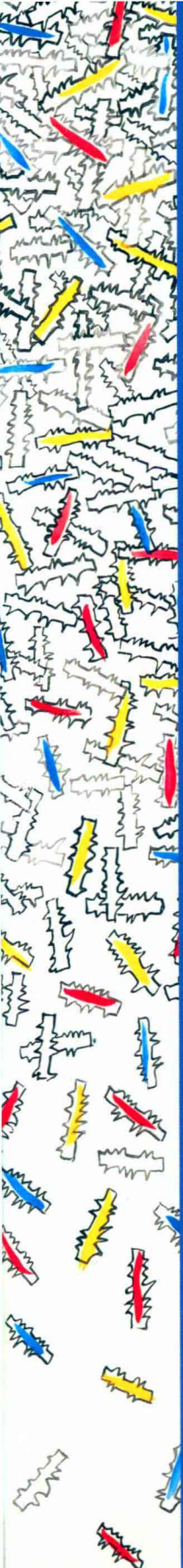

Corpora and Language Teaching

Edited by Karin Aijmer

Studies in Corpus Linguistics 33

JOHN BENJAMINS PUBLISHING COMPANY

Corpora and Language Teaching

Edited by

Karin Aijmer
University of Gothenburg

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.

Library of Congress Cataloging-in-Publication Data

Corpora and language teaching / edited by Karin Aijmer.
p. cm. (Studies in Corpus Linguistics, ISSN 1388-0373 ; v. 33)
Includes bibliographical references and index.
1. Language and languages--Study and teaching. 2. Grammar, Comparative and general--Study and teaching. 3. Corpora (Linguistics) 4. Second language acquisition. I. Aijmer, Karin.
P53.C67 2009
418'.0071--dc22 2008045267
ISBN 978 90 272 2307 4 (Hb; alk. paper)

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

Corpora and Language Teaching

Studies in Corpus Linguistics (SCL)

SCL focuses on the use of corpora throughout language study, the development of a quantitative approach to linguistics, the design and use of new tools for processing language texts, and the theoretical implications of a data-rich discipline.

Volume 33

Corpora and Language Teaching
Edited by Karin Aijmer

List of contributors

Karin Aijmer
English Department
Göteborg University
Box 200
40530 Göteborg
Sweden
karin.aijmer@eng.gu.se

Mia Boström Aronsson
Skövde University College
Högskolevägen
Box 408
541 28 Skövde

Winnie Cheng
Research Centre for Professional
Communication in English
Department of English
The Hong Kong Polytechnic University
Hung Hom
Hong Kong
egwcheng@inet.polyu.edu.hk

Signe Oksefjell Ebeling
Department of Literature, Area Studies,
and European Languages
University of Oslo
P.O. Box 1003, Blindern
0315 OSLO
Norway
s.o.ebeling@ilos.uio.no

Cécile Gouverneur
Centre for English Corpus Linguistics
Université Catholique de Louvain
Collège Erasme
Place Blaise Pascal 1
1348 Louvain-la-neuve
Belgium
line_gouverneur@yahoo.com

Solveig Granath
English Department
Karlstad University
65188 Karlstad
Sweden
Solveig.Granath@kau.se

Sylviane Granger
Centre for English Corpus Linguistics
Université Catholique de Louvain
Collège Erasme
Place Blaise Pascal 1
1348 Louvain-la-neuve
Belgium
granger@lige.ucl.ac.be

Hilde Hasselgård
Department of Literature, Area Studies,
and European Languages
University of Oslo
P.O. Box 1003, Blindern
0315 OSLO
Norway
hilde.hasselgard@ilos.uio.no

Jennifer Herriman
English Department
Göteborg University
Box 200
40530 Göteborg
Sweden
jennifer.herriman@eng.gu.se

Susan Hunston
Department of English
The University of Birmingham
Edgbaston
Birmingham B15 2TT
United Kingdom
s.e.hunston@bham.ac.uk

Stig Johansson
Department of Literature, Area Studies,
and European Languages
University of Oslo
P.O. Box 1003, Blindern
0315 OSLO
Norway
stig.johansson@ilos.uio.no

