Metadiscourse in L1 and L2 English

Annelie Ädel

Studies in Corpus Linguistics 24

JOHN BENJAMINS PUBLISHING COMPANY

Metadiscourse in L1 and L2 English

Annelie Ädel
University of Michigan, Ann Arbor

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 illustration from original painting *Random Order*
by Lorenzo Pezzatini, Florence, 1996.

Library of Congress Cataloging-in-Publication Data

Ädel, Annelie.
Metadiscourse in L1 and L2 English / Annelie Ädel.
p. cm. (Studies in Corpus Linguistics, ISSN 1388–0373 ; v. 24)
Includes bibliographical references and indexes.
1. English language--Discourse analysis. 2. Grammar, Comparative and general--Reflexives. I. Title. II. Series.

PE1422.A34 2006
420.1'41--dc22 2006045728
ISBN 90 272 2297 5 (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

Acknowledgements

There are several people without whom this book would never have been written. First and foremost, I want to express my deep gratitude to my husband Gregory Garretson for his support and criticism, especially during the long existence of this project as a Ph.D. dissertation. His sense of style and editing skills are extraordinary, and his magic wand (the red pen) transformed my scribblings into *text*.

Karin Aijmer and Bengt Altenberg supervised my dissertation project, out of which this book has grown, for which I am very grateful. Special gratitude is also extended to Anna Mauranen for taking on the role as official opponent at my doctoral defence at Göteborg University, Sweden, in September of 2003. I would also like to mention the members of my defence committee: Mall Stålhammar, Britt-Louise Gunnarsson and Hilde Hasselgård. Many thanks are also due to Jennifer Herriman, who, despite severe time constraints, read the whole manuscript in the summer of 2003.

More recent benefactors include Elena Tognini-Bonelli for suggesting that my dissertation be developed into a monograph and published in the *Studies in Corpus Linguistics* series. I am grateful to an anonymous reviewer for recommendations on the first draft of the text. My heartfelt appreciation also goes out to John Swales for reading and commenting on parts of the manuscript – and for his mentorship.

It must be said that all errors and unclarities that remain after the input of so many excellent scholars are my own.

In the course of my Ph.D. work, I was privileged to be awarded a generous grant by the Knut and Alice Wallenberg Foundation, as well as a scholarship from the Faculty of Arts and Humanities at Göteborg University, for which I am grateful. With the help of Cathy O'Connor, I was also fortunate enough to enjoy a two-year stay as a visiting scholar in the Applied Linguistics program at Boston University. At the time of writing, I am enjoying a privileged position as a post-doctoral research fellow in the English Language Institute at the University of Michigan, which has made it possible for me to find time to revise this manuscript.

On a more personal note, finally, I would like to acknowledge Ylva Hård af Segerstad Hasselgren's friendship, which was indispensable during the initial writ-

ing process. I am also indebted to Andreas Nordin, entertaining host and *amicus*. Last but not least, many thanks go to all of my fellow doctoral students at Göteborg University for providing such a pleasant and stimulating educational atmosphere.

Annelie Ädel
Ann Arbor, Michigan
December, 2005

Table of contents

CHAPTER 1

Introduction

1.1 Background

When we communicate, we not only talk about the world and about ourselves (many people's favourite topic), we also use language to *talk about talk*. Humans can explicitly refer to themselves not just as experiencers in the world, but also as communicators, and can comment not just on the topic of discussion, but also on the situation of communicating.

The term 'reflexivity' is usually defined as the capacity of natural language to refer to or describe itself (see Lyons 1977: 5), although it is of course *speakers* and *writers* who use language to refer to or describe it. All human languages, whether in spoken or written form, can be used reflexively by speakers to comment on linguistic matters. Not only is reflexivity a universal feature of language (Hockett 1977: 173), it is also unique to human communicative systems. Silverstein (1976: 16) claims that this property is "what makes language unique among all the cultural codes for social communication."

