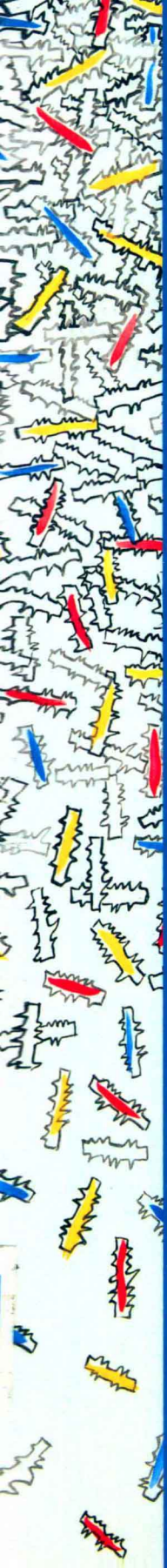

Corpus and Context

Investigating pragmatic functions in spoken discourse

Svenja Adolphs

Studies in Corpus Linguistics 30

JOHN BENJAMINS PUBLISHING COMPANY

Corpus and Context

Investigating pragmatic functions in spoken discourse

Svenja Adolphs
University of Nottingham

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

Adolphs, Svenja.
Corpus and context : investigating pragmatic functions in spoken discourse / Svenja Adolphs.
p. cm. (Studies in Corpus Linguistics, ISSN 1388-0373 ; v. 30)
Includes bibliographical references and index.
1. Discourse analysis--Data processing. 2. Corpora (Linguistics) 3. Speech acts (Linguistics) 4. Pragmatics. I. Title.
P302.3.A355 2008
401'.410285--dc22 2007045722
ISBN 978 90 272 2304 3 (Hb; alk. paper)

© 2008 – 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

Corpus and Context

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.

General Editor

Elena Tognini-Bonelli
The Tuscan Word Center/
The University of Siena

Consulting Editor

Wolfgang Teubert

Advisory Board

Michael Barlow
University of Auckland

Douglas Biber
Northern Arizona University

Marina Bondi
University of Modena and Reggio Emilia

Christopher S. Butler
University of Wales, Swansea

Sylviane Granger
University of Louvain

M.A.K. Halliday
University of Sydney

Susan Hunston
University of Birmingham

Stig Johansson
Oslo University

Graeme Kennedy
Victoria University of Wellington

Geoffrey N. Leech
University of Lancaster

Anna Mauranen
University of Helsinki

Ute Römer
University of Hannover

Michaela Mahlberg
University of Liverpool

Jan Svartvik
University of Lund

John M. Swales
University of Michigan

Yang Huizhong
Jiao Tong University, Shanghai

Volume 30

Corpus and Context. Investigating pragmatic functions in spoken discourse by Svenja Adolphs

此为试读,需要完整PDF请访问: www.ertongbook.com

Acknowledgements

The use of spoken corpora to study pragmatic patterns of speech act expressions was an idea that John Sinclair shared with me during a workshop at the Tuscan Word Centre in the 1990s. It later became the topic of my PhD thesis which forms the basis of this publication. Ronald Carter supervised my thesis and I am greatly indebted to him for his unstinting help, guidance and encouragement. He has been extremely generous with his time and ideas in providing feedback on this work and subsequent research. I have also benefited greatly from the detailed comments provided by the thesis examiners, Mike McCarthy and Guy Cook, and I am grateful to them for raising interesting issues and points of discussion in relation to this work.

I have been very fortunate to be able to draw on the expertise of colleagues here at the University of Nottingham and the IVACS research group while writing this book. I am grateful to them for providing helpful feedback on different ideas and projects that are discussed in this volume.

Special thanks to Dawn Knight who has been instrumental in the research on language and gesture covered in Chapter 6, and to Phoebe Lin who has been a great help in preparing the final version of the manuscript for publication.