Fanny Meunier
Centre for English Corpus Linguistics
Université Catholique de Louvain
Collège Erasme
Place Blaise Pascal 1
1348 Louvain-la-neuve
Belgium
fanny.meunier@uclouvain.be

Joybrato Mukherjee
Justus Liebig University Giessen
Department of English
Chair of English Linguistics
Otto-Behaghel-Str. 10B
35394 Giessen
Germany
Joybrato.Mukherjee@anglistik.uni-giessen.de

Ute Römer
University of Michigan
English Language Institute
500 E. Washington Street
Ann Arbor, MI 48104-2028
USA
uroemer@umich.edu

Table of contents

Introduction

Corpora and language teaching

Karin Aijmer

1. Introduction

Corpora have changed our views on language and language use and we can also expect to find them in the class-room. It is not only 'raw' corpora which are of interest but corpora come with user-friendly programs and software which makes them suitable for the use by learners. However it is clear that there are also problems and that we do not know enough about how learners and teachers experience the use of corpora in the classroom. When should corpora be used as part of the teaching of a language? How should they be used? What should be the proper balance between the corpus-based approach and more traditional classroom methods? Are corpora good for all kinds of students?

These questions are of great concern to teachers and scholars who share an interest in corpora and dedication to using corpora in the classroom. In December 2005 a symposium was organised at the University of Gothenburg in order to discuss such questions. A number of scholars with extensive experience of using corpora in their teaching and for applied linguistics research were invited to review the state of the art and to discuss the role and effectiveness of corpora and corpus-linguistic techniques for language teaching. The volume *Corpora and Language Teaching* contains a selection of the contributions from the symposium as well as some commissioned articles.

Corpus linguists are generally enthusiastic about what they have to offer the teaching profession. However the use of corpora in the EFL classroom is a rare occurrence and teachers are still unwilling to or lack the skill to use corpora as an aid to get new insights into English. On a pessimistic note Mukherjee and Rohrbach (2006:205) write that 'we have the impression that in EFL countries like Germany there is a widening gap and a widening lag between on-going and intensive corpus-linguistic research on the one hand and classroom teaching on the other'. The problem is how to reach students and teachers with information about corpora and what they can do. Although courses in corpus linguistics are

sometimes included in the university curriculum the direct exploitation of corpora in the EFL classroom is unusual and the impact of corpora on syllabus and materials design has been slight.

The articles in this volume are geared towards the applications of corpora in the classroom and for pedagogical research. They also deal with broader issues such as the relationship between corpora and second language acquisition or foreign language teaching. Applied corpus linguists and the average EFL teacher have different perspectives on language teaching and it is important to clarify how these perspectives differ. For example, teachers (and learners) look for simple answers to grammatical problems in terms of what is right and wrong and shy away from the fuzzy picture of language as used in the corpus concordance. Discussions of the pedagogical implications of corpora can take two forms. They focus on the use of corpora in the classroom. Moreover they deal with the use of corpora for applied linguistics research in particular the use of learner corpora to get a better picture of how advanced learners write and speak.

2. The use of corpora and second language acquisition

Two of the papers deal with the relationship between corpora and second language acquisition. In **Sylviane Granger's** state of the art paper she discusses several of the divisive issues in the field of learner corpora and in particular how we can establish a better link to second language acquisition research and more cooperation with second language acquisition practitioners. Corpus linguistics and second language acquisition need each other and learner corpora could be a meeting place for SLA practitioners and linguists interested in learner corpora.

There are several reasons why learner corpus research needs to join forces with SLA. As Granger writes (this work) 'a wide range of social, cognitive, and psychological factors that play a role in language learning have been extensively studied in SLA and familiarity with SLA findings will greatly help LC (learner corpus analysis) with a focus on learner production'. As a result there are now clear signs of the two fields coming closer together. Both sides have been quick to recognize the advantage of broadening the empirical basis for SLA analysis. It is for instance clear that learner corpora provide a wealth of empirical material making it possible to examine a number of different variables which have an effect on learner output. Differences between learners and native speakers can for example reflect a transfer effect which can be traced back to contrastive differences and be studied on the basis of multilingual corpora. There is a mutual give and take. If there are differences between the target and source language shown by the translations we can hypothesise that these will affect the way learners use L1. On the other hand,

deviations from the native speaker norm can at least sometimes be explained as transfer and be traced back to the contrastive analysis. The description of the results of learner corpus investigations can lead to improved language teaching materials which pay more attention to forms and structures which pose special problems for learners. The priorities for the future include both a wider range of learner corpora, standardized annotation and 'realistic' learner tools. Corpus annotation is important and PoS tagging is available for the written corpus.

