Research in English and Applied Linguistics
REAL Studies 5

Research in English & Applied Linguistics

Academic Writing in Europe: Empirical Perspectives

edited by
Josef Schmied

Cuvillier Verlag Göttingen
Internationaler wissenschaftlicher Fachverlag

Previous Titles

Vol. 1: English for Central Europe - Interdisciplinary Saxon-Czech Perspectives
Vol. 2: Stress Patterns of Nigerian English
Vol. 3: Complexity and Coherence: Approaches to Linguistic Research and Language Teaching
Vol. 4: English Projects in Teaching and Research in Central Europe

Academic Writing in Europe: Empirical Perspectives

Edited by Josef Schmied

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.
1. Aufl. - Göttingen : Cuvillier, 2011

978-3-86955-739-7

Europäische Union. Europäischer Fonds für regionale Entwicklung: Investition in Ihre Zukunft / Evropská unie. Evropský fond pro regionální rozvoj: Investice do vaší budoucnosti

Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

1. Auflage, 2011
Gedruckt auf säurefreiem Papier

978-3-86955-739-7

Preface

This volume on *Academic Writing in Europe: Empirical Perspectives* is a summary of my own research and teaching experience based on our recent projects and discussions with colleagues from the Czech Republic, Italy and other countries in Europe as well as China, Malaysia and Armenia.

The need for a comparative approach to academic writing has become evident to me during the development of new MA and PhD programmes in the so-called Bologna process, where academic writing components have to be included. This is not only because more and more students even at postgraduate levels seem to lack the skills that have been taken for granted for a long time at European universities or that were considered part of the autonomous efforts of young scholars themselves and not the responsibility of their teachers. This is also because with the further expansion of English as THE language of science and international cooperation during the last few decades, new challenges and opportunities have arisen for English departments and English graduates. On the one hand, there seems to be a standardising trend in international writing that discourages national styles and traditions in specific disciplines and genres that scholars need to be aware of if they want to be successful in international science discourse. On the other hand, new opportunities have arisen that English departments and English graduates can use to prove their "usefulness" in an ever more utilitarian society and view of universities and maybe even sell their "services".

In this light, a comparative view across disciplines, genres and national university traditions is useful. English departments may re-adjust their positions in their universities and societies. Many research traditions in English for Academic Purposes (EAP) and beyond can contribute to this comparative discussion. An empirical perspective may be even more influential than a theoretical one, since it cannot be misinterpreted as ideological, as its conclusions are based on current evidence and by no means fixed, since the serious discussion has just begun in research as well as in teaching. I firmly believe that the comparative research perspective documented and propagated in this volume will be fruitful for individual scholars, their students and their departments. We first have to take stock of what is happening in a community of practice before we can advise others how to participate successfully in the discipline-specific discourse community. This also has the advantage of combining research and teaching in an ideal way even from a student perspective. Students doing research on academic writing will hopefully be more aware of their own writing process and its products and thus successful novices in the academic English community.

Since this volume is an example of culture-specific writing itself, we refrained from "harmonizing" the volume and changing the personal style of individual authors. Unfortunately, it is too small for a comparative study of European writing, but maybe it shows some interesting differences in structure, argumentation and of course idiomaticity that go beyond individual writers. We hope to continue the discussion with more illustrations and statistics in this series.

Many contributions in this volume where first presented and discussed in the section 70 entitled "Empirical Approaches to Discipline, Culture and Identity in Academic Discourse" that I organized together with my colleague Marina Bondi (Modena, IT) for *The European Society for the Study of English* (ESSE) in Torino in August 2010. We hope that these European discussions will continue during the next ESSE conference in 2012 at Boğaziçi, Istanbul.

I wish to thank all my project collaborators in the Czech Republic, in Italy and other parts of Europe, in particular the Sächsisch-Tschechische Hochschul-Zentrum/Kolleg/Initiative (STHZ/K/I) for many years of continuous support and inspiration.

In particular, I wish to thank Susanne Wagner and Christoph Haase, who have contributed towards improving this volume in content and form.

This volume is only a temporary summary to initiate further debate and development of a fascinating topic.

Chemnitz, April 2011

Josef Schmied

Contents

Academic Writing in Europe: a Survey of Approaches and Problems

Josef Schmied

Abstract

This survey sketches the new understanding of academic writing that has developed over the last two decades, from a text-based to a writer- and reader-oriented perspective, from a prescriptive to an empirical discipline. It sets academic writing in a wider context (like English for Specific Purposes and English as a *lingua franca*) and clarifies the main concepts. From a constructionist perspective, a discourse community develops through common practice, using expected schemata for instance in genres like research articles. They can be analysed empirically in corpus- and text-linguistic approaches, where at least five dimensions can be compared in empirical research: genre, academic discipline, national culture, language tradition, and language features. The problems discussed range from fundamental ones (whether a *lingua franca* like English makes non-native users of English in Europe lose national traditions) to practical ones (to what extent the data available are compatible). Despite the problems, new opportunities arise for English departments in Europe when they include an empirical discourse- and genre-based approach in their research and teaching.

1. Introduction: Understanding academic writing

Academic writing has established itself almost as an independent discipline in applied linguistics, or at least as a research-led sub-discipline in English for Academic Purposes (EAP). There is much more to it than what was taught 20 years ago: Old essay-writing focussed on language-specific student errors or creative styles; old English for Specific Purposes (ESP) focussed on discipline-specific vocabulary. The understanding of academic writing has changed fundamentally from a formal text-based perspective to a functional perspective that concentrates on the writer and the writing process and, even more, on the reader and the cognitive construction of discourse in a community (cf. Hyland 2010a, Schmied 2008, Thompson 2001). This paradigm shift applies to teaching as well as to research: Text-oriented research would, for instance, measure syntactic complexity by number of words or clauses per T-unit, or the specificity of lexemes in ontological systems. Writer-oriented research has tried "think aloud protocols" or task observations including keystroke recordings. Reader-oriented research has emphasized the mediation between writers/institutions/cultures, and conventions "describing the stages which help writers to set out their thoughts in ways readers can easily follow and identifying salient features of texts which allow them to engage effectively with their readers" (Hyland 2010b: 194).

2. Key concepts of academic writing

2.1. Definitions of EAP and related terms

In this survey, I see academic writing as an important, if not the most important, part of academic language behaviour in a discourse community. This discourse community uses English for Academic Purposes in research and teaching/learning, not only in universities in native-speaker cultures but also in universities where English is used as an international language or *lingua franca* at levels of international cooperation, where researchers as well as teachers and students are non-native speakers of English.

Traditionally, discussions of language use have been seen as part of ELT (English Language Teaching), or TESOL (Teaching English to Speakers of Other Languages) and TEFL (Teaching English as a Foreign Language). Today these concepts are often seen as a wide field of related terms and acronyms like EAL (English as an Additional Language), EIL (English as an International Language), ELF (English as a Lingua Franca), ESP (English for Special Purposes, or English for Specific Purposes), etc., where overlapping notions are only a matter of perspective. EAP can be seen as the "higher end" of ELF (which, in contrast to "Tourist English", requires at least B2 in the European Framework of Reference for Languages, EFRL). EAP emphasises the common ground of specialised languages in terms of discourse or pragmatics, whereas ESP tends to emphasise the differences in terms of lexicon and idiomaticity. EAP also adds a theoretical framework to practical "writing classes", which have spread to universities in native as well as non-native countries, and which can be seen as part of professional writing in the academic world, just like professional writing in the domains of law (e.g. legal correspondence), journalism (e.g. reportage), engineering (e.g. technical reports), marketing (e.g. advertisements), entertainment (e.g. film scripts), and literature (e.g. "creative writing" of novels).

Within this wide field of EAP, at least three levels of communities can be distinguished, and thus three types of EAP defined:

- Student English: The academic 'novice' may come from an Anglophone background where English is used for a variety of intra-national functions including teaching at secondary schools. Still, academic writing requires additional training, for it necessitates the independent search for appropriate information, its critical evaluation and media-specific presentation. The traditional genre at this level is the academic essay of 2,000 to 5,000 words (occasionally also a corresponding media-supported oral presentation).
- Doctoral English: In contrast to student writing with its focus on digesting research by others, doctoral students have to develop their own ideas, to pursue their own research agenda and to write up everything in a major contribution, which is the result of some sophisticated innovative Ph.D. project that the writer takes a long time to accomplish.

- (International) Research English: Although the written exchange of research results has a long tradition (in Britain at least since the *Philosophical Transactions of the Royal Society* in the 17th century), the importance of international scholarly articles has increased enormously over the last decades, partly due to the increasing competition among universities and researchers and partly due to the new electronic media. This has led to the standardization of peer-review procedures and the corresponding discussion of subject- and genre-specific conventions.

In contrast to student English, the latter two categories, doctoral and research English, are more specialized in the sense that they (have to) follow more subject-specific conventions. This applies to individual research journals as well as whole research communities, e.g. in literary or social-science academic cultures (with their MLA and the ASA/APA conventions, respectively). Such conventions – together with the specialised terminology and argumentation procedures – have made (even sub-discipline-specific) "specialised" academic writing increasingly an in-group phenomenon. To balance this trend, a new EAP category has gained more and more importance: non-specialised writing for a general academic readership, which can be called "popular" academic writing or Popular Academic English. This has political implications, since societies demand increasingly to be informed about public investment in universities and other research institutions.

2.2. Academic writing in the discourse community

Since I emphasize that the key concepts of academic writing have to be made accessible to students, I will adopt a student perspective in this section. I will use entries in Wikipedia (just like many students do) as a starting point and scrutinize them from a perspective of knowledge transfer to see whether there are any major discrepancies between the popular academic representations of these concepts and my more specific academic conceptualisations. The Wikipedia entry for discourse community is quite specific and very suitable for our purposes – not surprisingly since it is based explicitly on Swales (1990):

A discourse community:
1. has a broadly agreed set of common public goals.
2. has mechanisms of intercommunication among its members.
3. uses its participatory mechanisms primarily to provide information and feedback.
4. utilizes and hence possesses one or more genres in the communicative furtherance of its aims.
5. in addition to owning genres, it has acquired some specific lexis.
6. has a threshold level of members with a suitable degree of relevant content and discoursal expertise.

http://en.wikipedia.org/wiki/Discourse_community (27/03/11)

The advantage of this entry is that it is very broad, but it also fits our concept of academic community very well, especially the emphasis on genres and lexis. The levels I have defined according to practice and expertise as student, doctoral, and research English above, each with specific genres and lexical complexity. The level-specific genres are constructed through university conventions and this construction is in line with current thinking on wider academic perspectives.

Over the last two decades, academic writing theory has been closely associated with social constructionism, and again we can use a well-founded Wikipedia entry as a starting point:

> A major focus of social constructionism is to uncover the ways in which individuals and groups participate in the creation of their perceived social reality. It involves looking at the ways social phenomena are created, institutionalized, and made into tradition by humans. Socially constructed reality is seen as an ongoing, dynamic process; reality is reproduced by people acting on their interpretations and their knowledge of it.
> http://en.wikipedia.org/wiki/Social_constructionism (29/03/11)

The two concepts discourse community and social constructionism in higher education can be combined in the concept of an academic community of practice:

> A **community of practice** (**CoP**) is, according to cognitive anthropologists Jean Lave and Etienne Wenger, a group of people who share an interest, a craft, and/or a profession. The group can evolve naturally because of the members' common interest in a particular domain or area, or it can be created specifically with the goal of gaining knowledge related to their field. It is through the process of sharing information and experiences with the group that the members learn from each other, and have an opportunity to develop themselves personally and professionally (Lave & Wenger 1991). CoPs can exist online, such as within discussion boards and newsgroups, or in real life, such as in a lunch room at work, in a field setting, on a factory floor, or elsewhere in the environment.
> http://en.wikipedia.org/wiki/Community_of_practise (29/03/11)

Again, this Wikipedia entry is useful since our academic community is constructed through "sharing information and experiences", like (sub-)discipline-specific conferences. Nowadays, the "written discussion" in scientific disciplines takes mainly place in academic journals or even on pre-publication servers, since the international academic discourse is accelerated enormously.

Although conference papers and journals are the central spoken and written genres in academic communities today, there are many others. The Wikipedia entry for genre gives a crisp summary:

> A text's genre may be determined by its:
> 1. Linguistic function.
> 2. Formal traits.
> 3. Textual organization.

4. Relation of communicative situation to formal and organizational traits of the text (Charaudeau & Maingueneau, 2002: 278-280).

http://en.wikipedia.org/wiki/Genre (27/03/11)

This is a good introduction, but for our purposes not explicit and detailed enough, especially since research over the last 20 years has provided us with so many insights into this central concept.

This exercise has shown that Wikipedia can be used as a resource in general (popular) academic discourse to introduce novices to the basic concepts of a field. Of course, this is not the case for all concepts we need for a scholarly discussion of academic writing. The Wikipedia entry for "Academic Writing" itself is little more than a number of lists of genres or key terms that does not really help the novice in the field.

2.3. Genres as expected schemata in communities of practice

Although genres are recognised subcategories of research discourse (Swales 2004), developing an awareness of their conventions is often difficult for students, since genres are abstractions of real texts, and students need to gain experience through repeated exposure. From a cognitive perspective, genres are schemata that help us engage actively in text comprehension as we develop a feeling for relating new information to existing knowledge and previous community discourse. We recognise prototypical genres as unmarked – and some novices' texts as unintentionally marked in community discourse, which may distract the reader from the intended message of the writer. Thus genres link users to their discourse community and they link texts to each other since real academic discourse is a constant development of intertextuality. For students, this means that they have to learn to select texts for their argumentation from the existing literature, digest them by integrating them into their own writing and continue the academic discourse by "spinning on the yarn".

But genres also activate situational contexts in academic discourse and help create the role of individual community members in the discourse. The students' task is then to be aware of the conventions involved in a project proposal or a BA thesis in their specialisation. Genres constitute the discipline as they form a network with "neighbouring" genres. This community of practice forms a network of members, who move "up" from novices to experts in their discipline through producing the expected situated texts in the different types of genres.

There is no conclusive and comprehensive list of academic genres and there is considerable overlap between the subgenres of academic books: introductory, textbook, research monograph, (research) article collection, handbook, encyclopaedia, etc. And even spoken and written subgenres may be related: a conference paper and the related article collection, the key-note (lecture) and the related handbook article, etc.

Thus genres are fuzzy concepts, but they are useful for empirical analyses of stratified data-bases and related interpretations as well as for teaching. The

advantages of genre-based instruction have been described persuasively by Hyland (2004: 10f):

> The main advantages can be summarized as follows. Genre teaching is:
>
> **Explicit.** Makes clear what is to be learned to facilitate the acquisition of writing skills
> **Systematic.** Provides a coherent framework for focusing on both language and contexts
> **Needs-based.** Ensures that course objectives and content are derived from student needs
> **Supportive.** Gives teachers a central role in scaffolding student learning and creativity
> **Empowering.** Provides access to the patterns and possibilities of variation in valued texts
> **Critical.** Provides the resources for students to understand and challenge valued discourses
> **Consciousness raising.** Increases teacher awareness of texts to confidently advise students on their writing.

These features make the concept of genre accessible to students and useful, since it allows them to meet the expectations of teachers, editors, and gatekeepers of all types in the academic community. Although it is not a formal checklist, it provides students and teachers with a frame they can use for self-study and for teaching.

3. Approaches to academic writing

3.1. Corpus- and text-linguistic approaches

Students and researchers who intend to study academic writing can choose from a wide range of approaches. Basically, I would like to distinguish between approaches that focus on central formal or functional features across texts, usually in stratified collections of academic texts (that is why I call them corpus-linguistic), and approaches that focus on the special or prototypical interplay of features in texts or text-types (that is why I call them text-linguistic). Of course, ideally both approaches overlap and a combination will provide us with the best insights into this complex phenomenon.

Corpus-linguistic approaches are the standard approach in this volume. This is partly due to the research networks in which this collection has been put together. However it also seems to be the prominent approach of our time, since more corpus collections and tools like AntConc give every researcher quickly a key-word-in-context and statistical survey as a starting point for thought and discussion. Even academic novices at BA level for instance achieve a satisfactory scholarly result. More difficult is the development of a simple formal and functional feature analysis into a factor analysis of multiple dimensions (often called Biber-type, since it has been used extensively by Biber, from Biber 1988 through Biber 2006 to Biber & Conrad 2009).

Examples of text-linguistic approaches can be found throughout the history of the analysis of academic writing. Halliday (1997/2004), for instance, uses different text types ranging from a *Microbiology* textbook to a *New Scientist* article to illustrate answers to the big question “how does the language of science

reconstrue human experience?" (ibid: 49). This may be too difficult for a student discussion and we rather illustrate the text-linguistic approach by discussing examples of student writing from the ChemCorpus (s. below).

The best top-down approaches in text-linguistics would be to use a text-processing system to show the systematic parallel structure of headlines or to devise a hyper-text system to allow the reader to follow the links (cf. Schmied 2005). For our purposes two small case studies may suffice to illustrate the holistic approach.

The first text-linguistic example is a distribution diagram of *may* in some (short) exam texts from the ChemCorpus (Figure 1).

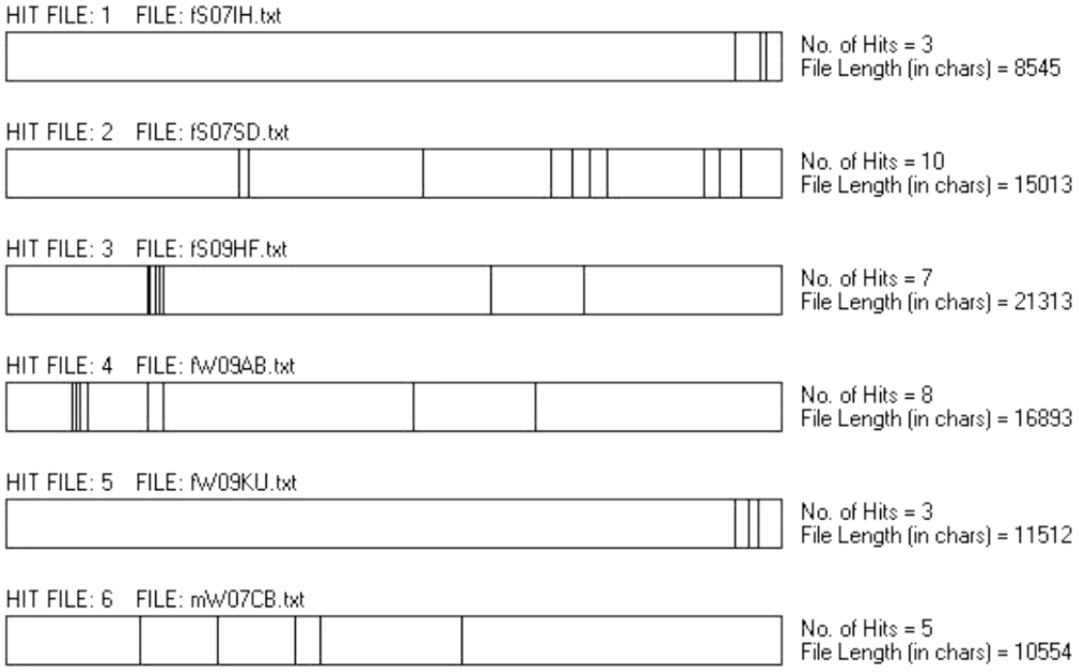


Figure 1: Distribution of *may* in selected ChemCorpus exam texts

The AntConc diagram clearly shows that modal auxiliaries like *may* cluster in specific parts of the text, they can be expected mainly in the second and the last sections, when the secondary literature is discussed critically and when research results are evaluated tentatively.