Thanks also to Elena Tognini-Bonelli, Kees Vaes and Martine van Marsbergen who provided huge and invaluable support for this project and who have been a real pleasure to work with.

The research discussed in Chapter 6 is funded by the UK Economic and Social Research Council (ESRC), e-Social Science Research Node *DReSS* (Grant N° RES-149-25-0035, http://www.ncess.ac.uk/research/digital_records), and the ESRC e-Social Science small grants project *HeadTalk* (Grant N° RES-149-25-1016). The discussion in Chapter 6 draws on the article 'Beyond the Word: New challenges in analysing corpora of spoken English' which has been previously published in the *European Journal of English Studies* 11(2). I am grateful to Ronald Carter and Dawn Knight for allowing me to use some of the material covered in this article for the purpose of this publication.

Most of the corpus examples in this book are taken from the Cambridge and Nottingham Corpus of Discourse in English (CANCODE). CANCODE is a 5-million-word computerised corpus of spoken English, made up of recordings from a variety of settings in the United Kingdom and Ireland. CANCODE was built by Cambridge University Press and the University of Nottingham and it

forms part of the Cambridge International Corpus (CIC). Sole copyright of the corpus resides with Cambridge University Press, from whom all permissions to reproduce material must be obtained. I am grateful to Cambridge University Press for allowing me to include extracts, frequency lists and concordance data from the CANCODE corpus in this book.

List of tables and figures

Tables

3.1 Collocates to the left of *why don't you* in CANCODE **59**
3.2 Collocates of *suggest* in CANCODE **64**
3.3 The top 10 collocates to the right of the speech act expression *why don't you* at N+1 **67**
3.4 Collocates to the right of the speech act expression *why don't you* at N+2 **67**
3.5 The overall functional distribution of *why not* according to speaker relationship category in CANCODE **73**
3.6 Collocational patterns of *why not* **73**
4.1 Absolute frequency of the speech act verb *suggest* **80**
6.1 O'Keeffe and Adolphs' (forthcoming) framework for classification and functions of backchannels **127**

Figures

3.1 Prototypical function of *Why don't you* according to speaker relationship category in CANCODE **65**
4.1 Absolute frequency of different speech act expressions and their function in the *intimate* category in CANCODE **79**
4.2 The relative distribution of the speech act verb *suggest* according to speaker relationship categories in CANCODE **79**
5.1 Hasan's (1999: 294) model of recounting events **118**
5.2 Reported suggestion episodes (based on Hasan 1999: 294, adapted and extended model) **118**
6.1 Screenshot of Digital Replay System **131**

Table of contents

Acknowledgements ix

List of tables and figures xi

CHAPTER 1

Introduction 1

1.0 Genesis of this book 1
1.1 Overview 2
1.2 Spoken discourse and corpus analysis 5
1.3 Pragmatics and corpus evidence 6
1.4 Speech act expressions 8
1.5 Corpus and context 10
1.5.1 Discourse Context 13
1.6 Why this book? 15
1.7 Organisation of this book 16

CHAPTER 2

Spoken discourse and corpus analysis 19

2.0 Introduction 19
2.1 Language as data 20
2.2 Language functions 22
2.3 Pragmatics and corpus linguistics 23
2.4 Speech Act Theory and corpus linguistics 25
2.4.1 Direct and indirect speech acts 26
2.5 Routinisation 28
2.6 Functional profiles 29
2.7 Speech act contexts 31
2.8 Spoken corpus design and contextual information: An example 33
2.8.1 Context-types 34
2.8.2 Interaction-types 36
2.8.3 Further issues in spoken corpus design 39
2.9 Summary 42