Stig Johansson stresses the importance of arranging controlled experiments where teaching based on different theories of foreign language teaching can be compared. The starting-point for Johansson's paper is an early experiment on explicit and implicit teaching methods which showed that the implicit methods were most successful at least with adult learners. The experiment also raised some questions on corpora and second language acquisition. To what extent is the use of corpora in language teaching tested experimentally? To what extent can the use of corpora be grounded in theories of language acquisition? The experiment reported by Johansson supports the idea that the learner is involved in hypothesis formation and hypothesis testing. There is just a short step from this to the idea that the same processes are involved when the corpus is used for language teaching. The frequency information in the corpus also supports a view of language acquisition where the learner goes to work on finding out the lexico-grammatical patterns helped by repetition and entrenchments of form-meaning links.

3. The use of corpora in the classroom – the learner as a researcher

The classroom provides the framework within which we can expect a lot of direct interaction between learner and corpus. The direct or data-driven application of corpora in the classroom implies that learners get their hands on authentic corpus material and are encouraged to discover things about language without any previous preconception about what they will find (Johns 1991; Bernardini 2004). The corpora can for instance be used to provide concordances or to select examples for learning activities. The articles by Granath, Ebeling and Römer in this volume give examples of different types of corpus-based learning in the classroom.

Solveig Granath shows how corpora can be an integral part of courses in grammar and in spoken and written proficiency. Corpora can for example be used to create exercises, demonstrate variation in grammar, show how syntactic structures can signal differences in meaning, to discuss near-synonyms and collocations.

The corpus techniques are especially relevant when the grammatical rules are very general and do not capture the way in which language is actually used. By

consulting the corpus students and teachers can for instance get a more varied picture of the use of concord with different types of collective nouns than you get by consulting a grammar.

Corpora can also find an answer to 'what teachers always wanted to know' and give informed answers to student questions (cf. Tsui 2004). Students sometimes ask questions about phenomena which are not mentioned in the grammar book and where corpora need to be consulted. In Granath's experience it takes time for students to become skilful corpus users. Many learners were unused to the inductive methods and therefore found this way of working difficult. It is also possible that corpus activities do not suit all types of learners (cf. Estling Vannestål & Lindquist 2007).

Signe Oksefjell Ebeling's paper describes an interactive web-based learning platform at the University of Oslo (Oslo Interactive English or OIE) with the aim to encourage more flexible learning by means of Information and Communication Technology. The OIE is also intended to serve as an introductory course to the use of corpora. The corpus which is used together with OIE consists of 7 million words of fiction and non-fiction and is a slightly slimmed version of the Longman/Lancaster English Language Corpus. The corpus is the basis for a large number of exercises for example multiple choice tasks, gap-filling exercises, error correction and 'open' choices (the students are free to write full answers). By using corpus evidence the OIE makes students reflect on language and critically examine the rules of grammar books.

The statistics on how often the interactive web platform had been used seemed to show that the OIE is popular among learners and that its popularity is increasing. However a closer look indicated there were many more people visiting the OIE's web pages than actually doing the exercises. The way to improve this situation will be to integrate OIE into on-campus teaching.

There are a number of useful corpora and corpus tools waiting to be used in the classroom but we need to know if they give the information teachers and students want and what they are looking for. **Ute Römer** reports on a survey among qualified English language teachers at secondary schools in Germany in order to learn more about the teachers' working situation and to collect comments on their problems and experiences. Together with an experienced practising teacher Römer devised a questionnaire designed to find out for example if the existing teaching material and handbooks gave sufficient support to the teaching of vocabulary and grammar and what the teachers' attitudes were towards corpora. Several informants pointed out that it took too long time to look up words in dictionaries and that they often wanted to consult a native speaker. However a majority of teachers did not see the consultation of one of the major language corpora as an alternative or supplement to the dictionary or grammar. The at-

titudes towards existing coursebooks were often negative which suggests that this is an area where corpora can make a difference. Teachers in general thought that the existing course books offered too few exercises or lacked interesting, authentic material. This was especially the case for spoken language.