One might think that reflexivity in language is restricted to the discourse of linguists or other professionals in society at large who spend a great deal of time thinking about language, such as translators, language teachers, lexicographers, editors, proof-readers, speech pathologists, or developers of speech recognition software. However, as pointed out by Roman Jakobson (e.g. 1980), it is actually a central feature of language in everyday use. Just to give some examples of the everyday flavour of some types of reflexive language, below is a list of expressions (from both spoken and written genres) that should be familiar to most speakers of English.

(1) I am writing to tell you the latest news...
I never said that!
Say it with flowers.
We will return to this issue in Chapter 6.
There are no words for how I feel about this.
It's just a manner of expression.
To cut a long story short, ...
Speak up, please!
Good question. I don't have an answer.

> You'll probably think I'm crazy for saying this, but...
> She loves digressions and can never keep to the point.
> What does 'gerrymandering' mean?
> **hy°per°bo°le** *n.* A figure of speech in which exaggeration is used for emphasis or effect, as in *I could sleep for a year.* [<Gk. *huperbole*, excess: HYPER- ? *ballein*, throw.]

Each of these examples is reflexive. Most of them are likely to be used in non-professional discourse. The last example represents a dictionary entry, which is something that lay people and professionals alike consult from time to time. So, having shown that reflexivity in language is not restricted to discourse for specific purposes, it seems appropriate to quote Verschueren (1999: 187–188, emphasis added), who recently argued that reflexivity may be "so central that it could be regarded as one of the original evolutionary prerequisites for the development of language. [In fact,] *all verbal communication is self-referential to a certain degree.*" Not only is reflexivity abundant in everyday discourse, but some scholars argue that it is fundamental to human communication.

One type of reflexivity in language is 'metadiscourse', a term which is usually reserved for written language. The present investigation will consider only written texts, but metadiscursive phenomena are of course an essential part of spoken texts as well (see for example Schiffrin 1980 on 'metatalk'). A shorthand definition of it is 'text about text'. Metadiscourse is discourse about the evolving discourse, or the writer's explicit commentary on her own ongoing text.[1] In this sense, it includes reflexive linguistic items that refer to the text itself as text or as language. In a wider sense, it refers to linguistic items which reveal the writer's and reader's (or speaker's and hearer's) presence in the text, either by referring to the organisation of the text or by commenting on the text in other ways. Metadiscourse thus helps a writer to "guide, direct, and inform" (Crismore 1989: 64) her reader about how she hopes he will respond to its content and is an important category both in creating and in reading text. As a kind of comment on the running text that is usually distinguishable from the content or subject matter of the text, metadiscourse has also been labelled 'non-topical material'.

Metadiscourse can take many different forms, ranging from morphemes, single word forms, phrases, clauses, to strings of sentences.[2] Morphosyntactically, it can be represented by a range of different structures. Adverbials form one of the most typical categories (e.g. *in other words, as noted earlier, stated formally, secondly*). If we look at metadiscursive units in terms of their internal syntax, they often contain an overt subject in the form of a pronoun plus a verb phrase, as in the more or less routinised forms *you know, as we have seen, I will discuss NP.* They may also be impersonal non-finite constructions, consisting for example of the infinitive form of *verba dicendi*, such as *to discuss, to conclude, to put it bluntly.*

Drawing on actual texts, Crismore (1989) provides a plethora of examples showing that Western writers from Antiquity to today, in areas as diverse as history, drama, handbooks, poetry, religion, biography and science, have found it natural to use metadiscourse in their texts. Despite their abundance, however, metadiscursive phenomena are far from fully explored. Researchers do not even agree on what metadiscourse is, nor have they sufficiently clarified how it works in general, how it varies across genres, or how it differs across languages.

By all appearances, however, metadiscourse is becoming a highly dynamic topic in text/discourse research. In Hyland's (1998:437) words, metadiscourse is "a relatively new concept but one which is increasingly important to research in composition, reading and text structure." Researchers are beginning to explore it from a wide range of perspectives in linguistics, for example in contrastive studies (Mauranen 1993; Markkanen et al. 1993), historical stylistic change (Taavitsainen 2000), pragmatics (Verschueren 1999; Hyland 1998), and genre studies, such as in Bäcklund (1998) and Bondi (1999).