CHAPTER 3
Pragmatic functions, speech act expressions and corpus evidence 43
3.0 Introduction 43
3.1 Lexico-grammar and speech acts 43
3.2 Defining speech acts: The example of suggestions 44
3.3 Speech act classification and discourse analysis 47
3.3.1 Problems with cost benefit scales 48
3.4 Speech acts as strategies? 49
3.5 Speech act idioms revisited 50
3.5.1 Speech acts in a corpus 52
3.6 Functional profiles of speech act expressions 53
3.6.1 Functional prosody 53
3.6.2 Collocation 56
3.6.2.1 Collocation and functional distribution 57
3.6.2.2 Collocation and interpersonal markers 58
3.6.2.3 Collocation and modality 59
3.6.2.4 Collocational patterns of *why don't you*, *why don't we* and *why not* 61
3.7 Summary 71

CHAPTER 4
Pragmatic functions in context 73
4.0 Introduction 73
4.1 Functional profiles and corpus-design 73
4.2 Text and context 76
4.3 Lexico-grammar, speech acts and context 78
4.4 Genre and text-type 79
4.5 Text-types and lexico-grammar 81
4.6 Speech acts and genre 81
4.7 Goals and discourse 82
4.8 Sample analysis 83
4.9 Summary 88

CHAPTER 5
Exploring pragmatic functions in discourse: The speech act episode 89
5.0 Introduction 89
5.1 Making suggestions: Units of analysis at the level of discourse 90
5.2 Static versus dynamic models 92
5.3 Text-types and goal-types 94

5.4 Identifying a speech act episode 95
5.4.1 Agency in suggestion episodes 97
5.4.1.1 Inclusive and exclusive agency 98
5.4.1.2 Solicited versus unsolicited advice 99
5.5 Pattern of problem solving 103
5.5.1 Suggestions and time reference 105
5.6 Categorising replies 106
5.6.1 Supporting moves 107
5.6.2 Challenging moves 109
5.6.3 Clarifying moves 111
5.6.4 Detaching moves 111
5.7 Reporting problems and reported suggestions 112
5.7.1 Integrating agency into story-telling genres 113
5.8 Summary 115

CHAPTER 6
Pragmatic functions beyond the text 117
6.0 Introduction 117
6.1 Multi-modal communication 119
6.1.1 Points of departure: From verbal to visual 120
6.1.1.1 Describing backchannels 121
6.1.1.2 Corpus-based analyses of backchannels 123
6.1.1.3 Merging verbal and visual 125
6.2 Head-nods and backchannels: An integrated approach 127
6.3 Summary 129

CHAPTER 7
Concluding remarks 131
7.0 Opportunities and limitations of spoken corpus analysis 131
7.1 Possible areas of application 133
7.2 The role of context 134
7.3 Future challenges 135

APPENDIX 1
Transcription conventions for the CANCODE data used in this book 137

References 139

Index 149

此为试读,需要完整PDF请访问: www.ertongbook.com

CHAPTER 1

Introduction

1.0 Genesis of this book

The research for this publication draws on a number of spoken corpus resources with a view to gaining a better understanding of the extent to which the analysis of spoken corpus data might support the link between form and utterance function.

When pragmatic and functional theories of language and associated methods of analysis were first developed, the technology to capture and store large samples of spoken discourse in digital format was not yet available. This may be part of the reason for the emphasis on wider contextual and cognitive phenomena in the models that set out to explain how we do things with language. In the absence of spoken corpora which could be used to provide evidence for recurring links between form and function, a substantial proportion of pragmatic theories refer to phenomena that are external to the actual text and discourse in their explanation of utterance force and function.

With the development of spoken corpora of naturally occurring discourse and the accompanying tools required to represent and search this kind of data, it has become possible to re-examine the possible relationship between lexico-grammar, utterance function and discourse context, and to explore possible patterns in this relationship which are not external to the discourse, but which can be described through recurrence of choices at these three levels.