4. The use of corpora for applied linguistics research

Learner corpora consist of the learners' own written or spoken production. They have been mainly used for research (what Granger refers to as delayed pedagogical research). An impressive amount of research has been carried out on the basis of the International Corpus of Learner English (ICLE) (see the articles in Granger 1998 and 2002). The ICLE project incorporates a number of national groups with different L1s collecting their own corpora according to shared design criteria such as the level of the learners (Granger 1998, 2002). The studies under the umbrella of the ICLE project have paid special attention to advanced learners.

Advanced learners make few morphosyntactic errors but they may use forms and structures in a non-native way. By comparing the essays produced within the ICLE Corpus with a native speaker norm we can discover subtle features such as overuse and underuse which account for the impression of near-nativeness of learners' essays. In this volume we find two examples of how the study of thematic variation has pedagogical implications for improving advanced learners' competence.

Jennifer Herriman & Mia Boström Aronsson's paper presents the findings from a comparison of how Swedish advanced learners of English and native speakers of English organize the information in argumentative writing using the Swedish component of the ICLE Corpus. The main focus is on the selection of theme and thematic variation. The comparison showed that the learners tended to thematize their opinions and attitudes to a much higher extent than native speakers. In particular learners overused *I think* and *what I want to say*. Swedish learners also overused clefts which were used to thematize new information and to express evaluative comments. The thematic choices made by learners may lead to a persuasive and emphatic style which is not characteristic of the argumentation of native speakers. The reasons for the overuse of certain types of theme include transfer of the native language structures or cultural conventions, general learner strategies such as the use of formulas and lack of knowledge about the conventions for argumentative writing in English.

Hilde Hasselgård explores the extent to which Norwegian learners apply Norwegian patterns in their choice of thematic structure on the basis of the Norwegian component of ICLE. It is shown that word order patterns are transferred

from Norwegian to English and that learners do not seem to have acquired the grammatical and stylistic norms in the relevant genres of English. In particular initial adverbials are overused in the Norwegian learner data compared to authentic English data. Norwegian learners overused extraposition but unlike the Swedish learners they didn't overuse clefts. Like the Swedish learners they often referred to themselves using *I think* and related structures. Another similarity is the overuse of (some) disjuncts such as *of course* and *probably.*

The conclusion we can draw from these two studies is that learners have a good mastery of thematic structures and of thematic variation but they do not know in what styles or registers it is appropriate to use them. Many of the features overused by non-native speakers were for example characteristic of speech rather than of writing.

Other studies focus on learners' problems in the area of phraseology and in particular how phraseology should be presented to learners. **Susan Hunston** discusses how information about corpus frequency can be linked to phraseology and how this information can be presented to learners. The suggested methodology (the 'corpus-driven' approach) gives priority to lexis and the phraseology associated with words. The focus of Hunston's study is on multiword units which are often ranked in the same way as single words with regard to frequency. An alternative view is to recognize that not only overall frequency plays a role but the strength of the collocation between the elements in the units. This results in fairly long phrases representing information about what is relatively often said rather than information about what is grammatically correct. Also when we consider the different forms subsumed under a lemma the probability of occurrence is affected. What we find is semantic sequences where a word form can be related to the complementation pattern and modal meaning by means of the probability of occurrence of each linguistic feature. However such sequences are hardly useful for prescription in the classroom. On the other hand, this type of sequences are important in the devising of teaching materials focusing on the functions, vocabulary and grammar items most needed by learners.