The recent interest in metadiscourse stands in marked contrast to the previous widespread neglect of it. As diplomatically described by Vande Kopple (1988:234), the overriding focus of pre-1980s research on text was on various aspects of the subject matter itself: "Much of the recent work on the nature of informative texts and on the processes that readers apply to them proceeds as if there is only one kind of meaning in such texts, the referential, ideational, or propositional". Also pointing to this neglect, Crismore (1989:vii) voiced her concern occasioned by "the lack of awareness of metadiscourse and lack of attention given it in the domains of reading, composition, educational psychology, language studies, social science, and science".

Another telling example of the earlier view of metadiscourse as not being worthy of much attention, whether in theoretical or in applied fields, is provided by Stubbs (1996:228), reviewing a study of the Canadian Hansard corpus[3] by Slembrouk (1992), who claims that "although Hansard is often regarded as a verbatim record of what is actually said in Parliament, considerable changes are made to the words spoken". These changes all concern metadiscourse and expressions of speaker attitude, such as *I hasten to stress, I think, I can only say* and *actually*, which have been deleted. In Stubbs' (ibid.) words, there is a "systematic removal of interpersonal and textual dimensions of utterances [...] from the written record".

In spite of such attitudes to metadiscourse as meaningless or superfluous verbiage of low status in linguistic research, however, the increasing number of studies of metadiscourse do bear witness to the fact that things are changing. This is not to say that we are anywhere near having a full account of what metadiscourse is and how it works in and varies across texts. One of the many areas in which studies of metadiscourse are lacking is L2 writing, which is where the present study enters the picture. This study aims to contribute to the study of L2 writing by analysing

the written production of advanced learners of English and comparing it with that of native speakers of British English (BrE) and American English (AmE).

At this time, there may be as many as 1.5 billion speakers of English in the world, counting liberally (see e.g. Strevens 1992: 28). Most of them are non-native speakers. In what Kachru (1992: 38) calls the 'inner circle', there are some 350 million speakers of English as a first language. It is much harder to estimate the sizes of the 'outer circle' of speakers for whom English is a second language, and the 'expanding circle' of speakers of English as a foreign language, but it is clear that these actually outnumber the inner circle of native speakers. In the 21st century, in the majority of communicative situations in which the English language is used, it is actually being used by non-native speakers.

Whether we like it or not, at this point, English is culturally, politically and economically one of the most important languages in the world. As a result, reading and writing in English have become important to a great number of people. Composition in English is important not only in academia, but also in the corporate world, and in semi-professional or non-professional areas – for example, to people who communicate electronically (and internationally) via newsgroups, chatrooms, or e-mail. A large number of these communicators are not native speakers, but do their reading and writing in English as a foreign language. This is a task that makes great demands on their linguistic abilities and communicative competence, extending beyond basic knowledge of grammar, vocabulary and spelling. The demands involve rhetorical skills in forming texts of various types, including, for example, knowledge of how to present facts effectively, how to argue one's case convincingly, and how to manage writer and reader visibility. We should not make the mistake of assuming that, as long as a learner is fluent in the foreign language and masters essential parts of its grammar and vocabulary, writing texts is a straightforward matter.

1.2 Aims

The overall aim of the present study is to compare the use of metadiscourse in written argumentative texts by advanced learners of English whose first language is Swedish with comparable texts written by native speakers of British and American English. The specific objectives are to contribute to (a) the theory of metadiscourse, (b) the application of computer-assisted methods to studies of metadiscourse, and (c) our knowledge of the use of metadiscourse in English argumentative writing (with special focus on learner writing).

While many researchers in this area note that metadiscourse is a fuzzy and heterogeneous category in need of better definition and clearer delimitation, they do not tend to make theoretical issues their primary concern. This is the first gap that

I intend to start to fill. The second gap is a methodological one; since the majority of investigations into metadiscourse have been carried out manually, there is a lack of systematic computer-assisted studies of metadiscourse. By testing the possibilities of applying computer-assisted methods to the study of metadiscourse, we stand to gain in our ability to generalise, in analytical consistency and in speed. The third gap concerns the fact that no previous research exists on the use of metadiscourse by L1 Swedish speakers writing in English, or on possible metadiscursive differences between AmE and BrE writing.