The past 10–15 years have seen the development of a number of spoken corpora, some of which are particularly concerned with representing language used in every-day life. They include the suite of spoken corpora collected at the University of Nottingham, such as the 5 million word Cambridge and Nottingham Corpus of Discourse in English (CANCODE) and, more recently, the Nottingham International Corpus of Learner English (NIC), as well as the Nottingham Health Communication Corpus (NHCC) and the Nottingham Multi-Modal Corpus (NMMC). The conversational data that have been collected as part of these corpus projects afford a new perspective on language function, as they provide the kind of evidence that has been largely missing from traditional theories developed in the area of pragmatics. It has thus become possible to re-investigate and re-evaluate some of the notions that have been developed in this area, such as the notion of the indirect speech act for example. The research presented in this

book draws on these corpora to further explore the link between speech acts and lexico-grammar (see also Adolphs 2001).

When studying corpus data, the close relationship between linguistic form and utterance function becomes very obvious to the point where even very minor variations in form can be linked to a particular variable in the function of the utterance. This observation is, of course, not new and especially in the area of large scale corpus-based and corpus-driven lexicography, the relationship between form and meaning has been well-documented and theorised (see for example Sinclair 1996). Similarly, in the study of spoken discourse and utterance force, it has been shown that certain conversational routines, such as *Why don't you X* for example, can be linked to a particular speech act, in this case the speech act of suggestion (see Aijmer 1996).

However, the nature of the relationship between such 'conversational routines', as Aijmer calls them, and the context in which they are used, remains relatively under-explored. Yet, in order for us to be able to understand the extent to which conversational routines are 'primed' (Hoey 2005) to be used in a particular context and within a particular co-text, it is necessary to investigate whether there are any patterns within the text itself that can be linked to the use of a particular routine. Such patterns may be visible within the immediate lexico-grammatical environment of the routine or they may be evident within the wider discoursal context.

Other patterns might not be as easy to track due to the fact that a transcript of a conversation cannot offer a systematic and searchable record of the actual event. Corpus linguistic techniques are currently not able to track some of the key elements for the construction of meaning such as gesture and intonation for example. These are part of the immediate context, and they play an important role in the interpretation of utterances. Although technology to capture video and audio records of discourse events in natural settings is developing fast, the exploration of multi-modal corpora is still in its infancy. Yet, early research in this area suggests that we may have to go beyond the transcript to make sense of some of the regularities that exist between the choice of particular words or sequences of words and specific functions in discourse (see Knight et al. 2006). In order to explore the relationship between different channels of interacting in a discourse event, this book will also report on recent developments and analyses of multi-modal spoken corpora.

1.1 Overview

The main development of corpus linguistics as a methodology for language description has largely taken place in the area of lexicography, especially in relation

to the ELT context, where patterns of usage of individual words or multi-word expressions are derived through the analysis of multi-million word corpora. Since such a focus requires a high degree of recurrence of the individual items that are being analysed, the size of the corpus resource has traditionally played a major part in achieving stable analytical results. As a consequence, the corpus resources that have informed this kind of research consist mainly of written language as this kind of variety is much easier to collect than spoken discourse, which has to undergo time and cost-intensive transcription before it can be subjected to corpus analysis.

There has thus been a noticeable focus on written language in the field and with written corpora now exceeding the one-billion-word mark, the possibilities for generating new insights into the way in which language is structured and used are both exciting and unprecedented. Spoken corpora, on the other hand, tend to be much smaller in size and thus unable to offer the same level of recurrence of individual items and phrases when compared to their written counterparts.

A further reason why the analysis of spoken discourse with the use of corpus-linguistic approaches has not received the same attention as the analysis of written discourse is because of the scepticism towards the applicability of corpus-based techniques to issues beyond the clause boundary (Conrad 2002:86). The analysis of discourse phenomena such as intonation, gesture or turn-taking for example requires different types of techniques compared to the study of lexico-grammar in a corpus as the units of analysis are not as clear-cut. Yet, those kinds of discourse level aspects of spoken interaction clearly form part of meaning construction and constitute one of the distinguishing features of spoken versus written discourse. It seems questionable then that the same techniques developed for written corpus analysis should be sufficient or appropriate for exploring spoken corpora, not least because discourse is an essentially collaborative event which is co-constructed by a number of participants in a discourse sequence where one contribution may directly influence the next. Given these issues, it would appear logical that when it comes to corpora of spoken discourse, the unit of analysis should be different to that of written discourse, although this is debatable and requires further exploration.