Corpora of learners' production are now also collected by teachers for use in the classroom. The object of **Winnie Cheng's** paper is to use such corpora to study the phraseology typical of the fields of economics and financial services in order to describe the content of the text in terms of 'aboutness'. Aboutness is especially clear when a genre is domain-specific. Central to the study of 'aboutness' is the establishment of collocational profiles on the basis of keywords and their collocates. The text-specific patterns which are found are then compared and their 'about distance' can be calculated by comparing it with a reference corpus such as the British National Corpus. The search engine ConcGram@ makes it possible to handle not only contiguous word combinations but variations in the patterns with

regard to position and constituency. The method and its findings have important implications both for ESP and LSP for example in raising language awareness and increasing knowledge about the 'aboutness' in discipline-specific discourse.

5. New types of pedagogical corpora

5.1 Textbook corpora

Corpora and corpus-based research can have an impact on syllabus design and on the preparation of textbooks, dictionaries, grammars and course-books. Dictionaries are for example generally corpus-based and oriented towards the learners' communicative needs and more and more grammars now base themselves on genuine corpus examples. In the new *Cambridge Grammar of English* (Carter & McCarthy 2006) all the examples represent natural English taken from a variety of written or spoken texts.

However, with a few exceptions (Broadhead 2003; McCarthy et al. 2005) textbooks still shy away from corpora. This is shown by a survey of English for General Purposes textbooks carried out by **Fanny Meunier** and **Céline Gouverneur**. Meunier and Gouverneur have compiled a new type of textbook material corpus (the TeMa Corpus) which contains over 700,000 words of textbooks which are popular on the international ELT market. The corpus was tagged pedagogically with tags referring to the type of exercise. The corpus can for instance be used to provide a list of all the words/expressions practised in the exercises at one level and compare it with other levels. Another use is to investigate the type of metalanguage used in the textbooks to see if the terms are used consistently.

5.2 Spoken learner corpora

Research on the basis of spoken corpora has shown that there is a 'grammar of conversation' with different forms and syntactic structures from what we find in written language. However the teaching of forms and structures which are typical of spoken language is still a neglected area. Textbook dialogues are generally stilted and written-like and lack the features which make texts come alive.

Joybrato Mukherjee focuses on the importance of syntactic features of conversation in advanced German learners' speech and discusses their language-pedagogical implications. The paper presents some findings from a number of case studies based on the German component of the Louvain International Database of Spoken English Interlanguage (LINDSEI). The first case study is concerned

with the differences between spoken and written (learner) language with regard to the number, kind and range of collocations. It is shown that learners are more restricted with regard to the range of verb-noun collocations they use in speech than in writing. Moreover many of the collocations used by the (German) learners were deviant.

Discourse markers such as *you know, well, sort of* are another relevant area for the teaching of spoken grammar. They occur frequently in spoken communication where they are used for interactional and interpersonal functions. Discourse markers are used by both learners and native speakers but they are used in different ways and with different frequencies (cf. also Hasselgren 2002; Müller 2005; Aijmer 2004).

In the third case study the comparison involves spoken performance phenomena such as repetition or pauses which have important roles in speech production. The different trends which can be observed when comparing such phenomena in the speech of learners and native speakers show that the planning pressure is higher for learners and can explain why learners' speech appears to be less fluent and spontaneous.

The upshot of the case studies is that it is necessary to pay more attention in the classroom to forms and structures which are typical of spoken language. These forms include discourse markers but also preconstructed phrases that can help learners to become more fluent (cf. De Cock 2004; Rühlemann 2006). For spontaneous spoken language we can envisage various DDL (data driven learning) scenarios with the purpose to raise the learners' awareness about how a particular marker such as *you know* or routinised phrases are used (cf. also Edmondson & House 1981). In order to become fluent speakers learners also need to practice dialogue techniques which force them to produce speech under real-time online production constraints.

6. Avenues for the future

The picture of the future for corpora in teaching is bright although tempered by what we know about attitudes of teachers and learners. As Römer points out (in this volume and in Römer 2006), corpus linguists have a tough job to meet the challenges from teachers and students who are used to more traditional methods. Corpora draw attention to complex patterns and phraseology rather than regularities and supports the view of language learning as a complex process involving hypothesis formation and testing.

Corpora have an obvious place in the classroom but cannot replace the teacher or language teaching. However the teacher has an important role to guide the stu-