When we read through the *may* usages in the following examples in context, it is clear that they are all used in epistemic function:

(1) This ***may*** lead to temporary or even permanent language attrition (fS07IH)
(2) These additions ***may*** take several forms (fS09HF)
(3) Firstly, they ***may*** be words that are completely new to English, words that (fS09HF)
(4) Secondly, they ***may*** be words new to the BrE variety, but (fS09HF)
(5) Thirdly, they ***may*** be words that have currency in BrE, but, in Australia (fS09HF)

(6) Finally, they ***may*** also be words that might be unfamiliar to speakers of Standard English (fS09HF)
(7) For some people it ***may*** seem that the Australian language travels from its roots, but (fS09HF)
(8) Many Australian[s] ***may*** be able to give a few examples, including (fS09HF)
(9) In return, children ***may*** also address and refer to their parents differently (fW09AB)
(10) At home, a child ***may*** say "mum", "mom" or "mummy" (fW09AB)
(11) On the phone with a friend listening you ***may*** address your mother by saying "mother" (fW09AB)
(12) When you are at home, talking to another family member and your mother is not present you ***may*** refer to her as "our mom" or "the old lady" (fW09AB)
(13) He ***may*** have coursed [caused] offence talking this way (fW09AB)
(14) In a meeting, for instance, where first names are generally used, the director ***may*** say (fW09AB)
(15) In everyday interactions, speech differences ***may*** also be reflected in people's social networks (fW090AB)
(16) To British ears a New Zealander's "bad" ***may*** sound like "bed" (fW09AB)
(17) the striking usage of 'be' which ***may*** support the creole hypothesis (fW09KU)
(18) Therefore, a middle way between these hypothesises ***may*** be advisable (fW09KU)
(19) I would conclude that AAVE ***may*** be an africanized language, which means (fW09KU)
(20) On the other hand it ***may*** be assumed that the text producer has a supertheme in mind (mW07CB)
(21) They ***may*** also be seen as linguistic principles (mW07CB)
(22) Cohesion ***may*** well be viewed as a phenomenon of surface structure, i.e. (mW07CB)
(23) This ***may*** be done via anaphora or cataphora (mW07CB)
(24) A specific seme ***may*** occur throughout the text (mW07CB)

All *mays* in this list are used to indicate tentative expressions, but even within this semantic space of cautious meta-discourse, we recognise a few patterns:

- *may* is serialised to list possibilities (*Firstly* to *Finally* in (3) – (6) in fS09HF),
- *may* collocates with verbs of thinking/seeing (*assumed/seen/viewed* in mW07CB),
- *may* is used with primary auxiliaries, especially *be* to express passive,
- *may* often precedes *also* (four times by three different students), which could be an advanced learner habit.

The second text-linguistic example is the introductory paragraph of a final exam essay (on a language variation project) to illustrate how inherent lexical structures could be made more explicit. First, we discuss the original text, then we construct a new text version that is explicit and systematic according to our principles:

> Language never stands still. It varies over time and is in a constant process of change, even though these processes might not be obvious. In many cases, linguistic variation is a result of internal linguistic factors. However, since the middle of the 20th century, many studies on language change have acknowledged that extralinguistic factors might have a considerable impact on the innovation and spread of new linguistic variants. Although these so-called sociolinguistic studies focused in the beginning predominantly on the category of social class, the awareness grew that it might in fact be the correlation of a variety of extralinguistic factors, such as age, ethnicity, and gender, that served best to explain the mechanism of linguistic variation and change. The subsequent chapters will focus in particular on the category of gender, as according to Labov (1990), the findings concerning the linguistic differentiation between men and women belong to the clearest and most consistent results of sociolinguistic research in the speech community. The first chapter will consider a number of methods and approaches that are commonly used to carry out sociolinguistic studies and obtain reliable results. As the argumentation will specifically deal with Great Britain, the correlation of social class and gender will need to be considered in particular. The second chapter will then discuss several generalizations that were made concerning gender-specific differences. The last part will finally present a number of phonological and grammatical variables that might be included in the study to analyse and reveal gender-based differences in linguistic variation. (fS10SK)

Apart from a "philosophical-essayistic" beginning, the text (fS10SK) consists of two parts: first, a discussion contrasting two approaches to language variation and change, and then a list of the sections that sketch the outline of this exam paper. The first part contrasts (through *however*) the traditional 19th century diachronic approach to language change with the modern sociolinguistic one since the mid-20th century. A further contrast is established between the old intra-linguistic and the new extra-linguistic factors, the latter expanding from social class to other variables like age, ethnicity, and gender. These contrasts can be made much more explicit in the re-written version (S10SK2) through the different type of contrasts: *intra-* vs. *extra-linguistic*, *19th* vs. *20th century*, *however* and *although* and the implicit *initially* vs. *grew*, as marked in the text below. Such lexical patterns in texts can be supplemented by grammatical patterns (like the *will* constructions above).

The second paragraph of the re-written text below is more clearly structured through lexical repetition of *section* and the near-synonym *part*. However, the contrast of the topics in the three sections (*methods and approaches*, *generalisations*, and *variables*) is not as clear as it could be. Most other changes in this introduction are simply structural simplifications that help the reader process the text more easily (*concerning* as a preposition). The reduction of tentativeness (i.e. auxiliaries, esp. *might*, the most "careful") may also be a point that has to be considered systematically at this advanced level of academic writing, since novice writers have to learn to develop their own stance. Such key concepts or guidelines can be deduced from this text example by the students themselves, so that (hopefully) they will be able to transfer their knowledge to similar texts later.

Language variation and change have been studied as a result of *intra-linguistic* factors since the *19th century*. Since the middle of the *20th century*, **however**, studies on language change have shown that *extra-linguistic* factors can have considerable impact on the innovation and spread of new linguistic variants. **Although** these sociolinguistic studies focused initially on the category of social class, the awareness grew that a combination of extra-linguistic factors, such as age, ethnicity and gender, may explain best the mechanism of language variation and change.
The subsequent *sections* focus in particular on the category of gender, as according to Labov (1990), the difference between men and women belongs to the clearest and most consistent results of sociolinguistic research in the speech community. The *first section* considers methods and approaches commonly used in sociolinguistic studies. As the project is situated in Great Britain, the correlation of social class and gender is focussed on first. The *second section* then discusses several generalizations on gender-specific differences. The *last part* presents phonological and grammatical variables that might be included in the study of gender-based differences in linguistic variation. (S10KS2)

The explicit use of cohesive devices is certainly a feature of advanced writers of English. This is partly teaching-induced, since students are told explicitly to pay attention to cohesive devices in their texts in today's writing classes, and this may thus change when students have learnt how to use less explicit devices at a more advanced level (approaching native-speaker level C2 in the EFRL). Of course, linguistics students who have worked through Halliday/Hasan's cohesive devices are more aware of the options and may thus tend to use them more often for a period of time in their writing.

The comparison of selected concessive and contrastive markers by Wagner (this volume) is an interesting case study that tests the usefulness of different types of data-bases for comparative research, including the ChemCorpus database.

3.2. Dimensions of linguistic analysis

When comparing databases for analyses of academic writing, we can distinguish at least five dimensions of factors (cf. Yakhontova 2006):

- Genre seems to be the dominant dimension in modern comparative research, and research articles the master genre (cf. Bondi 1999 and Hyland 2010a: 117).
- Academic discipline is the most hotly debated "cultural" component, because the different "cultures" of "humanities" and "sciences" have been discussed (e.g. Hyland & Bondi eds. 2006) since C.P. Snow's catch phrase of the "two cultures" (1959).
- National culture seems to be less prominent in the discussions now than during the contrastive period of linguistics, when English vs. German vs. Spanish texts were analysed. However, with English as a *lingua franca*, the issue is now whether German and other academic cultures can be expressed in English for several reader perspectives, the native German and the native English, and maybe others.

- Contrastive language cultures are less prominent today, but still important at least at the lowest levels of academic writing, when the International Corpus of Learner English (ICLE) provides a database of argumentative essays – irrespective of whether an "essay" is a realistic natural genre of academic writing.
- The language phenomena analysed have tended towards metalanguage in the last few years, focusing on interpersonal devices of proximity (cf. 4.2. below) from pronouns to modal adverbs.

The contributions to this volume (in alphabetical order in Table 1) are fairly representative of the current research focus in these five dimensions:

The dominant genre is clearly the research article. It is however interesting to see rare genres (like conference posters and course descriptions for students) analysed, since these are relatively new and thus allow us to see the establishment of formal conventions on the basis of functional needs.

The academic disciplines chosen for analysis show the usual contrasts between "hard" and "soft", with natural sciences usually assumed to be more standardised already. But it is by no means clear whether all disciplines will follow the same trends, since the diversity and artistic creativity in humanities may also be less suitable for standardisation and thus follow the trends less rigorously.

	genre	academic discipline	national "culture"	language phenomenon
D'Angelo	conference posters	physics, law	———	textual/ metadiscourse, semiotics
Diani	university lectures	linguistics, psychology, economics	English vs. Italian	person markers
Gesuato	course descriptions	biology, geography, history, journalism. law, literature, music, physics, psychology, statistics,	———	modality, tenses, lexicon, etc.
Haase	research articles	physics	———	modal adverbs/ auxiliaries
Hůlková	research articles	management, politics, sociology, adult education, psychology	———	conjunctive adverbs
Maláskova	research articles	humanities vs. social sciences	———	hedges
Provenzano	textbooks	finance	Western vs. Islamic	lexical, semantic
Wagner	student papers/theses	linguistics vs. methodology/cultural studies/literature	German vs. Czech	concessives/ contrastives

Table 1: Dimensions of linguistic analysis in the contributions in this volume

Similarly debated is the assumption that all national cultures will follow the Anglo-American model (cf. 4.1), so that even a comparison of "national" cultures of writing may be difficult, if it is not strictly tied to language properties, like etymological or typological contrasts between Germanic and Romance languages (thus the dimensions of national culture and languages are collapsed in Table 1).

The hotly-debated native-speaker issue is noticeably absent from this volume, probably because all contributions are written by non-native speakers and because it is difficult to find a suitably broad data-base that covers this aspect in addition to the discipline and level (but Wagner indicates the direction).

Finally, the language phenomena discussed seem to focus on those features that English traditionally has used for safe-guarding a good writer – reader interaction, whereas the analyses of argumentative structures are not easily comparable. The focus on metadiscourse is obvious in most contributions in this volume, but it is viewed from very different perspectives: from a functional perspective metadiscourse includes hedges (as in Maláskova), from a formal perspective modal/conjunctive adverbs (as in Haase and Hůlková) and modal auxiliaries (as in Haase and Gesuato); from a writer – reader discourse perspective it includes personal markers, like *I* and *we* (as in Diani), and of course, the "rare genres" studied by D'Angelo and Gesuato require specific variables, whereas the former concentrates on semiotics and multimodality (the text and image interface) the latter also covers grammatical (tenses) and lexical aspects (e.g. preferred key words from the academic word list). The culture-specific lexicon is the special focus in Provenzano's contribution. Thus the language features analysed in this volume give at least an impression of the many aspects that are still worth exploring systematically in this "dimension".

4. Problems of comparative research in academic writing

Despite the wide choice of approaches, some even easily accessible to novices, comparative research in academic writing has its problems. These challenges can however also be seen as opportunities that allow researchers to work on specific issues in this wide field.

4.1. Culture-specific traditions

Several authors have pointed out the "cultural baggage" of English (Wierzbicka 2006) in its lexicon (like *reasonable* or *fair*) and grammar (in causatives and epistemic adverbs), which makes English questionable as a neutral *lingua franca* for non-native speakers. It has also been argued (Thielmann 2009) that academic culture is linked closely to academic language and provocatively even that "one cannot do science in a pidgin". Detailed case studies can help clarify typological and pragmatic differences between English and, for instance, German, so that we can scrutinise academic argumentation styles (ibid.) on various levels of discourse:

Can we generalise that German style is more argumentative whereas English style is more persuasive? Can we prove that *weil* is used to focus on the writer's decisive argument that convinces the writer and, hopefully, the reader, whereas English *because* often refers to the argumentation by other writers? Is this "persuasive" reader orientation in English always positive or can it also be seen as negative, since this subjectification distracts from the force of the argument? Or should we distinguish between a preference for authority- and a preference for discourse-orientation in academic communities? Can we say that English has more lexical diversity that allows the writer to argue more subtly when using *because*, *since*, *as* instead of a simple *weil* or is this beyond the non-native writer already?

On top of all these multilingual/-cultural aspects, writers have to be aware of intercultural implications. The whole field of contrastive rhetoric and culture needs much more empirical research and rigorous conceptualisation (cf. Atkinson 2004).

Of course, the level of English in non-native academic writing and in non-native writing courses needs to be discussed. Unfortunately, many EAP classes still deal with prepositions and their use in idiomatic expressions and tense/aspect "rules" instead of concentrating on the specific features of academic language and academic English. Thus they never manage to address the following real issues:

Can we really criticize complex cognitive concepts like hedging and cohesion or argumentation structure below a proficiency level of C1 (in the EFRL)? Can we really write according to German academic traditions in English? Or is German academic style also becoming more reader-oriented due to American influence or due to changes in the staff-student relationship? Can a 'dual publication' be a compromise, as suggested by Hamp-Lyons (2011: 2) "that researchers who publish their work in languages other than English should be free to re-publish the same work in English translation, with full attribution to the original publication"?

Since academic knowledge creation is set in a specific learning situation, it also requires socio-cultural knowledge of the discourse community (like interpersonal conventions between writers and readers in terms of power-relationships and associated roles of formality, authority, intimacy, and others). Thus the conventions of English and American, German and French academic circles have developed differently in the national academic cultures.

Only in a comparative perspective can we decide whether current trends perceived by writers like Hyland (2005: 173f) in the Anglo-American tradition are universal.

> Over the past decade or so, academic writing has gradually lost its traditional tag as an objective, faceless and impersonal form of discourse and come to be seen as a persuasive endeavour involving interaction between writers and readers. This view sees academics as not simply producing texts that plausibly represent an external reality, but also as using language to acknowledge, construct and negotiate social relations. Writers seek to offer a credible representation of themselves and their work by claiming solidarity with readers, evaluating their material and acknowledging alternative views, so that controlling the level of personality in a text becomes central to building a convincing argument. Put succinctly, every successful academic text displays the writer's awareness of both its readers and its consequences.

If there is such a universal trend in academic writing, non-native writers may follow this trend in their mother-tongue or only in their English. More specifically, the interesting question here is whether academic texts by German or Italian students of English approximate the British and American models from their native German or Italian traditions or whether the non-native writings are closer to each other than to the source- and target-culture conventions. This idea is visualised in Figure 2, where the *lingua franca* Englishes are seen like inter-languages closer together than the source texts (German and Italian, for example). The arrows indicate that non-native speakers approximate Anglo-American standards, that there is considerable overlap and that the variation between source- and target-language is increasing, although the two-dimensional nature of the diagram may be misleading (it does not suggest that German writers approximate towards British and Italian towards American models). In fact, it could be argued that the distinction between British and American English is overemphasised in view of the increasing Americanisation world-wide and the growing acceptability of other target models from Australia, Canada, etc. The adaptation in other complex cultural writing spaces is a completely different discourse (see Nkemleke fc. on Cameroon).

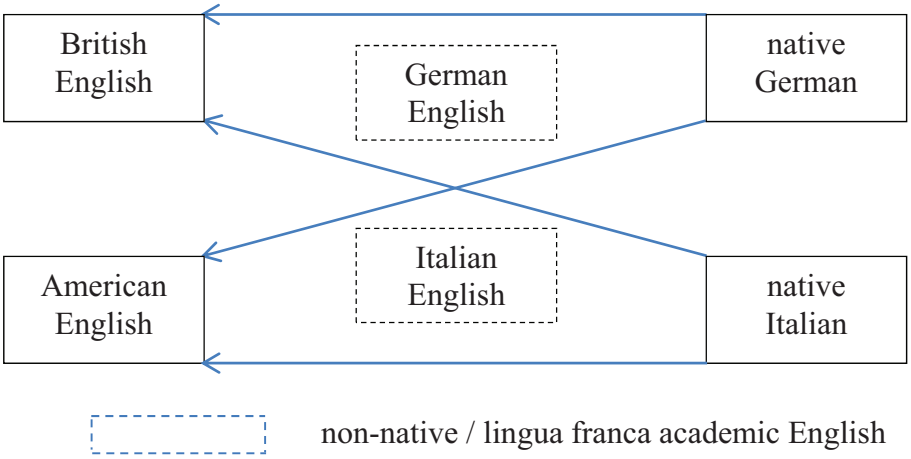


Figure 2: Approximation model for German and Italian academic writing styles towards native-speaker models

The main distinction in this approximation model is between native German, Italian, French, Czech, Chinese styles and the Anglo-American model. This does not imply that complete approximation is necessary or that a "neutral" *lingua franca* style cannot be accepted by non-native as well as native speakers of English.

Of course, the cultural influence of Anglo-American universities and Anglo-American publishing houses is dominant, but it does not mean that European publishes (like Elsevier, Benjamins or Mouton de Gruyter) or European editors as gate-keepers are not possible. If we take the concept of a (neutral, explicit, systematic) academic *lingua franca* seriously, this may even mean some training for native speakers of English, since this non-natural academic writing is more conscious and controlled, more reader- and culture-specific than usual, and it is at a high level, i.e. it does not violate basic grammar rules but rather tolerates less

stereotypical usages in lexicon, idiomaticity, metaphor, argumentation structure and other formal and functional conventions.