One of the aims of this book is to look more closely at the way in which smaller, spoken corpora might be analysed, not only in terms of lexico-grammatical patterns, but also in relation to issues which are at the heart of spoken discourse analysis, such as utterance function. This approach requires an investigation of a number of aspects of the discourse recorded in a spoken corpus. These include concordance level analysis and the description of co-textual patterns of selected words or phrases, as well as an analysis and grouping of contextual patterns which may be relevant in terms of a recurring match between form and function. The grouping and coding of contextual variables is an important step in this process and will be discussed in more detail throughout the book.

In order to explore further what a spoken corpus can offer in terms of describing and analysing utterance function, this book makes reference to four key areas of research: corpus linguistics, and more specifically the analysis of concordance outputs; pragmatics, with particular reference to utterance function and levels of directness; discourse analysis, with a focus on patterns and sequences in extended stretches of discourse; and context analysis, and the relationship between textual patterns and contextual variables. I argue that all four approaches are required to gain a better understanding of the scope of spoken corpus analysis in describing functional properties in spoken discourse, and that a combination of different approaches may assist us in the evaluation of specific concepts in each of the different areas.

To illustrate these points, and as outlined above, the book focuses on the description and extraction of functional properties of specific lexico-grammatical strings which have traditionally been associated with a particular speech act function, such as the string *Could you just* has been associated with the speech act of making a request. The book further explores the implications of using a spoken corpus in the analysis of speech act functions for some of the key notions within speech act theory, such as the distinction between a direct and an indirect speech act. In speech act theory, this distinction depends largely on the congruency of syntax and intended speech act function, i.e. on the match between sentence meaning and intended meaning, and on the use of modal items which may hedge the illocutionary force of an utterance. Following a corpus-based approach to analysing speech act function, this book considers whether the distinction between direct and indirect speech acts is useful and can be sustained, especially in the light of their association with wider social phenomena, such as politeness for example. In this context, it is argued that the notion of 'indirectness', as used in speech act theory, may have to be reconsidered and linked more explicitly to a larger discourse-based framework which takes account of the relationship between lexico-grammatical patterns and contextual variables as evidenced in language use.

One of the main questions that is explored in this book relates whether a pragmatic theory is necessary, or useful, when it comes to our understanding of utterance function, or whether a corpus-based analysis may provide a framework for describing such functions which relies mainly on a systematic extraction of text internal patterns. To deal with this question in some depth, the focus of the study has to be relatively narrow, and most of the chapters thus concentrate on one speech act function to exemplify the overall approach, that of making suggestions. In addition, the last chapter considers the function of signalling active listenership to illustrate the effect of working with multi-modal corpora in this field. While the use of backchannels in language is not a concern of traditional speech act theory, this area lends itself to illustrating the difference in scope when using spoken corpora that consist of transcripts only compared with multi-modal corpora. Signals

of active listenership or 'backchannels' have an immediate gestural correlate, i.e. nods of the head, and the analysis of multi-modal corpora allows us to highlight the importance of the role of the listener in the way in which we may assess the utterance function expressed by the speaker.