These three main lines of investigation are intended to work together and cross-fertilise in various ways. I make the assumption that metadiscourse is a textual parameter in which the corpora involved will differ in interesting ways: the non-native speakers' use of metadiscourse seems quite different from that of the native speakers. The corpora to be used consist of argumentative writing, which is an important text type in L2 English composition.[4] Essentially all previous research into metadiscourse investigates academic writing of various kinds, so there will be maximal opportunities to take advantage of previous findings on metadiscourse, which, after all, is a relatively unexplored area. Furthermore, it is often assumed that argumentative writing contains a great deal of metadiscourse. Crismore (1989: 70), for example, states that metadiscourse is "quite prevalent" in argumentative writing, since "authors refer quite frequently to the state of the argument, to the reader's understanding of it, or to the author's understanding of his own argument". In addition, argumentative writing is a highly interesting text type to study due to its general importance in society at large, ranging from the university to the media.

A further example of cross-fertilisation is that the methods can be designed to contribute to the theory of metadiscourse. The design of the present methods will allow me to take into account a wide variety of potential examples of metadiscourse, such that different subcategories as well as borderline cases can be identified more easily and in a more systematic manner.

Stated briefly, the main research questions are the following:

(a) What is the most appropriate definition of and theoretical model for metadiscourse? Can we find explicit features to use in defining what we mean by metadiscourse? Where do we draw the line between metadiscourse and neighbouring categories?
(b) How can computer-assisted methods and large collections of text be used for the analysis of metadiscourse? How can we find patterns for various aspects of the use of metadiscourse by applying computer-assisted methods?
(c) What differences are there between the use of metadiscourse by learners and the use of metadiscourse by native speakers of British and American English, as evidenced in the corpora? To what extent do the different groups of writ-

ers make visible the current discourse situation, the text itself, and the writer and/or the reader?

What types of discourse functions are fulfilled in the texts when the writer persona or the imagined reader are visible? Do the learners and the native speakers use metadiscourse for the same purposes and in similar proportions?

Where in the text does metadiscourse occur, i.e. what are the patterns of distribution, if any, across corpora? Are there tendencies for certain types of metadiscourse to occur in specific positions in a text? For example, can we find evidence that beginnings and endings are particularly suitable locations for writer and reader visibility?

Section 1.5 below gives an outline of the study, in which these research questions are presented in the order in which they will be dealt with. First, however, the next two sections will discuss the material and the methods used in the study.

1.3 Material

The texts used for the study are essays written by university students. These essays are available in an electronic format, collected within the *International Corpus of Learner English* (ICLE) project initiated and coordinated by Sylviane Granger at the University of Louvain in Belgium (see e.g. Granger 1993, 1998). The present investigation uses the Swedish learner subcorpus ('Swicle' for short).[5] In addition to the current 14 subcorpora of various language backgrounds, this international corpus project also offers a comparable corpus of native-speaker material, referred to by the acronym LOCNESS (the *Louvain Corpus of Native English Essays*; see Granger 1993).[6]

The design of the ICLE project is such that results based on any, or all, of the learner corpora may be evaluated against a norm or standard of comparison in the form of a native-speaker corpus. The general idea is that "[w]hen matched with comparable native-speaker texts, a learner language corpus provides the basis for revealing the characteristics of learner language, e.g. identifying interference from the mother tongue" (Johansson 1998: 7).

The learner subcorpora and the native-speaker corpus provide a basis for comparison not only of native language and interlanguage data, but also of parallel interlanguage data (produced by speakers with different L1s). The design of the ICLE corpus makes possible the study of a range of learner subcorpora (for comparison of different interlanguages, see Granger 1998: 12ff.). However, the present investigation is restricted to one specific group of learners, whose first language is Swedish, and only those learners' use of metadiscourse will be examined here.

That said, the methods developed here could easily be used to map metadiscourse in any of the other subcorpora. That way, the results would be directly comparable.