Since it is difficult to distinguish between Anglo-American and general style trends from the text- through the writer- to the reader perspective (see above), it may be worth summarizing explicit guidelines in current style guides or accepted university textbooks. Thus it has been argued (Swales & Feak 2000: 16) that "national" or cultural features of Anglo-American writing are:

1. be more explicit about its structure and purposes
2. be less tolerant of asides or digressions
3. use fairly short sentences with less complicated grammar
4. have stricter conventions for subsections and their titles
5. be more loaded with citations
6. rely more on recent citations
7. have longer paragraphs in terms of number of words
8. point more explicitly to "gaps" or "weaknesses" in the previous research
9. use more sentence connectors (words like *however*)
10. place the responsibility for clarity and understanding on the writer rather than on the reader

The empirical analysis of academic writing as exemplified in this collection will give us a good database for discussing acceptable usage practices by different academic communities in future.

4.2. Proximity as a reader-related text concept

A concept that has emerged as central in reader-oriented academic writing during the past few years is proximity (or approximation, when transferred from media studies). A very wide definition has been suggested recently by Hyland (2010a: 117):

> I use the term proximity here to refer to a writer's control of rhetorical features which display both authority as an expert and a personal position towards issues in an unfolding text. It involves responding to the context of the text, particularly the readers who form part of that context, textually constructing both the writer and the reader as people with similar understandings and goals. While it embraces the notion of interpersonality, proximity is a slightly wider idea as it not only includes how writers manage themselves and their interactions with others, but also the ways ideational material, what the text is 'about', is presented for a particular audience. It is concerned with how writers represent not only themselves and their readers, but also their material, in ways which are most likely to meet their readers' expectations.
>
> So proximity captures 2 key aspects of acting interpersonally. The first refers to what might be called the proximity of membership: How academic writers demonstrate their authority to colleagues through use of disciplinary conventions. What does the writer do to position him or herself as a disciplinary expert and competent colleague? The second concerns the proximity of commitment, or how the writer takes a personal position towards issues in an unfolding text. That is, what does the writer do to locate him or herself in relation to the material presented? One points to how we position ourselves in relation to our communities, and the other to how we position ourselves in relation to our text.

This covers almost all aspects of metadiscourse. If this is considered too wide, it could be restricted to "recipient-design", i.e. "an orientation and sensitivity to the particular others who are our co-participants through lexical choice, topic selection, conventions, of argument, and so on" (ibid.).

This concept is particularly useful in comparing specialised to popular academic writing, which is useful for EAP in journalistic writing:

> Science journalism illustrates the ways proximity (and interpersonality) work as writers set out material for different purposes and readers. Popularizations represent a discourse which establishes the uniqueness, relevance and immediacy of topics which might not seem to warrant lay attention by making information concrete, novel and accessible. Findings are therefore invested with a factual status, related to real life concerns, and presented as germane to readers with little detailed interest in the ways that they were arrived at or in the controversies surrounding them.
>
> Readers, in fact, experience the academic world and its discourses as a succession of discoveries in the relentless advance of inductive science. In sum, science journalism works as journalism rather than science. It is written in ways which make the research accessible and allow non-specialists to recover the interpretive voice of the scientist. (ibid: 126)

The popularisation of science has been a debated issue in English-speaking cultures for a long time. There have been many small studies to compare specialised and popular texts on the same topic, but it is not easy to ensure compatibility. The most common way is to trace back the origins of newspaper articles or science journals like *New Scientist* or *Scientific American* to the original publication in research articles or even on pre-publication servers (cf. Schmied 2009a,b,c). Haase (this volume) is a good example that illustrates the qualitative and quantitative options of research in this thrilling area. In a broader perspective, even films can be included in a multimedia-corpus of science texts and analysed, which may make it particularly attractive for students, but what can reasonably be compared is a real issue.

4.3. Compatibility of data

A major problem of comparative research in academic writing has been the availability of a compatible database. This can again be seen on three levels (cf. the case study by Wagner in this volume):

Academic writing on the web, as in Google Scholar, may be useful for finding usage patterns involving rare linguistic forms but the reliability of the-web-as-corpus in limited. Googlelabs currently includes 5.2 million books published in English between 1800 and 2000, approximately 361 billion words. The usefulness of this tool for historical comparisons (even of rare collocates) is demonstrated by Haase (this volume).

Taking the academic sections of national standard reference corpora, like the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA), is the better option, which has also been demonstrated by Mark Davies (n.d.) and by Wagner (this volume). The academic component is about 83 million words in COCA and 15 million in the BNC.

The third and most suitable and complex option are, of course, purpose-made corpora that are compiled for a specific piece of research and may be usable for another. Although the tradition of 'disposable corpora' has been very successful in translation studies, the small collections of academic writing that students and teachers have on their personal computers rarely add up to a coherent compilation. However, their advantage, that they are personal, should not be underestimated. If students really manage to develop a "detached" perspective from their own work, the combination of practical and theoretical work would be ideal, since they can learn from their own work (and maybe lecturer's corrections) to improve their future writing, esp. in their final theses. For, revising this decisive "masterpiece" of their studies is often most difficult because students are tempted to submit a first draft as a final version, since they are not used to revising a piece of writing thoroughly enough.

From a more scholarly perspective, all the more or less stratified collections of student work may not be ideal in many ways: in many English Departments male students are hard to find, the balance between linguistic, literature, cultural studies and methodology specialists is difficult, and the level of English generally very uneven. However, for some high-frequency language phenomena this may well be enough for a sophisticated analysis (as is shown by Wagner in this volume).

Krishnamurthy & Kosem (2007: 370f) have discussed the usefulness of existing corpora (up to the recently compiled British Academic Written/Spoken English Corpus, BAWE/BASE) for EAP pedagogy and they come to the following conclusion:

> The one thing that EAP seems to lack is a corpus that includes all levels of data—from pre-sessional students' writing and speech to academic lectures, Ph.D. theses, and published research articles and books. Such a corpus would need to include as many disciplines as possible, with sufficient detailed categorisation to enable the users (teachers or students) to select a customised set of corpus texts appropriate for their needs. If new corpora are created to agreed common designs, they could be accessed together, forming a richer and more extensive resource. The resulting corpus would need a user-friendly interface that is specifically designed for pedagogical use rather than for research.

The solution to all these problems is, of course, to compile our own corpus. However, corpus compilation is time-consuming and more resource-consuming than commonly assumed and funding options are limited. The BAWE corpus is the result of one of the few corpus compilation projects funded by a research agency (ESRC). A systematic set-up of a corpus (Table 2) would ideally comprise of enough texts written by male vs. female, linguistic vs. non-linguistic (cultural studies/literature) students, thus it would have 5 texts per category. Again ideally, the compilation would "accompany" students during their studies in regular developmental steps from the entrance examination to their BA and MA thesis, maybe with intermediate steps in texts from term papers in the 2nd year of their BA and again after the first year of their MA programme. If the number of words is also increased systematically, we would end up with a substantial corpus of more than 10 million words, which is not that far from the

15 million academic English in the BNC. This size would make our corpus comparable with the major native-speaker EAL corpora, the Corpus of British Academic Written English (BAWE) and its American model, the Michigan Corpus of Upper Student Papers (MICUSP), although they are far from ideally stratified and compatible either.

ChemCorpus	files	minimum words/text
entrance examination	20	1,000
BA2Year term paper	20	5,000
BA thesis	20	15,000
MA1Year term paper	20	8,000
MA thesis	20	25,000
total	**100**	**10,800,000**

Table 2: The ideal ChemCorpus of academic writing

Of course, such a collection would allow us to compare the students' developmental stages through specific writing classes and also before and after their year abroad (which is compulsory in the BA programme in the fifth semester at Chemnitz). Over the years, we could even take texts by the same students (from their "European Language *Portfolio*" in the CEFRL). Related research hypotheses to pursue would be, for instance, that advanced students of English move from overuse to appropriate use for specifically English features (in the case of continuous forms), from more explicit to more implicit marking (in the case of cohesion markers), from more extreme to more tentative forms (in the case of modal auxiliaries). If similar EAP corpora could be complied at other universities and even international partners (like Chemnitz and Brno), an interesting comparison would be possible. A major problem is, of course, that the study programmes are not compatible enough (despite "Bologna").

4.4. Applications in teaching

The advantages of genre-based academic writing can be directly derived from the definition: If academic writing constructs the discursive reality of a discipline, effective learning is also a social activity, it is a constant battle to meet (or challenge) expected outcomes in conventional genres.

Of course, effective learning must be needs-oriented, i.e. the first step is to identify students' needs in their academic life and afterwards, which may be partly different target situations (when an argumentative essay is the target in academic life and a presentation of a scientific problem to a general academic public is the target in a job afterwards). Obviously, determining students' needs is a continuous and changing process. Here the teacher is the facilitator who helps the students to achieve their own goals, and learning to write is a social

activity that helps in a social activity (of academic discourse) through the effective use of the tool language.

The following model for a teaching-learning cycle has been proposed (Feez 1998):

- *Setting the context* – revealing genre purposes and the settings in which a genre is commonly used
- *Modeling* – analyzing the genre to reveal its stages and key features
- *Joint construction* guided, teacher-supported practice in the genre
- *Independent construction* – independent writing monitored by the teacher
- *Comparing* – relating what has been learned to other genres and contexts

This cycle can be seen as a scaffold (according to Vygotsky 1978) that empowers students and raises their consciousness to learn cooperatively and independently in increasingly complex ways. So they can move from collecting to comparing texts, from investigating variation within genres and disciplines to differences between them, from discovering formal differences to explaining them through specific functional requirements.

The following checklist of good academic writing has been quoted frequently (e.g. by Hyland 2006: 221), because it seems to be general enough to appeal to teachers and exemplary enough to apply to students:

Texts are explicit, with clear discussion of data and results.

Texts follow an inductive "top-down" pattern, with topic sentences and an introduction to help readers see where the text will lead.

Texts contain metadiscourse, such as *to summarize*, *in conclusion*, *firstly*, *secondly*, etc., to help guide readers through the argument.

Texts are emotionally neutral and strive to appear objective.

Texts contain hedges like *probably* and *might* to avoid sounding too confident.

Texts are intertextual, drawing on other texts for their structure, form, and patterns of argument.

Texts adopt the right tone to show appropriate confidence and modesty.

Texts acknowledge prior work and avoid plagiarism.

Texts comply with the genre requirements of the community or classroom.

Figure 5.8. Feature of "academic writing" (Johns 1997)

Academic writing is gradually establishing itself as a central element in the new BA and MA degree programmes in European universities. The combination of empirical findings, their discussion and their contextualisation in personal and university experience with the help of appropriate teaching and learning models characterises the way forward towards a unified and effective European education space that may make learning and writing more effective for students and young researchers alike.

5. Conclusion

This survey has tried to provide a scholarly foundation for academic writing at European universities. It is based on a new understanding of writing, and academic language in general, as interaction between writer and reader through a text. Non-native academics have always had the problem of finding native speakers to "edit" their texts. This may become less important in the future if non-native standards become acceptable in Europe. The differences to native speakers may be less prominent in discourse pragmatics than in idiomaticity and usage variation and this may be less noticeable than, for examples, Asian discourse cultures. The conscious teaching of genre variation may allow non-native scholars to choose style features consciously and maybe make even native speakers aware of cultural preferences.

Detailed quantitative corpus-linguistic analyses of native speaker vs. non-native speaker writing may show "more than meets the eye". Such a data-based sensibility for writing conventions would make it possible particularly for non-native writers to increase their awareness of academic usages and thus be recognised as serious, committed and still careful researchers by the specialised discourse community.

Similarly, popular academic discourse also needs trained language specialists, for this new and conscious style of academic writing in all its 'translations' from specialised to journalistic discourse.

The specialisation in academic writing may be of particular relevance in countries like Germany where English specialists often do not find a job in state schools, because few teachers are needed due to population (and student) decline in the last few years. Thus language services of the future do not only offer teaching English generally or for specific purposes and translating different types of texts, but also for editing and web-publishing of academic texts. A scientist can write a first draft but it takes a real language specialist to improve it according to the conventions of the discourse community, including an analysis of the sub-discipline or even the specific journal. This consultancy on web publications for the different academic communities outlined here would be a new challenge but also a new opportunity.

Josef Schmied
English Language & Linguistics
Chemnitz University of Technology
Germany
josef.schmied@phil.tu-chemnitz.de

References

Atkinson, D. (2004). Contrasting Rhetorics/Contrasting Cultures: Why Contrastive Rhetoric Needs a Better Conceptualization of Culture. *Journal of English for Academic Purposes* 3, 277-289.

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, D. & S. Conrad (2009). *Register, Genre, and Style.* Cambridge: Cambridge University Press.

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Bondi, M. (1999). *English across Genres: Language Variation in the Discourse of Economics*. Modena: Il Fiorino.

Davies, M. (n.d.). The Corpus of Historical American English and Google Books / Culturomics. http://corpus.byu.edu/coha/compare-googleBooks.asp (18/04/11).

Feez, S. (1998). *Text-Based Syllabus Design*. Sydney: Macquarie University/AMES.

Halliday, M.A.K. (1997/2004). On the Grammar of Scientific English. In: C. Torsello (ed.), *Grammatica: Studi Interlinguistici*. Padova: Unipress, 21-38. Repr. 2004. in M.A.K. Halliday, *The Language of Science*. London: Continuum, 181-198.

Hamp-Lyons, L. (2011). Editorial: English for Academic Purposes: 2011 and Beyond. *Journal of English for Academic Purposes* 10, 2-4.

Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow: Longman.

Hyland, K. (2004). *Genre and Second Language Writing*. Ann Arbor: University of Michigan Press.

Hyland, K. (2005). Stance and Engagement: a Model of Interaction in Academic Discourse. *Discourse Studies* 7, 173-192.

Hyland, K. (2006). *English for Academic Purposes: an Advanced Resource Book*. London: Routledge.

Hyland, K. (2009). *Academic Discourse. English in a Global Context*. London: Continuum.

Hyland, K. (2010a). Constructing Proximity: Relating to Readers in Popular and Professional Science. *Journal of English for Academic Purposes* 9, 116-127.

Hyland, K. (2010b). Researching Writing. In: B. Paltridge & A. Phakiti (eds.) *Continuum Companion to Second Language Research Methods*. London: Continuum, 191-204.

Hyland, K. & M. Bondi (eds. 2006). *Academic Discourse across Disciplines*. Bern: Peter Lang.

Johns, A. M. (1997). *Text, Role and Context: Developing Academic Literacies*. Cambridge: Cambridge University Press.

Krishnamurthy, R. & I. Kosem (2007). Issues in Creating a Corpus for EAP Pedagogy and Research. *Journal of English for Academic Purposes* 6, 356-373.

Nkemleke, D. (fc.). *Exploring Academic Writing in Cameroon English*. Göttingen: Cuvillier.

Schmied, J. (2005). Hypertext and (Grammar) Learning. In: L. Mössner (ed.), *Anglistentag 2004 Aachen. Proceedings*. Trier: Wissenschaftlicher Verlag, 337-357.

Schmied, J. (2008). Academic Interaction: Adapting Complexity and Coherence to the Readership. *Topics in Linguistics 2. Politeness and Interaction.* Nitra: Constantine the Philosopher University, Faculty of Arts, 12-17.

Schmied, J. (2009a). Discourse Approaches to Specialised and Popular Academic English: Analysing Adverbial Clause Connections. In: O. Dontcheva-Navratilova & R. Povolná (eds.), *Coherence and Cohesion in Spoken and Written Discourse.* Newcastle upon Tyne: Cambridge Scholars, 167-181.

Schmied, J. (2009b). Hedges in Specialised vs. Popular Academic Interaction: A Case Study of Medical Texts. *Discourse and Interaction 1.2.,* Brno: Masaryk University, 85-98.

Schmied, J. (2009c). Discourse Approaches to Specialised and Popular Academic English: Comparing Stance, Hedges and Modality. In: M. Dynel (ed.), *Advances in Discourse Approaches.* Newcastle upon Tyne: Cambridge Scholars, 76-91.

Swales, J. (1990). *Genre Analysis.* Cambridge: Cambridge University Press.
Swales, J. (2004). *Research Genres: Exploration and Applications*. Cambridge: Cambridge University Press.
Swales, J. & C. Feak (2000). *English in Today's Research World: a Writer's Guide*. Michigan: Michigan University Press.
Thielmann, W. (2009). *Deutsche und englische Wissenschaftssprache im Vergleich. Hinführen - Verknüpfen - Benennen.* Heidelberg: Synchron Wissenschaftsverlag der Autoren.
Thompson, G. (2001). Interaction in Academic Writing: Learning to Argue with the Reader. *Applied Linguistics* 22, 58-78.
Vygotsky, L.S. (1978). *Mind and Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
Wierzbicka, A. (2006). *English: Meaning and Culture*. Oxford: Oxford University Press.
Yakhontova, T. (2006). Cultural and Disciplinary Variation in Academic Discourse: The Issue of Influencing Factors. *Journal of English for Academic Purposes* 5, 153-167.

Concessives and Contrastives in Student Writing: L1, L2 and Genre Differences

Susanne Wagner

Abstract

This study investigates the use of concessive markers in academic writing. Frequency differences between academic sub-genres can be established on the basis of large reference corpora such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). These are put in perspective through comparisons with specialised corpora of academic writing (Corpus of British Academic Written English, BAWE and the Michigan Corpus of Upper Student Papers, MICUSP). A learner component is introduced by analysing four markers, *although, however, nevertheless* and *while*, in student writing. The student data stem from two sources: German natives enrolled in English Studies programmes at Chemnitz University of Technology and Czech natives studying at Brno Masaryk University. Statistical analyses reveal striking significant differences both between native and non-native writing as well as between German and Czech production.