1.2 Spoken discourse and corpus analysis

Following on from the early developments of relatively small sized spoken corpora in the 1960s, such as the London-Lund Corpus for example, the past two decades have seen major advances in the collection and development of spoken corpora, particularly in the English language but not exclusively. Some examples of spoken corpora are the Cambridge and Nottingham Corpus of Discourse in English *CANCODE* (McCarthy 1998), a five million word corpus collected mainly in Britain,[1] the Limerick Corpus of Irish English *LCIE* (Farr, Murphy and O'Keeffe 2004),[2] the Hong Kong Corpus of Spoken English *HKCSE* (see Cheng & Warren 1999, 2000, 2002),[3] the Michigan Corpus of Academic Spoken English *MICASE* (Simpson, Lucka & Ovens 2000),[4] and the Corpus of Spoken Professional American English *CSPAE*.[5] In addition, there is a growing interest in the development of spoken corpora of international varieties of English and other languages, as well as of learner language (e.g., Bolton et al. 2003; De Cock et al. 1998). These corpora provide researchers with rich samples of spoken language-in-use which form the basis of new and emerging descriptions of naturally occurring discourse.

Current research outputs based on the analysis of spoken corpora are wide-ranging and include, for example, descriptions of lexis and grammar (e.g. Biber et al. 1999; Carter & McCarthy 2006), discourse particles (Aijmer 2002), courtroom talk (Cotterill 2004), media discourse (O'Keeffe 2006) and health care communication (Adolphs et al. 2004). Current research in spoken corpus linguistics covers phenomena at utterance level, as well as at the level of discourse. A number of studies start with the exploration of concordance outputs and frequency information as a point of entry into the data and carry out subsequent analyses at the level of discourse (e.g. McCarthy 1998), while others start with a discourse analyt-

1. www.cambridge.org/elt/corpus/cancode.htm

2. www.ul.ie/~lcie/homepage.htm

3. http://www.engl.polyu.edu.hk/department/academicstaff/Personal/ChengWinnie/HKCorpus_SpokenEnglish.htm

4. http://micase.umdl.umich.edu/m/micase/

5. www.athel.com/cpsa.html

ical approach followed by subsequent analyses of concordance data. There seems to be a common thread, then, in the research which uses spoken corpora, in that at least some consideration is given to the language that surrounds a particular word or sequence of words that goes beyond the span of +/– 4 at the concordance level. Recognition of the fact that spoken corpora are best analysed by using a combination of concordance level and discourse level description requires us to articulate clearly the relationship between the different frameworks that may be used to support such an approach.

The discourse level frameworks that may be relevant for the analysis of spoken corpora are not necessarily compatible with the kind of concordance-based and frequency driven analyses that are used in large scale lexicography studies. One of the key differences between spoken and written corpora is that most spoken discourse is collaborative in nature and as such it is more fluid and marked by emerging and changing orientations of the participants (McCarthy 1998). Yet, it is important to identify external categories for grouping transcripts in a corpus, especially where levels of formality and other functions are concerned which need to be judged against the wider context of the encounter. This process tends to be much more straightforward when dealing with written texts, as many of the genres that are used for written corpus analysis are well established, such as fiction versus non-fiction, letters versus e-mails etc. The group membership of such texts is more clearly demarcated than is the case with the majority of spoken discourse. The development of suitable frameworks for analysing spoken corpus data is thus particularly complex and further research is needed to assess whether the analysis of concordance data and discourse phenomena can ever be fully integrated.

1.3 Pragmatics and corpus evidence

The systematic exploration of utterance function has been a key concern in the area of pragmatics. One of the general assumptions in pragmatics is that the interpretation of language in use is based in part on the meaning of the actual words that are being used and in part on other sources of knowledge, including knowledge about contextual and cultural norms. As such pragmatic theories do not assume a one to one relationship between language form and utterance function, but instead are concerned with accounting for the processes that give rise to a particular interpretation of an utterance that is used in a particular context.

In doing so, a range of different methods and approaches have been developed which come under the very broad heading of Pragmatics. Spencer-Oatey and Zegarac (2002:84) distinguish between the cognitive-psychological approach which is marked by the analysis and discussion of decontextualised utterances and their possible interpretations, and the social-psychological approach. The latter tends

此为试读,需要完整PDF请访问: www.ertongbook.com