The ICLE corpora make it possible to determine which linguistic features are used considerably more or less by learners of various first language backgrounds, as compared to native speakers. As pointed out by Altenberg (1997: 121) "the difficulties advanced learners experience are mainly lexical and stylistic in character", and the non-nativeness or lack of idiomaticity of advanced learners' production is "not so much a matter of actual errors as of over- and underrepresentation of words and structures". At this advanced level, the focus of research will be on 'overuse' and 'underuse' of linguistic phenomena, rather than error analysis.

In fact, differences between advanced learners and native speakers can be subtle, and it may require a large amount of data to establish what they are. As Lorenz (1999: 53) puts it: "Where [advanced learners] do deviate from the norm, their deviations usually – in written production at least – concern finer points of lexico-grammar and style [, which] are difficult to pinpoint individually, yet in their accumulated effect [they] tend to generate an impression of 'non-nativeness' or 'lack of idiomaticity'". This 'accumulated effect' can be read in two ways: as referring either to individual essays or to a whole corpus-full of essays. In other words, the effects of non-nativeness may be evident within a single essay, but only after reading most or all of it; or they may not be evident in individual essays, but emerge as a general picture when a large number of essays are examined at the same time.

Levenston (1971: 115), discussing 'over-indulgence' and 'underrepresentation' in non-native language use, points out that "[a]lthough the L2 clause structure produced by group level translation may be completely acceptable, both grammatically and lexically, it can still produce effects by no means intended by the user: verbosity, formality, informality, and so on". This is what mastering a language (whether a first or foreign language) is about: the ability to produce intended effects in ways that are efficient and appropriate to the context, and to avoid producing unintended effects.

For the present study of the use of metadiscourse, the learner texts were analysed for interference at the level of discourse. Since discourse competence (or the ability to compose text efficiently) is one aspect of a speaker's communicative competence (see Hymes 1992), we could say that the learners' communicative competence is being investigated.[7]

The corpora used in the present study will be briefly described in the following. For more detailed analysis of the corpora, see Appendix 1, which includes a discussion of the comparability of these corpora, based on Bhatia's (1993) genre criteria, and Appendix 2, a general discussion of the control corpus and the norm.

The learner material amounts to approximately 205,000 words, or 350 essays. The essays are full-length, with an average size of about 560 words. The essay writ-

Table 1. Sizes of the corpora

Corpus	Variety	No. of words	No. of essays
Swicle	L2 English (L1 Swedish)	205,000	350
Locness AmE	American English	150,000	175
Locness BrE	British English	95,500	118

ers are all advanced learners of English: students in English departments at three different Swedish universities (Lund, Göteborg and Växjö). The native-speaker corpus consists of around 247,000 words altogether and comprises some 290 essays written by university students, half speakers of American English (AmE) and half speakers of British English (BrE).

For present purposes, the Locness was divided into two different subcorpora on the basis of regional variety. It is assumed that cultural background may be an important factor and that essays written by British and American writers may exhibit differences in the use of metadiscourse. The 175 AmE essays have an average of 856 words per essay and 150,000 words in total. The 118 BrE essays have an average of 810 words per essay and approximately 95,500 words in total.[8]

The three corpora that the present investigation sets out to contrast are summarised in Table 1.

1.4 Method

The methods used in this study are primarily comparative and corpus-based. The comparative method will be used in order to locate differences in the use of metadiscourse between learners and native speakers of English, but also to identify possible differences between the different native-speaker groups. In choosing to do a computer-assisted study of somewhat sizeable corpora, I embrace the empirical principle that the more material the analysis is based on, the safer the conclusions drawn will be, and the more generalisable the results will be.

Figure 1 illustrates the three-way comparison to be made between the learner English, AmE and BrE corpora.

This comparison can be described as a form of Contrastive Interlanguage Analysis, or CIA. This is a term coined by Granger (1993:43), who states that, unlike classical CA, which systematically compares various levels of the linguistic systems of two or more languages, "CIA does not establish comparisons between two different languages but between native and learner varieties of one and the same language". CIA studies contrast "what non-native and native speakers of a language do in a comparable situation" (Pery-Woodley 1990:43; as quoted in Granger 1998:12), and identify the characteristics of a particular interlanguage.