1. Introduction

It is a well-known fact that academic writing is very different from other genres in a number of respects. One major difference concerns the complexity of the language used by the author(s), with 'complexity' a difficult notion in itself (cf. e.g. Schmied 2007). In the present context, complexity is understood as both formally and cognitively complex, thus clearly including the marking of concessive and contrastive relationships (cf. Champaud & Bassano 1994: 415). Consider the examples in (1):

(1) a. **Although** fifteen states have blasphemy statutes, the Supreme Court effectively nullified them in a 1951 case involving the film The Miracle, in which the Court held that … (COCA 2010 Acad Humanist)

b. For many years after the war he refused to visit Germany, **although** he continued to write in German. (COCA 2010 Acad AmerScholar)

c. **However** school faculty may sometimes try to improve civility by using methods shown by research to be ineffective. (COCA 2010 Acad Education)

d. The status-of-forces agreement signed by the United States and Iraq in 2008 does seem to have diminished some of Iran's concerns, **however.** (COCA 2009 Acad ForeignAffairs)

e. **Nevertheless**, Turkish people have adapted to shopping malls. (COCA 2009 Acad Adolescence)

f. While opting for one over the other was inevitable, given that the IRGC owes its allegiance to the Islamic Republic, rather than to the president, Ahmadinejad **nevertheless** took a calculated risk; … (COCA 2010 Acad MidEastQ)

g. **Whereas** the attenuated whites in the matrix are clearly just that – white – zombies and vampires are often more suggestively so. (COCA 2010 Acad AnthropolQ)

h. In this review, I concentrate on the distance approach because it is most appropriate for analyzing the rate of species turnover, **whereas** the raw data approach produces an analysis of community composition that explains the degree of species turnover… (COCA 2010 Acad Bioscience)

The four markers in (1), *although, however, nevertheless* and *while*, will be at the heart of the present study. The examples already indicate one of their common properties: they can occur both sentence/ clause-initially (1a,c,e,g) and intra-sententially (1b,d,f,h). It is indicative of the complexity level of academic writing that the majority of the randomly selected examples are – even in shortened form – longer than two lines.

From a processing perspective, concessives include complex forms of (non-)entailment, which is one of the major reasons for their cognitive complexity. Following König (1985, 1986, 1988), Vergaro (2008a: 99) posits that the underlying conceptual structure of a concessive can be said to consist

> in the assertion of two situations (facts) against the background of a certain conflict or incompatibility. […] Even though P, Q. And the underlying relation is If P [protasis], then not-Q [apodosis], where P and Q represent prototypical states of affairs or, more precisely, states of affairs that are prototypically in contrast.

Quirk et al. (1987: 1098) state the following on concessive meaning in general:

> Concessive clauses indicate that the situation in the matrix clause is contrary to expectation in the light of what is said in the concessive clause. In consequence of the mutuality, it is often purely a matter of choice which clause is made subordinate:
>
> *No goals were scored, although it was an exciting game.*
> *It was an exciting game, although no goals were scored.*
>
> Often they also imply contrast between the situations described by the two clauses.

In a first language acquisition context, the complexity of concessives has long been established. Champaud & Bassano (1994: 416) refer to numerous previous studies which have shown that

a) concessive markers are acquired later than markers of other semantic relationship,
b) 'real' concessive markers (excluding the *but*-type, cf. Scott 1984) are usually quite uncommon (i.e. infrequent) in young children's utterances, pointing to a rather slow adoption of an adult-like production,
c) concessive markers are also among the last to be processed correctly (evidence from psycholinguistic studies involving tasks such as comprehension, interpretation, acceptability judgements, completion or repetition of sentences and lexical decision).

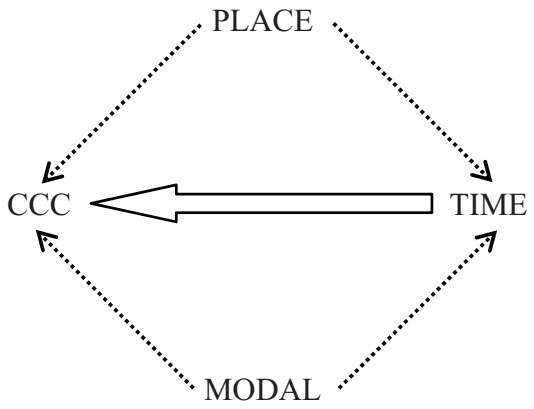


Figure 1: The macrostructure of the semantic space of interclausal relations (adapted from Kortmann 1997: 178)

Studying adverbial subordination from a typological perspective, Kortmann (1997) shows that the polyfunctionality of most adverbial subordinators (e.g. English *while* both temporal and adversative/concessive) is shared by a wide range of languages from very different language families. The typical paths of semantic change he establishes are shown in Figure 1 (see Kortmann 1997: 178; 'CCC' stands for causal, conditional, concessive). Numerous other arguments also point to an ordering of adverbial subordinators on a cognitive gradient; of relevance here are in particular the parallels in synchronic complexity and diachronic development: concessives are the end point of grammaticalisation; they develop late in the history of a language; they are acquired last by children. Moreover, the acquisition of adverbial relations by children seems to follow a generalisable order (cf. Kortmann 1997: 155f.):

time > cause, result, purpose > hypotheticals, counterfactuals > contrastives, concessives

Kortmann states that "[t]he explanation of theses ontogenetic facts standardly is that the general course of language acquisition reflects increasing degrees of cognitive complexity inherent in the different circumstantial relations these connectives express." (1997: 156f.)

It is probably because of the complexity of concessive relations both from a processing and production perspective that large quantitative studies of first language acquisition are not widely available. The situation is even more dire for second language acquisition data. The present study intends to add to the discussion in providing quantitative as well as qualitative evidence of key problem areas for two learner groups, namely German and Czech university students enrolled in English Language & Literature programmes in Germany (Chemnitz University of Technology, CUT) and the Czech Republic (Brno Masaryk University, BMU). Since concessive/contrastive markers are very rare in spoken language (see e.g. Povolná's 2009 study on MICASE), the focus will be on written language exclusively.

By first looking at native speaker usage of selected concessive markers in large L1 reference corpora (*British National Corpus*, BNC; BYU Version, see Davies 2004-; *Corpus of Contemporary American English,* COCA; see Davies 2008-), we will gain insight into general distributional patterns in academic writing, before zooming in on student writing as a sub-genre. An L1 database provided by parts of the *Corpus of British Academic Written English* (BAWE[1]) and the *Michigan Corpus of Upper Student Papers* (MICUSP) will be compared with texts written by German natives at different stages of their studies (ranging from second to fifth year and later) as well as similar texts produced by Czech natives. The Czech and German data also offer the opportunity for a further genre comparison, since the texts stem from two to three different sub-disciplines (namely linguistics, literature plus methodology for the Czech and cultural studies for the German students).

2. Data

2.1. BNC and COCA

At 100 and over 410 million words, the BNC and COCA are the largest reference corpora currently publicly available. Both contain different genres, both contain a sizeable amount of spoken material (10% in the BNC, some 20% in COCA). For the present study, the focus will be on those texts classified as 'academic' by the corpus compilers. These add up to almost 83 million words in COCA (82,914,544) and some 15 million (15,331,668) in the BNC.

sub-discipline	words
Engineering	678,621
Humanities	3,296,072
Law	4,615,173
Medicine	1,412,808
Natural Sciences	1,104,527
Social Sciences	4,224,467
total	**15,331,668**

Table 1: Academic sub-disciplines and respective word totals in the BNC

The texts stem from various academic disciplines, and here we are already confronted with the first problem when comparing frequencies across corpora. Table 1 lists the academic sub-disciplines identified in the BNC, Table 2 gives those of COCA. It is obvious that the overlap is not perfect – while some sections

[1] BAWE was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800); see http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/.

are identical (e.g. Humanities, Medicine), others clearly are not, and the differences seem large enough to make generalisations across sub-disciplines impossible.

sub-discipline	words
Education	8,995,502
Geology/Social Sciences	18,355,867
History	13,023,995
Humanities	12,754,473
Law/Political Science	9,755,165
Medicine	6,001,786
Miscellaneous	4,393,986
Philosophy/Religion	7,536,860
Science & Technology	15,157,304
total	**82,914,544**

Table 2: Academic sub-disciplines and respective word totals in COCA

2.2. BAWE and MICUSP

BAWE and MICUSP were used as reference corpora for British and American student writing 'standards' respectively. Unfortunately, the web-based interface of MICUSP does not allow easy retrieval of statistical data except for overall frequencies, which is why analyses will be restricted to this type of data. The following quotes from the respective corpus websites provide background information on the range and type of data included in the two corpora.

> The BAWE corpus contains 2,761 pieces of proficient assessed student writing, ranging in length from about 500 words to about 5,000 words. Holdings are fairly evenly distributed across four broad disciplinary areas (Arts and Humanities, Social Sciences, Life Sciences and Physical Sciences) and across four levels of study (undergraduate and taught masters level). Thirty-five disciplines are represented. (adapted from http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/)

> The Michigan Corpus of Upper-level Student Papers (MICUSP) is a collection of around 830 A grade papers (roughly 2,6 million words) from a range of disciplines across four academic divisions (Humanities and Arts, Social Sciences, Biological and Health Sciences, Physical Sciences) of the University of Michigan (U-M), Ann Arbor. (adapted from http://micusp.elicorpora.info/)

BAWE is distributed via the Oxford text archive in individual files together with a very helpful overview spreadsheet, allowing researchers to compile their own sub-corpora according to specific criteria. For the present study, four such corpora were compiled. The first contains all files which give "English" as the L1 of the student author. The second contains all non-L1 files regardless of native language; the third is made up of all L1 files identified as "Linguistics" (discipline), the fourth of those tagged as "English" (discipline).

MICUSP is available online and offers a number of tools to narrow searches (student levels – nativeness – textual features – paper types – disciplines).

Unfortunately, the non-native component of MICUSP is – despite its 147 files – not very reliable for comparative studies, as token frequencies of the items under investigation here are < 5 for many, sometimes most, of the disciplines. Moreover, when comparing the frequencies of the selected items, the non-L1 papers included in MICUSP behave very exceptionally, with differences of factor 5 in relation to the native-speaker texts not uncommon.[2] It was thus decided not to introduce a native/non-native contrast in the American data, though this would have been ideal for comparison with the British data.

The MICUSP interface allows the original papers to be browsed in html format, and they can also be downloaded in pdf format. However, the online search tool only enables 'normal' word or phrase searches; wildcard searches for parts of words (e.g. certain inflectional or derivational morphemes) are impossible. For a full-blown analysis, the researcher would thus have to download/convert the papers they are interested in into text files which can then be used as input for a concordancing tool for closer analysis. This was done with a sub-component of the linguistics papers, namely all research papers and all argumentative essays by native English speakers, resulting in 19 files (62,307 words). These will be used in the qualitative part of this study (see 3.2). It should be mentioned here that this sub-corpus is far from representative: a closer look at the contributors reveals that at least 4 of the research papers (which are the longest papers in the corpus and thus responsible for a large share of the overall words) were written by the same individual. These contributions add up to almost 12,000 words, i.e. almost 20% of the total corpus. Table 3 provides an overview of the different (sub-)corpora of BAWE and MICUSP used in the present study.

corpus	files	words
BAWE L1 English	1,956	4,608,528
BAWE non-L1	807	1,963,641
BAWE Linguistics (L1 only)	76	174,828
BAWE English (L1 only)	93	232,544

Table 3: BAWE and MICUSP sub-corpora

As for genre differences, which will also play a role in the analysis, Figure 2 reveals that MICUSP paper types are very different in such 'close' disciplines as Linguistics and English. While research papers and reports constitute the (quantitative) majority of writing in Linguistics, they play negligible roles in English, where Argumentative Essays (which in turn are only rarely found in Linguistics) dominate by a wide margin. This should be kept in mind for the results section.

[2] An overview of the frequencies per 10,000 words in the L1 and non-L1 sections of MICUSP is presented here:

	even though	*although*	*nevertheless*	*whereas*	*however*	*on the other hand*
MICUSP all L1	0.67	3.53	0.38	0.62	9.58	0.75
MICUSP non-L1	0.18	0.73	0.11	0.17	2.12	0.28

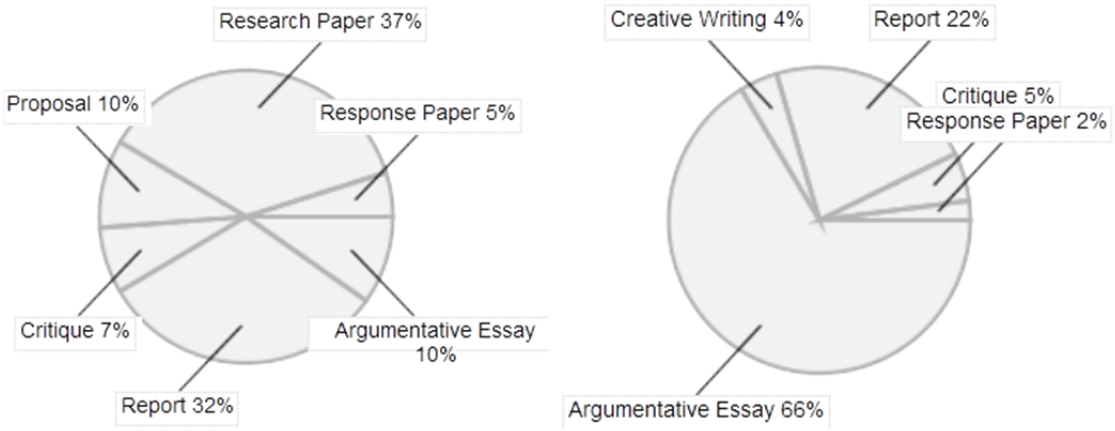


Figure 2: Distribution of paper types (MICUSP, 'Linguistics' and 'English')

2.3. CUT and BMU

The material from Germany and the Czech Republic was collected in part in connection with a joint Czech-German research project on Academic English.[3] On the German side, student writing of different levels and types is included, ranging from term papers at the end of year 2 to *Magister* theses (usually written after 4 or 5 years of study, but often later). Data from 2 different disciplines, namely Linguistics and Cultural Studies/Literature is included. All quotes from secondary sources were excluded to avoid 'outside'/secondary influence in the analyses. Unfortunately, the relative rarity of theses written in English outside Linguistics (particularly in the Literature section) makes a more fine-grained comparison or further breakdown into disciplines impossible. The original ChemCorpus consisting of *Magister* Theses and *Magister* Exams (4-hour exams written by students of the *Magister* programme on a topic of their own choice, but without knowing the concrete research question) is supplemented by term papers written in a linguistics course at the end of year 2 in the Bachelor programme, as well as by Bachelor Thesis (linguistics only). All texts were produced between 2001 and 2011. Taken together, 115 files result in 1,268,358 words (cf. Table 4). More details on the ChemCorpus and its future can be found in Schmied (this volume).

Student writing by Czech natives is represented by 10 theses each from 3 different sub-disciplines, namely Linguistics, Literature and Methodology (Masaryk University's programmes are mostly teacher training programmes), produced between 2005 and 2010. Thirty files add up to 560,404 words (cf. Table 4).

[3] In part funded by the German Academic Exchange Service (DAAD).

corpus	# files	# words
Bachelor Year 2 Term Papers Linguistics CUT	18	75,528
Bachelor Theses Linguistics CUT	11	143,692
Magister Exams Linguistics CUT	52	103,732
Magister Theses Linguistics CUT	24	652,869
Magister Theses Cultural Studies/Literature CUT	10	292,537
total CUT	**115**	**1,268,358**
MagTheses Brno Linguistics	10	178,243
MagTheses Brno Literature	10	178,207
MagTheses Brno Methodology	10	203,954
total BMU	**30**	**560,404**

Table 4: German and Czech student writing – corpus details

3. Methodology

3.1. Quantitative methodology

To gain an overview of the distributions of different markers of concessive/ contrastive relations, word queries were run as an initial step for the following items in each of the corpora / sub-corpora mentioned in Section 2: *although, however, nevertheless, on the one hand, on the other hand, (even) though, whereas*.

Frequencies per 10,000 words were calculated for each query, which will serve as basis of comparison for the remainder of this study. Markers were not scrutinised qualitatively; i.e. for those markers that display a certain degree of polyfunctionality, the different meaning(s) were neglected. Such a procedure is justified by the generally negligible frequencies of the non-clausal meaning facets. Just to give one example, *however* + adjective (e.g. *however important politics might be …*) yields only 1,080 hits in COCA academic, with most combinations occurring only once (with 21 hits, *however important* is the most frequent collocation). The rates for the BNC are similar, with a total of 488 hits (*however small* ranks first in the collocations list with 13 hits). These frequencies contrast with 75,385 (COCA academic) and 18,722 (BNC academic) overall hits for *however*, clearly supporting the negligible nature of the adjective qualification meaning.

The comparisons of the large reference corpora then formed the basis of further investigations of particularly interesting distributions. One of the most surprising findings in this respect concerns the relationship between *although* and *even though*, which will be scrutinised further in Section 4.2.1. Statistical analyses also revealed that for the corpora in question, the markers *however, nevertheless* and *whereas* are the most interesting ones, since their distribution in the non-native corpora (German and Czech student writing) deviates most strongly from the native speaker reference data. These markers will be discussed individually with the help of more qualitative analyses in Sections 4.2.2–4.2.4, focussing in particular on the collocational differences which may help explain the findings.

3.2. Qualitative methodology

In addition to the quantitative evidence from frequency patterns, three markers (which revealed the most striking asymmetries in distribution) were also analysed qualitatively. For *however, nevertheless* and *whereas,* collocational patterns were established with the help of the concordancing programme *AntConc*[4]. The most frequent collocations for each of the sub-corpora were compared with each other to discover L1 vs. L2 or possible genre differences.

4. Results and Discussion

4.1. Overall distribution of markers in the corpora

4.1.1. Reference Corpora

A first step in the reference corpora analysis was the exclusion of *on the one hand* – the phrase is practically non-existent in native-speaker data: 28 hits in the L1 papers from MICUSP (none in Linguistics, 5 in English) and 50 in L1-data from BAWE (no hits in either Linguistics or English). This finding is a surprise in so far as a plethora of academic writing teachers and guidelines (attempt to) reinforce a strict "if you use *on the other hand,* it has to be preceded by *on the one hand*" rule.

The ratio between the two phrases in L1 data tells a different story: in MICUSP, 28 *on the one hand* contrast with 219 *on the other hand*, and 50 *on the on hand* in BAWE with 345 *on the other hand.* It is noteworthy that this asymmetry is much more pronounced – in fact more than twice as strong – in student writing than in academic English in general, as a comparison with the ratios for the BNC and COCA (academic) reveals (see Table 5). It should also be noted that the asymmetry is very likely even stronger than indicated here, as many instances of *on the other hand* will be reduced to *on the other* if preceded by *on the one hand.*

corpus	*on the one hand*		*on the other hand*		total
	#	%	#	%	#
BAWE	50	12.7	345	87.3	395
MICUSP	28	11.3	219	88.7	247
BNC academic	629	26.1	1,784	73.9	2,413
COCA academic	2,362	24.7	7,184	75.3	9,546

Table 5: Frequency of *on the one hand* and *on the other hand* in BAWE, MICUSP, BNC and COCA

[4] Information about AntConc is available at Laurence Anthony's, its programmer's, website: http://www.antlab.sci.waseda.ac.jp/software.html.

