistic correspondence. Broadly speaking, these can be said to range from highly recurrent 'expected' translation equivalents to a bewildering variety of 'unexpected' renderings, many of which cross the boundaries between linguistic subsystems and at first sight seem to defy classification. It may be tempting to dismiss such 'unexpected' cases as products of the translator's 'performance', but there is generally a good reason behind the choice of translation. It is the task of the contrastive researcher to evaluate the corpus data as far as possible

and try to see the patterns lurking behind the translator's resourcefulness and behind the most 'unexpected' renderings that turn up in translation corpora.

4. Types of cross-linguistic correspondence

Languages divide up semantic space in different ways. This is a natural consequence of the fact that the conceptual world evolves differently in different languages, for historical, cultural, geographical and social reasons. As a result, complete equivalence between words and expressions in different languages is rather unusual, just as it is unusual to find exact synonyms within one language. This lack of cross-linguistic correspondence is manifested in different ways. The number of concepts encoded in the vocabulary may differ from one language to another. Moreover, the conceptual systems may differ in structure. Familiar examples of this are the ways in which colours and kinship are encoded in different languages. Swedish, for example, has no common term corresponding to English *uncle* or French *oncle* but has to make a distinction between *farbror* 'father's brother' and *morbror* 'mother's brother'.

One consequence of this is that words that are treated as translation equivalents in bilingual dictionaries tend to have different ranges of meaning. An example of this is the relationship between the French, English and German words *bois*: *wood*: *Holz* and *forêt*: *forest*: *Wald* (see Svensén 1993: 141). *Bois* has a wider meaning than *wood*, and *wood* a wider meaning than *Holz*; conversely, *Wald* has a wider meaning than *forest*, and *forest* has a wider meaning than *forêt*. As a result, the meanings of *wood* and *Wald* only partly overlap, and the same is true of *forest* and *bois*. In other words, there is not complete equivalence between any of the words. Partial overlap of a similar kind is revealed by Teubert (1996) in his analysis of English *diary* and *calendar* and German *Tagebuch*, *Kalender* and *Almanach*.

The divergent meaning extensions that have evolved in different languages are especially striking in high-frequency words expressing certain basic meanings. This is clearly illustrated by verbs of motion, perception, and cognition, which occur in most languages with roughly the same basic meanings. At the same time, they are highly polysemous owing to various types of universal and language-specific meaning extensions (see e.g. Viberg 1996a).¹³ The complex cross-linguistic differences these give rise to can be described in terms of such general processes as lexical specification (or elaboration), schematisation (or abstraction), grammaticalisation, metaphorical extension, and idiomatisation.

Cross-linguistically, these developments result in complex patterns of partially overlapping polysemy. Differences of this kind are not only a major problem for language learners, they have also become one of the major stumbling blocks for machine translation and one reason why the lexicon is often described as the 'bottleneck' of natural language processing (see e.g. Calzolari 1996:3 and Sinclair et al. 1996:174). To identify and describe these patterns is a challenge for lexical CL.

However, cross-linguistic equivalence is not only a matter of semantic content. Since the meaning of words is also determined by their grammatical and lexical environment (syntagmatic relations like colligation and collocation) as well as by the situation in which they are used (style, pragmatics), similarities and differences in these respects must also be considered when cross-linguistic equivalence is determined. In other words, equivalence is a complex phenomenon: it involves several levels of linguistic description, and both paradigmatic and syntagmatic relations. We shall not attempt to give a detailed description of various types of cross-linguistic correspondence here. Instead, we shall make a broad distinction between three types of cross-linguistic relationships:

- (a) overlapping polysemy (items in two languages have roughly the same meaning extensions)
- (b) diverging polysemy (items in two languages have different meaning extensions)
- (c) no correspondence (an item in one language has no obvious equivalent in another language)

It should be added that polysemy is not a clear-cut notion. Whether a lexical item can be assigned a certain number of meanings (polysemy) or should be regarded as vague or underspecified with regard to particular items in another language is often difficult to determine. However, it is obvious that translation corpora offer a fertile basis for exploring issues of this kind. In the rest of this section we shall give examples of some recent studies that have explored various types of correspondence. Since overlapping polysemy (in its strictest sense) is relatively uncommon (see however Alsina and DeCesaris, this volume), we shall concentrate on the last two types distinguished above. The difference between paradigmatic and syntagmatic relations will be discussed separately in Section 5.

4.1 Diverging polysemy

Diverging polysemy is a very common phenomenon in contrastive studies of lexis. In a series of studies focusing on high-frequency verbs with similar basic meanings in English and Swedish, Viberg (1996a and b, 1998, 1999, this volume) has explored the divergent patterns of polysemy characterising verbs of motion (such as *go* : *gå* and verbs for 'running', 'putting' and 'pulling') and physical contact verbs (verbs for 'hitting') in the English-Swedish Parallel Corpus. Using a general typological framework, partly inspired by Miller and Johnson-Laird (1976), Talmy (1985) and the frame semantics model proposed by Fillmore and Atkins (1992), he demonstrates that verbs that are usually treated as translation equivalents in dictionaries display surprisingly low mutual correspondence in the corpus, a fact which is due to their various divergent meaning extensions and reflected in a wide range of translations in both languages. Viberg's studies are a good illustration of how theory and cross-linguistic data can interact in a fruitful way. The data serve to test the validity of a language-independent semantic framework, while the framework provides a stable basis for refined descriptions of language-specific and typological lexical differences, as well as of universal semantic categories and principles of meaning extension.

In his contribution to the present volume Viberg compares the Swedish possession verb *få* with its closest English equivalent *get* and, more briefly, with its equivalents in Finnish and French. Starting from basic sense distinctions of *få* and *get* established on the basis of the original texts, he uses their translation equivalents to determine their degree of cross-linguistic correspondence. Viberg finds great conceptual similarities, as regards both their basic and their extended meanings, but the lexicalisation patterns are very language-specific and their mutual translatability low. Another important finding is that the meanings of both verbs can to a large extent be disambiguated by the syntactic frames in which they occur. Some meanings, however, have to be inferred from semantic and pragmatic cues in the linguistic and extra-linguistic context.

A good example of the complexity of cross-linguistic (and intralinguistic) lexical relationships are the multiple correspondences revealed by Chodkiewicz et al. (in this volume) in their comparison of the French legal term *procédure* and the English term *proceedings* in the French and English versions of the European Convention on Human Rights. Both terms are highly polysemous and consequently have multiple equivalents in the other language: *proceedings* has no less than twelve translation equivalents in the French sub-