A second very striking finding concerns the concessive marker *even though*: it occurs exactly once(!) in the academic part of the BNC, and there are only 17 hits in total in the whole corpus. It is unclear why this marker is so rare in the British data; with almost 37,000 hits in COCA (some 7,000 of which in the academic section), it is far from exceptional in the American counterpart.[5] The overall distribution of the selected markers is given in Table 6. Neither the sum of differences nor any of the individual differences reach statistical significance (chi-square test), indicating that there are no variety-specific patterns for these markers.

	BNC academic	COCA academic
though	3.52	2.87
(even though)	1 hit! (0.0007)	0.84
although	6.95	5.81
nevertheless	1.55	0.81
whereas	1.45	1.05
however	12.24	9.09
on the other hand	1.16	0.87

Table 6: Frequencies per 10,000 words of selected concessive/contrastive markers (BNC academic, COCA academic)

Table 7 presents the findings for the seven markers in the academic writing corpora MICUSP (AmE) and BAWE (BrE).[6] Once more, none of the (individual or compiled) differences reach statistical significance, but the figures for *however* are close ($p = 0.07$). Table 7 also shows the distribution in the two sub-disciplines of Linguistics and English for each of the L1 corpora. No statistically significant differences can be found within the Linguistics (AmE vs. BrE) or English (AmE vs. BrE) sub-corpus nor between the two disciplines (AmE – linguistics vs. English; BrE – linguistics vs. English) either. It is interesting to note however that differences between AmE and BrE within the discipline of Linguistics are larger than any of the others[7]. Despite their lack of significance, the most interesting contrasts are indicated by boldface font in Table 7. *Though* is considerably more frequent than average in both the American Linguistics and English sub-corpora; *whereas* seems overrepresented in the British Linguistics sub-corpus; and the contrasts which BrE and AmE display on a general level for *however* disappears in the sub-corpora.

[5] A look at other sources is not very helpful either as e.g. Biber et al. do not discuss *though* and *even though* separately (cf. e.g. Biber et al. 1999: 842).

[6] A table with absolute and relative frequencies for all corpora under discussion can be found in the Appendix.

[7] P-values: MICUSP Linguistics vs. English: 0,98394; BAWE Linguistics vs. English: 0,82585; Linguistics – AmE vs. BrE: 0,72461; English – AmE vs. BrE: 0,89301.

	MICUSP average	BAWE average	linguistics		English	
			MICUSP	BAWE	MICUSP	BAWE
though	3.18	2.96	**5.11**	2.00	**8.53**	3.96
even though	0.66	0.76	0.99	0.92	1.05	1.03
although	3.53	7.38	5.05	10.13	3.00	6.32
nevertheless	0.38	0.58	0.35	0.34	0.54	0.47
whereas	0.62	1.58	0.70	**3.15**	0.67	0.99
however	**9.58**	**19.26**	11.33	10.24	9.44	12.95
on the other hand	0.75	0.75	0.93	0.34	0.71	1.25

Table 7: Frequencies per 10,000 words of selected concessive/contrastive markers (MICUSP, BAWE)

4.1.2. Genre differences in L1 corpora

Before looking at the distributions of the selected concessive/contrastive markers in different genres, a word of caution is in order. Unfortunately, despite the meticulous categorisation of texts into academic sub-genres, the BNC and COCA academic sections are not really comparable. This is owed to the different number and very different classification of genres. Thus, 6 categories (engineering – humanities – law – medicine – natSci – socSci) in the BNC contrast with 9 in COCA (Education – Geog/SocSc – History – Humanities – Law/PolSci – Medicine – Misc – Phil/Rel – Sci/Tech). This should be kept in mind for the following presentation of results.

It is remarkable that the selected items practically only occur in extremes. Only 3 values are neither bold nor in italics in Tables 8 and 9[8]: however, the values for *whereas* are also noteworthy in both corpora, since they are clearly below average with medicine leading the ranking for both BrE and AmE. As for the remainder of the markers, a very interesting picture emerges which hints at genre- and/or variety-specific differences. *Although* in non-initial position occurs at average frequencies in the humanities section of the BNC, but far below average in COCA. The varieties agree as to the initial use of *although*, which figures at frequencies very close to or at the lowest value in both corpora. This corresponds with Biber et al.'s findings: "Concessive clauses show a slight preference for final position; this preference is shared across all registers." (1999: 833) However, while the ratio there is 2:3, the figures here point towards an even stronger preference (approximately 1:3).

Not very surprising given the comparatively non-formal nature of much humanities writing, *though* is represented with the highest values in both corpora. A similar picture emerges for *nevertheless*. Because of the non-existence of *even though* in the BNC, nothing can be deduced from the COCA findings alone except to note that it is not very prominent in the humanities.

[8] The following conventions hold for Tables 8 and 9: if humanities = highest value (or very close): bold; if lowest (or close): italics.

	average academic	lowest value	highest value	humanities
although	6.95	5.70 (engineering)	9.13 (medicine)	6.90
although (initial)	2.30	1.91 (humanities)	4.20 (medicine)	*1.91*
though	3.52	1.42 (medicine)	5.73 (humanities)	**5.73**
nevertheless	1.55	0.63 (engineering)	2.00 (humanities)	**2.00**
whereas	1.45	1.00 (law)	2.29 (medicine)	1.33

Table 8: Frequencies per 10,000 words of selected concessive/contrastive markers by genre in BNC academic

	average academic	lowest value	highest value	humanities
although	5.81	4.03 (Phil/Rel)	6.02 (medicine)	*4.26*
although (initial)	2.52	1.53 (Phil/Rel)	2.88 (Geog/SocSc)	*1.68*
though	2.88	1.12 (medicine)	3.31 (humanities)	**3.31**
even though	0.84	0.51 (medicine)	0.97 (Law/PolSc)	*0.68*
nevertheless	0.81	0.46 (education)	1.00 (history)	**0.94**
whereas	1.05	0.57 (Law/PolSc)	1.13 (medicine)	0.77

Table 9: Frequencies per 10,000 words of selected concessive/contrastive markers by genre in COCA academic

Overall, we can observe that the humanities behave more 'extremely' in the BNC. It is possible that this is a result of the different genre classifications. Thus, lowest or highest values occur in 3 out of 5 categories in the BNC, but only in 1 out of 6 in COCA. The almost complementary distribution of humanities and medicine in the BNC is remarkable and an interesting topic for further research; COCA sub-disciplines are more evenly represented.

The overall spread of values is much broader in BNC, indicating less strict/conventionalised styles in BrE in comparison with AmE. It would be interesting to attempt to collate these with possible in-house policies for academic writing, which tend to be more extreme in US publishing companies, as many academics can confirm.[9]

4.1.3. L2 Corpora (German & Czech student writing)

As indicated already, the detailed comparisons will focus on a selection of markers, namely *although, even though, however, nevertheless, on the other hand* and *whereas*. All statistical analyses are based on these markers only. Needless to say, contrasting them in any combination of the L1 reference corpora reveals no significant difference whatsoever, with p-values between 0.72 and 0.99. In addition to an overall comparison of these markers, the relationship between *although* and *even though* will be analysed in more detail.

[9] A cliché example involves the automatic 'correction' of every restrictive relative pronoun to *that* and every non-restrictive one to *which*, regardless of the resulting (un)grammaticality.

4.1.3.1. CUT student writing vs. reference corpora

For all sub-corpora of Chemnitz student writing, chi-square values were calculated for comparisons of the six concessive/contrastive markers *although, even though, however, nevertheless, on the other hand* and *whereas* (per-10,000-word frequencies). The results can be found in Table 10. Chi-square values approaching significance are highlighted in boldface. Generally, no clear patterns can be deduced. Some generalisations are possible, however:

- CUT students show an affinity to American rather than British writing (smaller differences to MICUSP than to BAWE data in all sub-corpora), particularly at BA level (Magister: more British)
- CUT student writing is closer to general L1 writing than to specialised linguistics writing (smaller differences in all sub-corpora)
- the best-planned type of writing, namely the Magister Theses, shows the lowest values for 'non-nativeness' (0.56, comparison with BAWE non-L1)

(sub-)corpus	BAWE L1	BAWE non-L1	BAWE Ling.	MICUSP L1	MICUSP Ling.
BA term papers year 2	0.22	0.64	0.30	0.61	0.47
BA Theses Linguistics	0.33	0.81	**0.11**	0.73	0.53
Magister written exams	0.24	0.77	**0.14**	0.64	0.46
Magister Theses Linguistics	0.24	0.56	0.36	0.60	0.47

Table 10: Chi-square values CUT vs. Reference Corpora (six concessive/contrastive markers)

When zooming in on the markers with the most striking differences in average frequencies, *however, nevertheless* and *whereas*, a clearer – though not necessarily more helpful – picture emerges (Table 11): for all sub-corpora, the differences are the smallest for the comparison with the British linguistics sub-corpus and the largest in contrast with British L1 writing in general. Interestingly, the figures for non-native British writing and specialised American writing (linguistics sub-corpus) are similar to each other. For the comparison with American writing, the reverse scenario of the British observation holds: CUT writing is closer to generalised rather than specialised writing.

This last finding in fact nicely reflects the every-day goings on at CUT. We can safely assume that Chemnitz students are (or were) confronted with general American academic writing more often than with particularly linguistic styles, based on the departmental competences: the practical language programme used to be mostly 'American', while linguistics classes (including the literature in question) have a more British focus (with texts on World Englishes often being more British in style than English English natives'). Note that at the level of Magister Theses, the difference to BrE L1 writing reaches statistical significance ($p = 0.04$).

reference corpus	BA year 2	BA Theses	Magister Exams	Magister Theses
BAWE L1	0.08	0.28	0.18	**0.04**
BAWE non-L1	0.22	0.54	0.41	0.18
BAWE Linguistics	0.56	0.78	0.72	0.28
MICUSP L1	0.21	0.46	0.35	0.16
MICUSP Linguistics	0.15	0.37	0.27	0.10

Table 11: CUT student writing vs. reference corpora: *however, nevertheless, whereas* (chi-square values)

4.1.3.2. BMU student writing vs. reference corpora

The same procedure adopted for CUT student writing will also be followed for the Czech data. For BMU, only one type (level) of texts is available, namely the final year theses. The first thing one notices in Table 12 is that Czech student writing is very different from native British English production, almost reaching significance both for the general reference corpus (BAWE L1, p = 0.07) and the specialised linguistics one (p = 0.06). In comparison with non-native data, Czech students are not 'typical' either, however, with a p-value of 0.5 suggesting that the non-native writing in BAWE is not very 'Czech'. This is not surprising in light of the fact that only 5 files (out of 807) with a total of 9,963 words (i.e. less than 0.5% of the word total) are classified as 'Czech L1' in BAWE (moreover, all files stem from the discipline 'Business'). The comparison with American student writing places Czechs closer to that than to BrE, but the distance is still quite large.

reference corpus	Czech Magister Theses
BAWE L1	**0.07**
BAWE non-L1	0.50
BAWE Ling.	**0.06**
MICUSP L1	0.45
MICUSP Ling.	0.35

Table 12: Chi-square values BMU vs. reference corpora (six concessive/contrastive markers)

Table 13 presents the chi-square values for the three markers *however, nevertheless* and *whereas*. Except for one value, namely the one for the comparison with specialised BrE (BAWE Linguistics), all differences become more pronounced, indicating that it is those three markers in particular which distinguish Czech writing from native writing most strongly.[10]

[10] Why the difference to non-native BrE also becomes more pronounced is unclear at this point and should be investigated more closely.

reference corpus	Czech Magister Theses
BAWE L1	**0.06**
BAWE non-L1	0.23
BAWE Linguistics	**0.09**
MICUSP L1	0.23
MICUSP Linguistics	0.17

Table 13: Chi-square values BMU vs. reference corpora: *however, nevertheless, whereas* (chi-square values)

4.1.3.3. CUT vs. BMU student writing

This section attempts to put the results from the two previous ones into perspective, comparing German and Czech student writing in relation to native-speaker production. Table 14 summarises the chi-square values relevant for the discussion. Since the types of texts produced in both countries are not 100% comparable, figures for both the BA and the Magister Theses on the German side are included. The Czech theses are assumed to be in between these in terms of level (based on length of studying).

The most obvious finding is the general 'distance' of the Czech data to all of the reference corpora as well as the German texts, with the values for the BrE L1 comparisons approaching significance. Of the L1 corpora, student writing is closest to American general academic writing for both Germans and Czechs. Germans are more typically 'non-native British', indicated by the highest chi-square value of 0.81 in the table in comparison to BAWE non-L1 data. This, as already mentioned above, is most likely owed to the under-representation of Czech speakers in BAWE (there are 57 texts produced by German natives, amounting to over 146,000 words, i.e. ca. 15x the Czech word number).

reference corpus	chi-square		
	CUT BA Theses	CUT Magister Theses	BMU Magister Theses
BAWE L1	0.33	0.24	**0.07**
BAWE non-L1	0.81	0.56	0.50
BAWE Linguistics	0.11	0.36	**0.06**
MICUSP L1	0.73	0.60	0.45
MICUSP Linguistics	0.53	0.47	0.35
BMU Magister Theses	0.16	0.37	—

Table 14: German and Czech student writing in comparison – chi-square values (six concessive/contrastive markers)

Concentrating on *however, nevertheless, whereas* (Table 15) once more reveals striking contrasts between Czech and German student writing. Except for one value, namely the comparison with specialised (linguistic) BrE writing, it is the German students who produce more 'a-typical' instances of the three markers in question, with one value (BAWE L1) even reaching significance at $p = 0.04$. As in most of the tables discussed so far, the increase in distance does not only

concern distance to the native reference corpora, but also to the non-native BAWE texts. It seems likely that this is an artefact of the corpus (imbalance and consequential non-representativeness in terms of L1s and English levels) rather than a real difference. Consequently, the CUT data were also compared with the German-L1 sub-component of BAWE only.

reference corpus	CUT BA Theses	CUT Magister Theses	BMU Magister Theses
BAWE L1	0.28	**0.04**	**0.06**
BAWE non-L1	0.54	0.18	0.23
BAWE Linguistics	0.78	0.28	**0.09**
MICUSP L1	0.46	0.16	0.23
MICUSP Linguistics	0.37	0.10	0.17

Table 15: German and Czech student writing in comparison – *however, nevertheless, whereas* (chi-square values)

As can be clearly seen in Table 16, the CUT data and the German data included in BAWE are very similar with regard to their use of the markers in question. For practically all comparisons, the distance (in terms of chi-square values) is the smallest between these two. The one value 'out of line' is the Magister Theses value, which can be explained with the longest / most elaborate level of planning involved in this type of text (6 months vs. 18 weeks for BA Theses).

CUT sub-corpus	BAWE L1 German
BA year 2 term papers Linguistics	0.73
BA theses Linguistics	0.95
Magister written exams Linguistics	0.87
Magister Theses Linguistics	0.46

Table 16: CUT student writing vs. L1 German texts (BAWE) – six concessive/contrastive markers (chi-square values)

4.2. Qualitative differences in individual markers

4.2.1. Although, though and even though

The differences between *although, though* and *even though* are discussed in a number of publications and are also of concern to grammar writers. Biber et al. (1999) in particular are useful in the present context as their grammar includes comments on genre-based distributions and differences which may or may not be mirrored in the present study.

Commenting on the difference between *although* and *even though*, Lewin, Fine & Young (2001: 141) note that "*although* selects 'approval unmarked' compared to *even though* which selects for 'disapproval'." Quirk et al. (1985:

1099) similarly argue that "[*e*]*ven though* and *even when* are more emphatic forms of *though* and *when*, the modifying *even* also expressing unexpectedness."

Despite claims as to their synonymous status when used as subordinators (Biber et al. 1999: 845), the difference between *though* and *although* is clearly one of genre. Biber et al. can show that (*even*) *though* is a marker of fiction, while *although* marks academic writing (1999: 842). Furthermore, "[i]n academic prose, *although* is about three times as frequent as *though*. *Although* seems to have a slightly more formal tone to it, fitting the style of academic prose" (845). "The greater use of *although* by writers of academic prose may also result from an attempt to distinguish this subordinator from the common use of *though* as a linking adverbial in conversation" (846). As to the polyfunctional status of *though* as subordinator and what they call "linking adverbial", they explain further:

> *Though* is much more common overall in conversation and fiction than in news and academic prose. In the use of *though*, conversation is very different from all three written registers. Most occurrences of *though* in conversation are as a linking adverbial. In the written registers, the vast majority of the occurrences are as a subordinator. (Biber et al. 1999: 850)

The following can be gleaned from the corpus analyses[11]:

- *though* as conjunction in first position:
  - *not* a spoken phenomenon (as one might think) – most frequent in magazine writing and fiction (COCA; similar in BNC);
  - also a feature of academic writing – but very rare in comparison to 'official' uses (COCA: 0.34 per 10,000 words; BNC: 0.37)
- *though* as sentence-final conjunction (, *though*.):
  - historically, steady rise in frequency throughout the 19th (COHA) and 20th (COHA, TIME) century (more than 10x the frequency now than it started out with; colloquialisation?; cf. Biber & Finegan 1989, Mair & Hundt 1995, Mair 1997); very much a fiction feature in BrE;
  - rare in magazines, academic writing AND speech – BUT: in COCA, about as frequent in spoken language as in fiction
- *even though*
  - dramatic increase historically – from 0.11 per 10,000 words in 1810s to 0.97 in 2000s (COHA)
  - most frequent in speech, but only small differences between genres (COCA: 0.94 per 10,000 words in spoken vs. 0.93 in fiction, 0.84 in academic, 0.83 in newspaper and 0.81 in magazine writing)

What is most surprising when comparing the L1 reference corpora with the corpora of student writing are the ratios of *even though* in relation to *though* and

[11] The historical data cited here is based on frequencies from the POS-tagged *Corpus of Historical American English* (COHA; Davies 2010-) and the *TIME Magazine Corpus* (TIME, Davies 2007-).

although: in native speaker production, *even though* clearly represents a marked option in comparison to both *though* and *although*. This is to be expected given the extra 'strength' that *even* adds. Table 17 also shows that overuse of *even though* is a non-native phenomenon in both MICUSP and BAWE, with percentages rising more than 10% in comparison to the L1 figures. Differences between BrE and AmE can also be observed: while both Linguistics and English as sub-disciplines of academic writing display lower frequencies of *even though* in the American data, the opposite holds for the British data. A nice parallel can be seen in the relationship between *even though* and *although* in the L1 and Linguistics sub-components of both varieties: figures are identical, but American writers use *even though* twice as often as *although* compared with their British counterparts.

reference corpora	*though*	*even though*	% *even though*	*although*	ratio *even though* : *although*
MICUSP all L1	1,008	213	21.13	1,081	0.2
MICUSP Linguistics (L1 only)	88	17	19.32	87	0.2
MICUSP English (L1 only)	253	31	12.25	89	0.35
MICUSP all non-L1	159	48	30.19	211	0.23
BAWE total	2,023	617	30.5	4,500	0.14
BAWE Linguistics (L1 only)	35	16	45.71	177	0.09
BAWE English (L1 only)	92	24	26.09	147	0.16
BAWE all L1	1,364	348	25.51	3,401	0.1
BAWE all non-L1	660	271	41.06	1,115	0.24
COCA academic	23,853	6,981	29.27	48,206	0.14
BNC academic	5,400	1	0.02	10,656	0

Table 17: Distribution of *even though, though* and *although* in the reference corpora

A comparison of the relative frequencies (cf. Appendix) is also indicative of a typically German overuse of *even though*: while native speaker rates are generally below 1 occurrence per 10,000 words, German rates tend to be higher than 2, i.e. more than twice as high as those of the natives.

Table 18 displays the same data as Table 17, focusing on the German and Czech corpora. The differences are more than striking. First, both German and Czech students heavily overuse the marked *even though* in relation to both *though* and *although*. In some cases, *even though* is responsible for almost two thirds to three quarters of all instances of *though* in the respective corpus. Interestingly, the difference is least pronounced in the Linguistics Theses at the highest level (Magister) in both languages; in fact, at 44.2 and 48.3% respectively, it is even close to the BAWE Linguistics Corpus (41.1%). A further parallel between the German and Czech data can be seen in the non-linguistic sub-components: both Cultural Studies/Literature texts from CUT as well as Brno Literature and Methodology theses display the most pronounced overuse of *even though.*

As for the ratio of *even though* to *although*, both German and Czech natives only rarely approach a native speaker distribution. Instead, the marked variant

even though is used much more frequently than in the comparable L1 data. For most sub-corpora, the distribution is around 1:2 (i.e. *even though* occurs about half as much as *although*). Recall that the same ratio in the L1 corpora was around 1:5 (AmE) and 1:10 (BrE). Once more we can see that both German and Czech students are closer to AmE than to BrE in this respect; in fact, the most advanced type of German student writing (Magister Theses) displays a 'native-like' production (ratio of 1:5), at least in the Linguistics sub-component. The same holds for the Brno Literature Theses. It should be mentioned however that neither the AmE nor the BrE corpora show any indication of a native/non-native distinction with regard to this feature; we can equally claim that CUT and BMU students follow a non-native pattern.

German and Czech student writing	*though*	*even though*	% *even though*	*although*	ratio *even though : although*
BA year 2 CUT Ling. term papers	33	20	60.61	42	0.48
BA Theses CUT Linguistics	83	41	49.4	69	0.59
Magister Theses CUT Linguistics	236	114	48.31	579	0.2
Magister Exams CUT Linguistics	30	22	73.33	40	0.55
Magister Theses CUT Cult&Lit	88	67	76.14	168	0.4
MagTheses Brno Linguistics	86	38	44.19	99	0.38
MagTheses Brno Literature	43	26	60.47	120	0.22
MagTheses Brno Methodology	52	34	65.38	66	0.52

Table 18: Distribution of *even though, though* and *although* in the German & Czech corpora

A comparison of the position of the three markers in the sentence also proves interesting. Unfortunately, low token frequencies (< 5) for initial *though* make it impossible to discuss positional differences for *though* in general, so the focus will be on *although* and *even though*. In Table 19, a preference for initial markers (> 50%) is highlighted in grey. A comparison of native and non-native data shows a pronounced preference for non-initial markers in L1 writing, while non-native authors tend to use them more frequently initially than elsewhere. This contrast is even stronger in the distribution of *even though.* However, the table also shows clear differences within the respective non-native data: for Czech writing in particular, there are considerable differences between Linguistics, Literature and Methodology texts, with Linguistics being closest to native-like production. A similar picture emerges for the German data, but here the reverse holds: texts from the linguistic sub-corpus are in line with the non-native findings, while Cultural Studies/Literature texts are more native-like.

		although			*even though*			*though*		
		initial	non-initial	total	initial	non-initial	total	initial	non-initial	total
BAWE L1		1,537	1,864	3,401	94	254	348	250	860	1,110
	%	45.19	54.81		27.01	72.99		22.52	77.48	
BAWE Ling. (L1)		70	106	176	6	11	17	3	29	32
	%	39.77	60.23		35.29	64.71		9.38	90.63	
BAWE Engl. (L1)		76	71	147	8	16	24	7	69	76
	%	51.7	48.3		33.33	66.67		9.21	90.79	
BAWE Grm. (L1)		55	47	102	16	11	27	3	29	32
	%	53.92	46.08		59.26	40.74		9.38	90.63	
BA yr. 2 CUT Ling. TPs		21	21	42	10	10	20	2	31	33
	%	50.0	50.0		50.0	50.0		6.06	93.94	
Mag. Exams CUT Ling.		20	20	40	12	10	22	3	17	20
	%	50.0	50.0		54.55	45.45		15.0	85.0	
Mag. Theses CUT Ling.		303	276	579	63	51	114	26	159	185
	%	52.33	47.67		55.26	44.74		14.05	85.95	
Mag. Th. CUT Cult&Lit		110	58	168	24	43	67	4	41	45
	%	65.48	34.52		35.82	64.18		8.89	91.11	
MagTheses Brno Ling.		39	60	99	13	25	38	2	59	61
	%	39.39	60.61		34.21	65.79		3.28	96.72	
MagTheses Brno Lit.		84	36	120	22	4	26	4	35	39
	%	70.0	30.0		84.62	15.38		10.26	89.74	
MagTheses Brno Met.		41	25	66	19	15	34	2	35	37
	%	62.12	37.88		55.88	44.12		5.41	94.59	
COCA acad. (after ".")		21,277	28,016	49,293	1,330	5,786	7,116	2,917	18,609	21,526
	%	43.16	56.84		18.69	81.31		13.55	86.45	
BNC acad. (after ".")		3,523	7,133	10,656	(not enough tokens)			571	4,829	5,400
	%	33.06	66.94					10.57	89.43	

Table 19: Position of *although, even though* and *though*

4.2.2. However

Concerning the distribution of *however* in relation to *though,* Biber and colleagues summarise that

> [T]he linking adverbial *however* is far more common than *though* in academic prose and carries the same contrastive meaning [...]. The use of *though* appears to be a marked stylistic choice by certain authors. (Biber et al. 1999: 851)

> None of these linking adverbials associated with final position in conversation occur frequently in academic prose. Instead, the common linking adverbs in academic prose – *therefore*, *thus*, and *however* – tend to occur in medial positions (when not in initial position). In particular, these forms often occur immediately following the subject" (*ibid.*: 892)

A comparison of the position of *however* (initial vs. non-initial) in the corpora investigated here reveals no striking differences. There is however a general tendency in terms of genre preferences: both native and non-native texts from the literature and cultural studies sub-corpora show a stronger preference for initial *however* than linguistics texts and general corpora (65-78% vs. 48-62%).

It is noteworthy that both COCA and the BNC academic components actually disprefer initial *however* as the only two corpora in the sample. This could be caused by the query run to identify the relevant tokens. The architecture of the corpora does not allow case-sensitive searches; thus, the pattern ". * *however*" was run and the tokens for ". *however*" were subtracted from the total. This excludes all passage-initial instances of *however*, which may skew the results.

In their study on Finnish academic writing, Ventola & Mauranen (1991: 462) found that Finnish writers overuse *however* at the expense of other markers: "*however* represented the whole category of adversatives for many writers". This cannot be confirmed for the corpora used in this study. We can observe though that German student writing shows relatively high frequencies of 15-20 occurrences per 10,000 words, while Czech students only use *however* app. 10 times per 10,000 words (see Appendix).

This once more suggests German overuse, but a look at the native speaker frequencies tells a different story: *however* is twice as frequent in BAWE L1 data (ca. 20/10,000) as in the comparative MICUSP sample (ca. 10/10,000). For the British data, there is a huge genre difference between the Linguistics and Literature data, with linguistic frequencies around 10 (i.e. half the corpus average). There is no such difference in the AmE data; in fact, the tendency points in the other direction (Ling. > Lit.; cf. Appendix).

One final tendency to be investigated further concerns the collocates of *however*, particularly those occurring immediately to the right of sentence-initial *however*. In native speaker writing, a combination of *however* with another concessive/contrastive clause (or phrase) figures five times among the 20 most frequent collocates (e.g. BAWE L1 in descending order of frequency: *as* – 212*, if* – 121*, when* – 82, *despite* – 61, *although* – 49). Some typical examples can be found in (2). While the top ranking collocates are shared in both native and non-native data (*the, it, this, in, there*) with only minor differences, this more complex type of 'double marking' seems to be a characteristic of L1 authors.[12]

(2) a. ***However, although*** she may be Knightley's equal in prosperity, social status and intelligence, he is far superior to her in reason, and until she becomes his equal in this respect, he cannot truly love her. (BAWE 0229b)

b. ***However, as*** data2 takes account only of the provided experimental values whereas data1 uses the average values over time, it is likely that data1 would provide a much better fit after a significant period of time. (BAWE 0263a)

[12] The 9 instances of *as* in the Czech sub-corpora are produced by only 5 of the 30 authors. Similarly, 6 out of the 11 instances of *however, if* in the CUT Magister Theses corpus stem from one author (6 authors in total).

c. ***However, despite*** the personal nature of some of the experiences, this introduction to tourist behaviour theory suggests that academics' analysis has often been accurate. (BAWE 3042e)

d. ***However, if*** the person with the higher status wanted to change domains to one where status was not so important, they could do so by switching to Guarani. (BAWE 6062d)

e. ***However when*** posed as a means of freedom from the kind of ideology Habermas identifies, it seems less convincing. (BAWE 0010e)

4.2.3. Nevertheless

Historically, a look at the frequencies of *nevertheless* in COHA reveals that the form is becoming less frequent. The trend started in the 1950s and is led by magazine and fiction writing. In addition – and not very surprisingly –, *nevertheless* is a typical feature of academic writing, both in British and American English, as witnessed by the respective per-million-word rates in the BNC and COCA.

Bell (2010) investigates the use of *nevertheless* in a corpus of 8.5 million words "made up of one million words each of academic writing, newspapers, and fiction" (1915). He also compares written and spoken discourse, but given the scope of this paper, the latter will not be discussed here. His relative frequencies (Bell 2010: 1916, Table 1) are of interest for the sake of comparison: at 0.97 occurrences per 10,000 words, his rate of *nevertheless* is slightly higher than the COCA academic rate (0.81, see Appendix). It is however considerably higher than the American rate of 0.38 per 10,000 words (MICUSP, see Appendix). Since the regional provenance of the academic sources (or rather the authors of the chosen articles) cannot be deduced from the information given in the article, it is impossible to say whether Bell's data are more 'British' or more 'American'. It should be noted though that his spoken data stem from US sources exclusively, complicating the comparison.

The author emphasises the positional difference/variability between *however* and *nevertheless*, with some contexts allowing both forms interchangeably while others do not (cf. Bell 2010: 1917). A genre comparison shows that *nevertheless* occurs preferably in intersentential position in academic writing while an intrasentential position is dominant in fiction (*ibid.*: 1916). If disciplines are teased apart, humanities texts favour intrasentential and non-initial *nevertheless* (1918).

A look at *nevertheless* in the L1 and L2 corpora of this study cannot confirm Bell's observation. In fact, the opposite seems to be the case: in all corpora, initial *nevertheless* is the dominant option by far (Table 20). While German student rates fluctuate between 70 and 83%, rates for all disciplines of Czech student writing are above 80%. The rates of initial *nevertheless* for both the BNC and COCA academic are much lower, suggesting that neither linguistics nor English writing is typical of academic writing in general, which seems to prefer non-initial *nevertheless.* An additional search was thus performed within the

academic sections of COCA and the BNC, this time selecting only those papers tagged as 'humanities'. The rates provide an explanation for the deviation between Bell's and our findings: American writers in the humanities seem to prefer non-initial *nevertheless*, but British authors select initial *nevertheless* more frequently. Unfortunately, the MICUSP interface does not allow case-sensitive searches, and a comparison with the linguistics sub-corpus compiled for this study yields a total of only 4 tokens of *nevertheless*, too few to draw any conclusions (but note that 3 of the 4 occur in initial position).

corpus	*nevertheless*				
	initial	non-initial	total	distribution (%)	
BAWE L1	192	74	266	72.18	27.82
BAWE Linguistics (L1 only)	13	2	15	86.67	13.33
BAWE English (L1 only)	8	3	11	72.73	27.27
BAWE German L1	21	7	28	75.0	25.0
BA year 2 CUT Ling. TPs	14	6	20	70.0	30.0
Mag. Exams CUT Linguistics	19	5	24	79.17	20.83
Mag. Theses CUT Linguistics	297	107	404	73.51	26.49
Mag. Theses CUT Cult&Lit	69	14	83	83.13	16.87
MagTheses Brno Linguistics	71	14	85	83.53	16.47
MagTheses Brno Literature	45	11	56	80.36	19.64
MagTheses Brno Methodology	35	6	41	85.37	14.63
COCA academic (after full stop)	3,572	3,272	6,844	52.19	47.81
COCA acad. humanities (after full stop)	466	753	1,219	38.23	61.77
BNC academic (after full stop)	1,419	956	2375	59.75	40.25
BNC acad. humanities (after full stop)	378	288	666	56.76	43.24

Table 20: Position of *nevertheless*

4.2.4. Whereas

Of all the markers investigated here, *whereas* is very clearly the 'most academic' one. In terms of its history, both COHA and the TIME corpus show a steep rise in frequency until the 1930s, followed by a decrease ever since. It seems possible to link this decrease with a perceived colloquialisation of English in the past decades (cf. e.g. Biber & Finegan 1989; Mair & Hundt 1995, Mair 1997).

The large reference corpora also hint at another difference: *whereas* is much more common in spoken BrE than AmE, a finding that actually goes against the colloquialisation trend, at least if one assumes that AmE is generally less formal than BrE. As for frequencies in general, *whereas* is clearly one of the most over-used markers in German student writing: while L1 frequencies hover around 1 instance per 10,000 words, German authors use *whereas* six times as much (cf. Appendix). Interestingly, both the native and non-native data show discipline differences as well: Linguists use *whereas* much more frequently (at a rate of about 2:1) than students from other sections. This is true for CUT writing in general as well as for BAWE (L1 vs. Linguistics; L1 vs. German L1). It does not hold for MICUSP or the Czech data, though. For the latter, there are in fact

hardly any differences discernible between the three sub-disciplines (Linguistics, Literature, Methodology); if anything, then Methodology Theses display an under-use of *whereas*.

corpus	*whereas*				
	initial	non-initial	total	distribution (%)	
BAWE L1	104	624	728	14.29	85.71
BAWE Linguistics (L1 only)	9	48	57	15.79	84.21
BAWE English (L1 only)	2	21	23	8.7	91.3
BAWE German L1	20	41	61	32.79	67.21
BA year 2 CUT Ling. TPs	6	45	51	11.76	88.24
Mag. Exams CUT Linguistics	14	47	61	22.95	77.05
Mag. Theses CUT Linguistics	85	320	405	20.99	79.01
Mag. Theses CUT Cult&Lit	32	64	96	33.33	66.67
MagTheses Brno Linguistics	11	49	60	18.33	81.67
MagTheses Brno Literature	5	15	20	25.0	75.0
MagTheses Brno Methodology	2	10	12	16.67	83.33
COCA academic (after full stop)	1,439	7,456	8,895	16.18	83.82
COCA acad. humanities (after full stop)	271	726	997	27.18	72.82
BNC academic (after full stop)	318	1,903	2,221	14.32	85.68
BNC acad. humanities (after full stop)	93	346	439	21.18	78.82

Table 21: Position of *whereas*

In Table 21, the distribution of initial vs. non-initial *whereas* in the corpora is compared. All corpora show a clear preference for non-initial use, regardless of native speaker status, genre or sub-discipline. Although low token frequencies in some corpora make generalisations impossible, some tendencies can be observed: first, the large reference corpora point to a discipline effect, with a less pronounced trend to initial *whereas* in the humanities. Second, for both the German and Czech corpora, the rates of non-initial *whereas* are higher in the linguistics sub-components than in the literature texts. This trend is not shared by the native speaker reference data (BAWE Linguistics & English), which points in the opposite direction (but note the low token frequencies).

5. Summary and implications for teaching

In the preceding sections, some of the most common markers of concessive/contrastive relationships have been analysed both quantitatively and qualitatively. Native and non-native as well as genre and discipline differences could be established. Among all of the investigated features, the following stand out most strongly:

- a general overuse of a number of explicit contrastive markers by non-native writers of English
- with regard to the investigated markers, German student writing tends to be closer to AmE than to BrE
- Czech students are closer to native patterns for some features, but display striking differences both to native and German writing for others

Some of the findings point to a link with teaching. In classes on academic writing, instructors tend to heavily emphasise the importance of discourse markers in general and of those establishing contrast in particular. It is not impossible that the differences observed here are in fact 'taught'. A next step in the analysis would certainly benefit from discussions with instructors and students to find out more about active teaching methods that might help explain the present findings. Subsequently, it would be desirable to raise students' awareness of alternatives to mark concessive and contrastive relationships. Another avenue of further research involves a more fine-grained distinction of academic sub-disciplines. Even the small-scale comparisons presented here point to major differences in preference of certain patterns in such 'close' disciplines as Linguistics and Literature.

Susanne Wagner
English Language & Linguistics
Chemnitz University of Technology
Germany
susanne.wagner@phil.tu-chemnitz.de

References

Barth-Weingarten, D. (2003). *Concession in Spoken English: on the Realisation of a Discourse-Pragmatic Relation*. Tübingen: Narr.

Bell, D. M. (2010). Nevertheless, Still and Yet: Concessive Cancellative Discourse Markers. *Journal of Pragmatics* 42, 1912-1927.

Biber, D. & E. Finegan (1989). Drift and the Evolution of English Style: a History of Three Genres. *Language* 65, 487-517.

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Champaud, C. & D. Bassano (1994). French Concessive Connectives and Argumentation: an Experimental Study in Eight- to Ten-Year-Old Children. *Journal of Child Language* 21, 415-438.

Davies, M. (2004-). *BYU-BNC: The British National Corpus*. Available Online at http://corpus.byu.edu/bnc.

Davies, M. (2008-). *The Corpus of Contemporary American English (COCA): 410+ Million Words, 1990-Present*. Available Online at http://www.americancorpus.org.

Davies, M. (2010-). *The Corpus of Historical American English (COHA): 400+ Million Words, 1810-2009*. Available Online at http://corpus.byu.edu/coha.

Davies, M. (2007-). *TIME Magazine Corpus (100 Million Words, 1920s-2000s)*. Available Online at http://corpus.byu.edu/time.

Defrancq, B. (2008). Contrastive Corpus Analysis of Cross-Linguistic Asymmetries in Concessive Conditionals. In: R. Xiao, L. He & M. Yue (eds.), *Proceedings of The International Symposium on Using Corpora in Contrastive and Translation Studies (UCCTS 2008)*, Hangzhou: Zhejiang University, 25-27 September 2008. Available from http://www.lancs.ac.uk/fass/projects/corpus/UCCTS2008Proceedings/papers /Defrancq.pdf. Accessed 15.04.2011.

König, E. (1985). On the History of Concessive Connectives in English. Diachronic and Synchronic Evidence. *Lingua* 66, 1-19.

König, E. (1986). Conditionals, Concessive Conditionals and Concessives: Areas of Contrast, Overlap and Neutralization. In: E. Traugott, A. ter Meulen, J. Snitzer Reilly & C. A. Ferguson (eds.), *On Conditionals*. Cambridge: Cambridge University Press, 229-246.

König, E. (1988). Concessive Connectives and Concessive Sentences: Cross-Linguistic Regularities and Pragmatic Principles. In: J. Hawkins (ed.), *Explaining Language Universals*. Oxford: Blackwell, 145-166.

Kortmann, B. (1997). *Adverbial Subordination: A Typology and History of Adverbial Subordinators Based on European Languages*. Berlin/New York: Mouton de Gruyter.

Mair, C. (1997). Parallel Corpora: A Real-Time Approach to the Study of Language Change in Progress. In: M. Ljung (ed.), *Corpus-Based Studies in English. Papers from the 17th International Conference on English Language Research on Computerized Corpora (ICAME 17)*. Amsterdam: Rodopi, 195-209.

Mair, C. & M. Hundt (1995). Why is the Progressive Becoming more Frequent in English? A Corpus-Based Investigation of Language Change in Progress. *Zeitschrift für Anglistik und Amerikanistik* 43, 111-122.

Lewin, B.A., J. Fine & L. Young (2001). *Expository Discourse. A Genre-Based Approach to Social Science Research Texts*. London/New York: Continuum.

MICUSP – *Michigan Corpus of Upper-level Student Papers.* (2009). Ann Arbor, MI: The Regents of the University of Michigan. http://search-micusp.elicorpora.info/simple/

Povolná, R. (2009). On Contrastive Relations in Academic Spoken Discourse. *Brno Studies in English* 35, 103-115.

Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Schmied, J. (2007). Complexity and Coherence in English Student Writing, Especially in Hypertext Learning Systems. In: J. Schmied, C. Haase & R. Povolná (eds.), *Complexity and Coherence. Approaches to Linguistic Research and Language Teaching*. Göttingen: Cuvillier, 13-30.

Scott, C. M. (1984). Adverbial Connectivity in Conversations of Children 6 to 12. *Journal of Child Language* 11, 423-452.

Ventola, E. & A. Mauranen (1991). Non-Native Writing and Native Revising of Scientific Articles. In: E. Ventola (ed.), *Functional and Systemic Linguistics*. Berlin: Mouton de Gruyter, 457-492.

Vergaro, C. (2008a). Concessive Constructions in English Business Letter Discourse. *Text & Talk* 26, 97-118.

Vergaro, C. (2008b). On the Pragmatics of Concessive Constructions in Italian and English Business Letter Discourse. *Multilingua* 27, 255-283.

Appendix – absolute and relative frequencies for selected concessive / contrastive markers

Reference Corpora

	though	*even though*	*although*	*nevertheless*	*whereas*	*however*	*on the other hand*
MICUSP all L1	1,008	213	1,081	115	184	2,685	219
per 10,000 words	3.18	0.67	3.53	0.38	0.62	9.58	0.75
MICUSP Linguistics (L1 only)	88	17	87	6	12	195	16
per 10,000 words	5.11	0.99	5.05	0.35	0.7	11.33	0.93
MICUSP English (L1 only)	253	31	89	16	20	280	21
per 10,000 words	8.53	1.05	3	0.54	0.67	9.44	0.71
BAWE total	2,023	617	4,500	597	1,081	12,189	836
per 10,000 words	3.08	0.94	6.85	0.91	1.65	18.57	1.27
BAWE Linguistics (L1 only)	35	16	177	6	55	179	6
per 10,000 words	2	0.92	10.13	0.34	3.15	10.24	0.34
BAWE English (L1 only)	92	24	147	11	23	301	29
per 10,000 words	3.96	1.03	6.32	0.47	0.99	12.95	1.25
BAWE all L1	1,364	348	3,401	266	728	8,874	345
per 10,000 words	2.96	0.76	7.38	0.58	1.58	19.26	0.75
BAWE all non-L1	660	271	1,115	330	358	3,334	489
per 10,000 words	3.36	1.38	5.68	1.68	1.82	16.98	2.49
COCA academic	23,853	6,981	48,206	6,738	8,693	75,385	7,177
per 10,000 words	2.88	0.84	5.81	0.81	1.05	9.09	0.87
BNC academic	5,400	1	10,656	2,375	2,221	18,772	1,784
per 10,000 words	3.52	0.0007	6.95	1.55	1.45	12.24	1.16

German & Czech corpora

	though	*even though*	*although*	*never-theless*	*whereas*	*however*	*on the other hand*
BA year 2 CUT Ling. term papers	33	20	42	20	51	109	17
per 10,000 words	4.37	2.65	5.56	2.65	6.75	14.44	2.25
BA Theses CUT Linguistics	83	41	69	37	95	338	43
per 10,000 words	5.78	2.85	4.8	2.57	6.61	23.52	2.99
Magister Theses CUT Linguistics	236	114	579	404	405	1123	123
per 10,000 words	3.61	1.75	8.87	6.19	6.2	17.2	1.88
Magister Exams CUT Linguistics	30	22	40	24	61	176	33
per 10,000 words	2.89	2.12	3.86	2.31	5.88	16.97	3.18
Magister Theses CUT Cult&Lit	88	67	168	83	96	404	73
per 10,000 words	3.01	2.29	5.74	2.84	3.28	13.81	2.5
MagTheses Brno Linguistics	86	38	99	85	18	178	86
per 10,000 words	4.83	2.13	5.56	4.77	1.01	9.99	4.83
MagTheses Brno Literature	43	26	120	56	20	202	41
per 10,000 words	2.41	1.46	6.73	3.14	1.12	11.34	2.3
MagTheses Brno Methodology	52	34	66	41	12	199	50
per 10,000 words	2.55	1.67	3.24	2.01	0.59	9.75	2.45

Modal Indeterminacy and Evidentiality in Adverbial Expressions: A Culturome in Academic Writing?

Christoph Haase

Abstract

In this study a corpus of academic English is queried for modal items in permutations of adverbs and auxiliaries. The resulting data will enable a quantitative view on collocational strength between these items which further support a theory of a complex relationship between truth-value commitment on the side of the author (also known as modality) and evidentiality as an independent (i.e. not commitment-dependent) way to express a scientific fact. The corpus is comprised of natural-science texts in academic and popular-academic variants; for the collocational strength a new measure is introduced. A possible shift in academic culture(s) is briefly contemplated.

1. Introduction

This contribution looks into collexematic distributions of the three most frequent modal adverbs in English, *certainly, probably* and *possibly*. These adverbs do not exhaust the framework of truth-value-modifying adverbs in English (in fact, they are all relatively positive and range +>50% in commitment) in which a number of others like *hardly, barely* or *impossibly* explore the lower end of the probability spectrum. The adverbs in the following study were selected because they are similar in scalarity, they appear alternatively with sentence scope and with propositional scope (in clause-initial and in clause-medial positions), and they are the most frequent of all modal adverbials. Moreover, they are unambiguous as all of them all epistemic, a feature they share, among all modal markers, only with the class of epistemic parentheticals like *That's right, I guess.*(Brinton & Brinton 2010: 171), also known as comment clauses.

Furthermore, the three modal adverbials targeted in this study are also chosen by Nuyts as essential epistemic markers in his standard treatment of epistemic modality (Nuyts 2001). As sentential adverbs they form a "tight lexical field of epistemic sentence adverbs" (Aijmer 2002: 17). This function raises the question whether (academic) cultures can be profiled by finding shifts in usage frequency in the respective text types.

Modal adverbs can be deontic only in combination with other deontic modal markers such as auxiliaries, cf.

(1) a. Certainly, you must leave now

and even here, the deontic meaning depends on the knowledge of the speaker. When they are semantically linked to a modal auxiliary, they add to the meaning not in a componential way, the overall truth value is not a computation of the truth values of the modal elements (see Allerton 2010: 149). Otherwise,

(1) b. ???Possibly, you must leave now

and the numerous permutations that can be formed, would be acceptable.

For the study this means that a narrow window of truth-preserving utterances is expressed by these adverbs and that the subsequent verbal action needs closer investigation should the choices speakers make in distributing these items be dependent on what is expressed by the action they have scope over. In the study I try to show that modal adverbials in academic texts are a pervasive and argumentation-structuring means of the commitment of the author, a means whose relevance is still on the increase and that bridges a linguistic gap between different text types in academia. The study looks at means to find a semantically plausible way of combining the modality of adverbial expressions with auxiliary expressions and suggests a novel way to explain evidentiality. This further supports a profiling of science cultures and the information transmission within (the culturome).

2. Modal adverbials in academic writing

2.1. Distributional features

Modal adverbials quantify and modify the degree of likelihood of a verbal utterance to come true due to the speaker' knowledge (epistemic modal adverbials) or due to the way the speaker came to this assessment (evidentials) (Cornillie 2010: 301). *Certainly*, *probably* and *possibly* are considered

> the 'purest' expressions for epistemic modality, in the sense that they are the most precise and specific means available for marking the degree of likelihood of a state of affairs: on the epistemic scale, certain(ly) is at the extreme positive end, probable/probably is more or less in the middle on the positive side of the scale, possible/possibly is near or at the neutral point, in the middle between the positive and the negative side of the scale, etc. (Nuyts 2001: 55).

Their use in academic prose is well-documented (cf. Haase 2008) and undisputed. In sentence-medial positions, the distribution of these adverbials is time-course-balanced and different genres show unspecific distribution profiles, as the Google *n*-gram survey shows (cf. Figure 1).

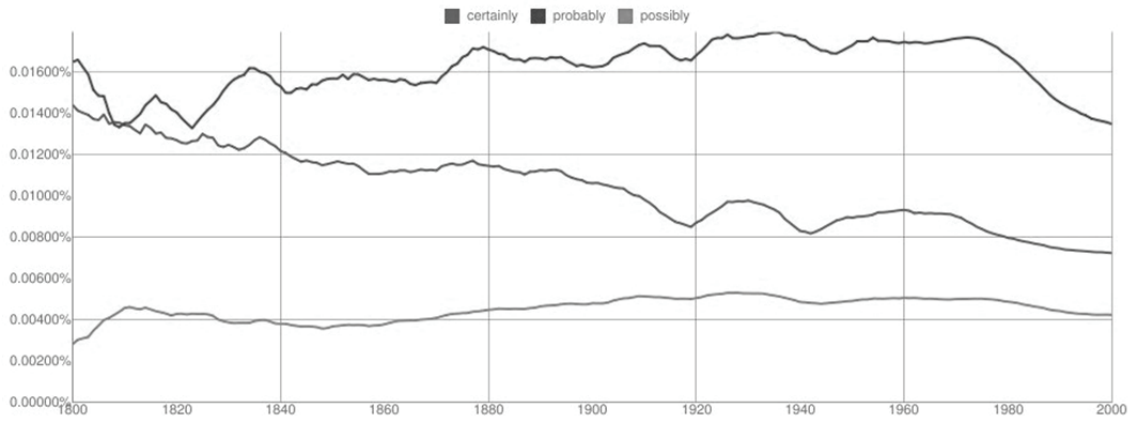


Figure 1: Distribution of epistemic modal adverbials 1800-2000 (http://ngrams.googlelabs.com)

The figures for the books queried include 5.2 m books published in English between 1800 and 2000, a corpus with approximately 361 billion words (Michel et al. 2010: 1). The books were supplied by 40 university libraries and contain academic texts but also special interest publications (but not fiction). The graph shows a marked decrease of *certainly* for which a monocausal explanation could be an increase in more differentiated, less polemic argumentation of modern academia. This change of culture corresponds with a change in a "culturome" and is epitomized in the recognition of the two cultures by Snow (for an account of the "two cultures" in conceptual metaphor in corpora cf. Haase 2010). A culturome is an information package with Dawkinsian properties of self-replication defined parallel to the more conventional terms of genome (for genes) and meme (for ideas).

This discussion falls within influential trends in the philosophy of science, summarized under the term supervaluationism (cf. Akiba 2000: 364) in which an attributed vagueness is in fact rather "in the world" than "in language" and any expression in language is scoped by a tacit (and therefore often unexpressed) quantifier of modal indeterminacy. This is a reformulation of a sorites paradox that operates with a Fregean law of the excluded middle and in which the balance from complete rejection to complete conviction is tipped by just one single item of data (see also Keefe 2003: 22). However, the spell-outs of this indeterminacy can be discerned in the peaks of this change in academic culture. This explains the highly frequent use of epistemic modal adverbials rather than that of evidentiality markers (such as *seemingly, evidently, obviously*) although the rise of *obviously* begins around 1910 (not graphed) and in current writing ranks between *certainly* and *possibly*.

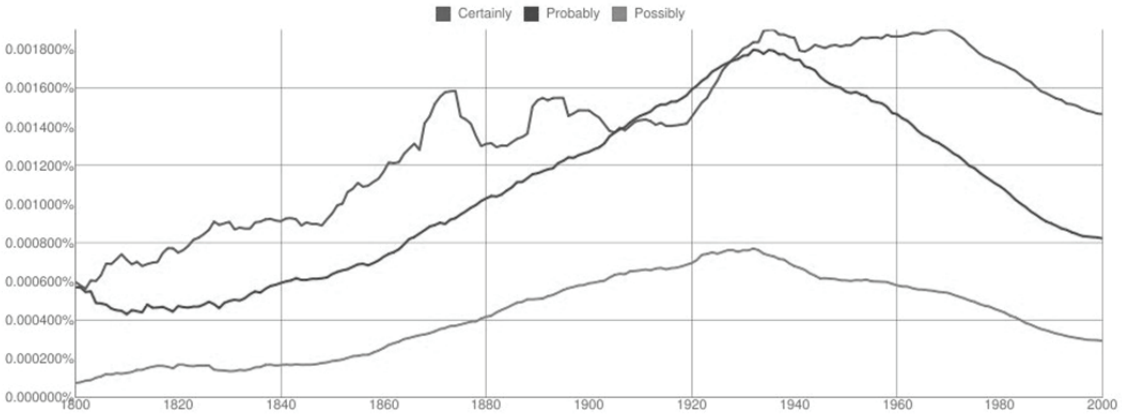


Figure 2: Distribution of epistemic modal sentence adverbials 1800-2000 (http://ngrams.googlelabs.com)

Thus, we can ascertain that the culturome changes again by blending epistemic and evidentiality expressions as the "official epistemology" of modern science is empiricism – "the doctrine that our knowledge is justified by experience – observation, data collection, experiment" (Rosenberg 2005: 138).

The evidentiality markers support this by opening the positivist window on research that is at the same time objective but allows disagreement for want of better results or methodologies. Thus, the use of *obviously* signals that the researcher looks at data rather in a metaphorical way.

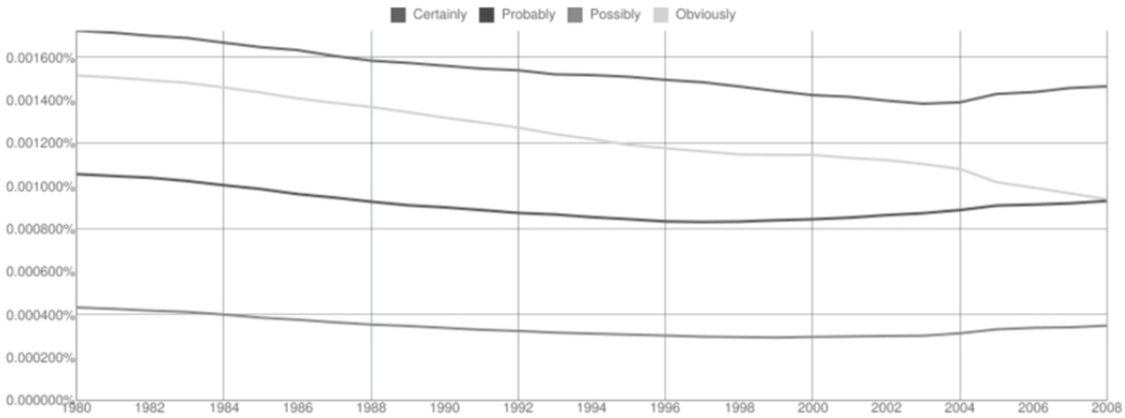


Figure 3: Distribution of epistemic and evidential modal sentence adverbials 1980-2008 (http://ngrams.googlelabs.com)

The effect classified in example (1a.) is frequent in spoken speech and can also be found in all types of non-academic writing because the combination of epistemic and deontic modality in fact creates evidentiality, = the epistemic conviction creates a rationale for the deontic (or root) meaning of the utterance. In any academic argumentation, this rationale seems ill-founded on the grounds of logic: ('It must be true because I know/am convinced it is so').

This breaks down into the following tasks answerable by the corpus:

Find all modal adverbial *n*-grams with left or right bounding modal auxiliaries.

This enables us to look at expressions like

(2) a. Certainly, this trend should stabilize
b. This trend should certainly stabilize

in which a "harmonic" use of a double-pattern construction of coherent modality whereas a so-called "disharmonic" use creates examples with borderline acceptability:

(3) a. Possibly, the data must show
b. The data must possibly show

This unusual combination can therefore not be called epistemic or double-epistemic. A treatment of these forms will be suggested in the final section.

2.2. Hypotheses

Harmonic and disharmonic uses are hypothesized to be present in the corpus in different distributions. The default, harmonic use should occur with higher frequency. The disharmonic use should be used for stylistic and/or evidential effect. The ratios may indicate characteristics of a culturome in academic writing.

The "Harmonic use" (in a terminology suggested by Halliday and Lyonsian, quoted from Cornillie 2010: 302) leads to the following queries:

a) *may/might/can + possibly*
b) *will/would/should + probably*
c) *can/shall/will/must + certainly*

Disharmonic uses to be queried are:
d) *may/might + certainly*
e) *must/shall/will + possibly/probably*
f) *can + certainly* will be excluded from query due to its deontic meaning.

The corpus is queried for collexematic expressions of permutations of harmonic and disharmonic uses of modal adverbials with modal auxiliaries in which the syntactic requirement is that both occur in the same clause but in which the modal adverb does not have necessary scope over the auxiliary. Thus, a stacking of modal operators as it is allowed in some languages (but not English) does not lead to a differentiated computation of modality.

3. Data and discussion

3.1. Data materials

The corpus queried for the study is called SPACE, the corpus of Specialized and Popular ACademic English, a project that was started in 2007 at Chemnitz University of Technology. In its present gestalt the corpus is comprised of nearly 900,000 words from different academic fields, primarily from the "hard" (physics) and the "soft" sciences (biology) with sub-domains in quantum physics, cosmology and particle physics on the hard science side and in microbiology, biochemistry and genetic engineering in the biological/life sciences (cf. Table 1).

subcorpus	descriptors	word count
arXiv	physics, astrophysics, computer science, quantum mechanics	288,861
New Scientist – physics	physics, astrophysics, computer science, quantum mechanics	38,315
Proceedings of the National Academy of Science (PNAS)	biochemistry, genetics, genetic engineering, microbiology	270,669
New Scientist - biosciences	biochemistry, genetics, genetic engineering, microbiology	31,125
Public Library of Science – Medicine (PLoS), New England Journal of Medicine, Journal of Clinical Investigation	medicine, virology, clinical psychology, public health	217,254
New Scientist – medicine	medicine, virology, clinical psychology, public health	17,050
total		**863,247**

Table 1: SPACE corpus 2007-2010

The unique feature of the corpus is the parallelism of a number of original science papers from pre-print servers like arXiv with an equal number of popularizations of these papers published by science journalists in the New Scientist. For this study, the components on medicine were ignored to ensure the hard/soft parallelism which makes the total of words considered 628,970 words. These components are comprised of 45,683 different words; the type-token-ratio is 0.08 or 7.6%.

Several studies (cf. Haase & Schmied 2008, Haase 2008) investigated details of distributional features. The current study falls within a larger project that studies markers of author commitment and hedging in academic writing.

Hedging in academic texts is a common means of speaker distancing from the truth-value of his/her utterance proposition as in many cases the results of research discussed depend on a narrow set of circumstances which may not hold

for a generality of conditions. Therefore, a hedged statement signals a cautious, responsible investigator. This implies a strategic use of vagueness. Thus, as Jucker et al. observe, "(v)arying the level of vagueness may help guide the addressee to make the intended representation of entities and events and to draw intended implications from them" (Jucker, Smith & Lüdge 2003: 1739).

The role of hedging changes when the author is a science journalist who is not him/herself responsible for the research results. Here, hedging signals the not immediately affected observer who merely summarizes other people's results. Both functions will be discussed in the following paragraph.

3.2. Case studies of English modal adverbials

The less prominent semantics of hedging in *certainly* expresses a clear author commitment for the truth value of the verbal action that follows. This is overly clear in all examples with sentence-scope (0047NS) but also when the adverbial is left-scoping to the verb. In the cases of the adverbial post-positioned to the verb, the authors saw the need for downtoning the statement in "almost certainly".

(4) They also have important implications for understanding - and copying - the chemistry of water splitting. "**Certainly**, several mechanistic proposals in the literature can be thrown in the bin," says Styring. (SPACE 0047NS)
(5) Organic techniques **certainly** have their benefits, especially for poor farmers. (SPACE 0056NS)
(6) ... one-size-fits-all approach to farming whether the rigid application of organic standards or an insistence on large-scale, high-input cropping is almost **certainly** a bad approach. (SPACE 0056NS)
(7) ... gray matter, but they do not differ by enough to account for our results, even if they varied systematically with brain size, which they almost **certainly** do not. (SPACE 0054PN)
(8) Translation increasingly shows itself to be an RNA-defined mechanism (3033). Its primordial form almost **certainly** arose in some sort of RNA-world context. (SPACE 0075PN)
(9) Most of them were nonneuronal (endothelial cells and cells in the white matter), but neurons were **certainly** labeled, especially in the hippocampus and cerebral cortex. (SPACE 0095PN)

It is interesting to note that in the randomly selected examples from the academic part of the corpus (PN) the downtoning is more systematic.

Probably marks probability with a clear positive likelihood that locates the author commitment close to certainty but allows for deviation in the results. The belief expressed to the best knowledge corresponds with auxiliary *will* or should. In sentence-initial position, the entire proposition is subjected to this given (and unmitigated) level of probability (see examples 10-14).

(10) These parts of spacetime are in the region with the curvature greater than the Planck value. **Probably** the Classical General Relativity is not applicable here (see below, section 6). (SPACE 0032AX)
(11) This means that from the point of view of semiclassical physics a singularity arises here. **Probably** at this stage the black hole has the characteristics of an extreme black hole, when the external event horizon and internal Cauchy horizon coincide. (SPACE 0032AX)
(12) We assume that M = N = 1000M, since on average web pages **probably** do not contain more than 1000 Google search terms. (SPACE 0032AX)
(13) From computer models based on what's known about the behaviour of protons and neutrons in nuclei, the researchers predict that tetrahedral nuclei **probably** pop up all over the periodic table. (SPACE 0021NS)
(14) We discovered that any driver of acceleration other than a cosmological constant can **probably** allow life to persist indefinitely," says Freese. Krauss and Starkman's pessimism isn't beaten yet, though. (SPACE 0023NS)

The semantics changes with the right-scoping adverbials. Here, the difference is that in postposition to an auxiliary, the truth-value of the proposition is enhanced by modifying the likelihood projected by the auxiliary. This is an asymmetric construction since *probably can* is much less likely to occur (and in fact to logically compute) than *can probably*.

The epistemic value of *possibly* is obviously related to the presence of unreliable conditions, insufficient data or information with temporary validity. This can be summarized as "inexact knowledge" (Keefe 2003: 64-65) which means "there is a margin for error principle at work. If your true belief that *p* is to count as knowledge, it should not be true just by luck." (ibid).

(15) ... small fold changes between 1.3 and 1.1 (representing 10-30% decreases in mRNA levels), but were statistically significant (P<0.05, and pp = 60%). **Possibly**, these alterations reflect decreased protein synthesis as an adaptation to impaired ATP production. (SPACE 0085PN)
(16) As a whole, CR affects the expression of more genes in the heart than the aging process (21% vs. 10%). **Possibly**, alterations in the expression of genes involved in specific transcriptional classes, such as ... (SPACE 0085PN)
(17) ... claimed that in three weeks, a family group could gather more grain than it could **possibly** consume in an entire year (28). (SPACE 0094PN)
(18) It's also solid, but has crystallised differently, **possibly** reflecting two episodes of inner core development, the researchers write in Proceedings of the National Academy of Sciences. (SPACE 0077NS)
(19) His-337 and Asp-189 (or His-190) are also **possibly** coordinated to the Mn cluster; weak electron density connections between them and the Mn cluster can be seen at a 0.2level. (SPACE 0047PN)
(20) The results suggest that establishment of an altered cellular environment by PBA, albeit for a limited period, can extend the lifespan of flies, **possibly** by inhibiting the accumulation of damage, and or stimulating repair mechanisms. (SPACE 0049PN)

The examples show two isolated uses of *possibly* as a sentential adverb. Both come from the same text (i.e. the same author(s)) so that an ideolectal use can be attested. The other modal uses are all pre-positioned to the main proposition,

often in participial or gerundial variation. Collexematic occurrences with *could* (0094PN) and *can* (0049PN) are frequent.

3.3. Data Discussion

The study considers at first baseline data from the BNC which show the ranks and absolute frequencies of the four items. The distribution of modal adverbials in a representative subsection of English as a means of comparison shows a first surprise in the case of the evidential (cf. Table 3).

item	*probably*	*certainly*	*obviously*	*possibly*
rank	364	560	915	1342
absolute freq	27,303	18,647	11,014	7,211

Table 3: Absolute frequencies and ranks of adverbial items in SPACE

The evidential has the lowest mean frequency in the SPACE subcorpora. The relative frequencies correspond with Zipf's law applied to the BNC counts and to the SPACE counts gives an indication of normalized frequencies. Zipfian distributions are especially interesting for comparably small corpora and infrequent phenomena (Gilquin & Gries 2009: 9) so the data in Table 4 are revealing.

item	*probably*	*certainly*	*obviously*	*possibly*
BNC	0.0027	0.0018	0.0011	0.0007
SPACE	0.0013	0.0003	0.0002	0.0006

Table 4: Normalized frequencies of adverbial items, BNC and SPACE

With the exception of *possibly* (the low counts of *possibly* surprise; in the SPACE corpus we are looking at 2 (two) occurrences with possibly as a sentential adverbial, both in the same text), all epistemic markers are two to six times more frequent in the BNC than in SPACE, the evidential is five times more frequent.

The relationship between modals and evidentials observed by Flowerdew suggests a differentiation into written and spoken genres. In his study on biology texts and lectures he shows that the "measure of involvement (expressed by modal items – CH) is much stronger in lectures" - with a modal written/spoken ratio of 0.6 (calculated from Flowerdew's data) (Flowerdew 1993: 83).

Conversely, this could mean that evidential are an increasingly written phenomenon – a surprise, given that in a lecture there are demonstrative aspects involved and an item like *obviously* can have abstract bus also very concrete referents.

Remarkable support for this comes from Hyland's broadband study of features (Hyland 2010) in which he, among other features trained on Biber's feature groupings (Biber 1988), summarizes modal auxiliaries under hedge expressions. Hyland's data compare nicely with SPACE as he maintains a parallelism between masters dissertations and doctoral dissertations, which in text type correspond roughly with two-way specialized and popular-academic stratification of the SPACE corpus. The following table compares the ratio of evidentials and hedges from Hyland's data (Hyland 2010: 132) with the corresponding text types in SPACE.

	Hyland *f* per 10,000			SPACE *f* per 10,000			
	masters	doctoral	all	popular	specialized	all	
evidentials	40	76.2	64.1	0.14	0.32	0.3	*obviously*
hedges	86.1	95.6	92.4	5.9	2.64	3.00	epistemics
ratio	0.46	0.79	0.69	0.024	0.12	0.1	ratio

Table 5: Hyland data for hedge–evidentiality ratios vs. SPACE data modality–evidentiality ratios

The relevant figure in Table 5 is the ratio. Hyland considers all hedge expressions and all evidentials whereas the SPACE study registers 3 (three) epistemic modal adverbials and one evidential adverbial. The ratio is further tilted because by design, 3 times more epistemics than evidentials were considered than epistemics. Thus, the higher relative frequency of epistemics for popular texts surprises as in Hyland's data, the doctoral dissertations show a slightly elevated modal signature.

Multi-word expressions with at least two modal items (one adverbial, one auxiliary) can be ranked according to their collocational strength. For large corpora, *MI* scores are probably the best metric (see Gilquin & Gries 2009). In this study, the attraction/repellation indices for the SPACE data did not render significant results due to the relatively small size of the corpus and the relatively low number of collexematic occurrences. As an alternative metric, Daudaravicius' and Marcinkevicienė's lexical gravity was used. Lexical gravity *G* yields a better scoring for small data sets and more precise collocational attraction. As real gravity, *G* expresses an attractive force, unlike real gravity, it can also be negative, i.e. repelling. Lexical gravity is calculated

$$\text{Gravity } G(\text{word}_1, \text{word}_2) = \log(f(\text{word}_1, \text{word}_2)(\text{type } f(\text{afterword}_1)/f(\text{word}_1))+ \log(f(\text{word}_1, \text{word}_2)(\text{type } f(\text{afterword}_2)/f(\text{word}_2))$$

(Daudaravicius & Marcinkevicienė 2004: 331)

The query set creates a 3x8 matrix of permutations of harmonic and disharmonic uses. The raw counts for all collocations with a span of one (left and right of the target word) are, at first glance, depressingly low (cf. Table 6).

	possibly	*probably*	*certainly*
can	3	3	1
could	4	1	0
may	2	1	0
might	n.d.	n.d.	n.d.
must	0	0	2
should	0	1	1
will	1	6	0
would	0	3	2

Table 6: Modal item permutations in SPACE (raw); *n*=+/-1

An intuitive trend that follows from this is that *possibly* likes to go together with *can* and *could*, *probably* beyond significance with *will* and *would* and *certainly* repels most direct modal neighbors. Extending the collexical frame is a natural next step. Here, decisive and in fact conclusive evidence can be obtained when we extend the frame and raise the *n*-gram threshold to *n*=4; a threshold suggested by Gries & Mukherjee (2010: 522) we obtain the following values for collocational strength (cf. Table 7).

	possibly		*probably*		*certainly*	
	attract	**repel**	**attract**	**repel**	**attract**	**repel**
can	5.45	0.027	2.04	0.479	18.033	0
could	6.29	0.023	0	127.90	2.91	0.04
may	4.45	0.033	0	30.17	0	0
might	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
must	0	0	2.89	900.04	3103.93	0
should	0	0	2.04	0	0	11.70
will	3.14	0.047	0	0	5.55	0
would	0	0	2.89	0	1.00	0

Table 7: Lexical gravity and negative gravity for modal items in SPACE

The results show a considerable skew in the data towards a collexematic attraction of *can*, *could* and *may* to *possibly*, an extreme outlier in *must* and *certainly* (not entirely unexpected) but raised beyond a "normal" attraction of 5.5 (suggested as the value du jour by Gries & Mukherjee 2010: 530; without much rationale). The most interesting negative *G*s are those between *must* and *probably* and to considerable surprise, *could* and *probably*. Overall, the harmonic use can be empirically attested. Disharmonic uses like *may* + *certainly* have zero

attraction as have *must* + *possibly*. Elevated antigravity persists further between *should* + *certainly*. Few figures exceed the level of 5.5.

4. Conclusion

The examples discussed and the quantitative evidence strongly suggests that the harmonic use does not change the level of commitment in the modality of the proposition. The disharmonic uses are statistically found to be rare as is evidentiality (recall Table 5). The rise of *obviously* is matched by a rise in evidentiality by disharmonic forms in academic writing. A specific evidentiality profile of a culturome in academia cannot be attested. The disharmonic use is still a marginal phenomenon, the results, though statistically speaking significant (cf. the results from the collocational gravity tests in Table 7) represent simply a too-miniscule ratio of evidentiality.

Furthermore, the low counts for virtually all modal adverbs in sentence-initial position shows that this type of adverb has lost its function to give modal scope over the entire proposition.

The use of *obviously* in almost all cases for a science journalist is not an evidentiality marker. Instead, this *obviously* is an epistemic marker for the narrowly entrenched field of popularized research in which the results at hand are not investigated, in they do not depend on the direct and original research but on the domain knowledge of the popularizer who does not him/herself questions the nature and quality of the results.

A similar phenomenon was observed by Holmes, for *sort of*, as quoted in Aijmer (2002: 191):

> Holmes (1988a:94) regards sort of as 'a particularly interesting pragmatic particle because of the wide range of meanings it may convey' and introduces a useful distinction between the epistemic modal (evidential) and affective (interpersonal) function of *sort of*'.

This gives rise for the need of a necessary terminological split of evidentials into real evidentials and pseudo-evidentials.

Christoph Haase
English Department
Purkyne University
Usti nad labem/Aussig
christoph.haase@ujep.cz

References

Aijmer, K. (2002). *English Discourse Particles: Evidence from a Corpus*. Amsterdam: John Benjamins.

Akiba, K. (2000). Vagueness as a Modality. *The Philosophical Quarterly* 50 (200), 359-370.

Allerton, D. J. (2006). Verbs and Their Satellites. In: B. Aarts & A. McMahon (eds.), *The Handbook of English Linguistics*. Malden, MA/Oxford: Blackwell, 146-179.

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Brinton, L. J. & D. M. Brinton (2010). *The Linguistic Structure of Modern English*. Revised Edition. Amsterdam: John Benjamins.

Butler, C. S. (2008). 'Basically Speaking'. A Corpus-Based Analysis of Three English Adverbs and their Formal Equivalents in English. In: M[a]. de los Ángeles Gómez González, J. L. Mackenzie & E. M. González Álvarez (eds.), *Current Trends in Contrastive Linguistics. Functional and Cognitive Perspectives*. Amsterdam: John Benjamins, 147-176.

Cornillie, B. (2010). On Conceptual Semantics and Discourse Functions: The Case of Spanish Modal Adverbs in Informal Conversation. *Review of Cognitive Linguistics* 8 (2), 300-320.

Daudaravicius, V. & R. Marcinkevicienė (2004). Gravity Counts for the Boundaries of Collocations. *International Journal of Corpus Linguistics* 9 (2), 321-348.

Flowerdew, J. (1993). Variation across Speech and Writing in Biology: A Quantitative Study. *Perspectives* 5, 75-88.

Gilquin, G. & S. T. Gries (2009). Corpora and Experimental Methods: A State-Of-The-Art Review. *Corpus Linguistics and Linguistic Theory* 5 (1), 1-26.

Gries, S. T. & J. Mukherjee (2010). Lexical Gravity across Varieties of English. An ICE-based Study of n-grams in Asian Englishes. *International Journal of Corpus Linguistics* 15 (4), 520-548.

Haase, C. (2008). Linguistic Determinants of English for Academic Purpose. In: A. Frankenberg-Garcia et al., *Proceedings of the 8th teaching and language corpora conference*. Lisbon: Associaocao de Estudos de Investigacao Cientifica, 128-132.

Haase, C. & J. Schmied (2008). Clause Linking in Specialized and Popular Academic English: an Investigation into Scientific Texts from the Space Corpus. In: B. Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools, and Applications - State of the Art*. Frankfurt/New York: Peter Lang, 381-399.

Haase, C. (2010). Mediating between the "Two Cultures" in Academia: The Role of Conceptual Metaphor. *Discourse and Interaction* 1 (3), 5-18.

Hyland, K. (2010). Metadiscourse: Mapping Interactions in Academic Writing. *Nordic Journal of English Studies* 9 (2), 125-143.

Jucker, A. H., S. W. Smith & T. Lüdge (2003). Interactive Aspects of Vagueness in Conversation. *Journal of Pragmatics* 35, 1737-1769.

Keefe, R. (2003). *Theories of Vagueness*. Cambridge: Cambridge University Press.

Michel, J. B. et al. (2010). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331 (6014), 176-182.

Nuyts, J. (2001). *Epistemic Modality, Language, and Conceptualization. A Cognitive-Pragmatic Perspective.* Amsterdam/Philadelphia: John Benjamins.

Rosenberg, A. (2005). *Philosophy of Science* (2nd ed.). New York/London: Routledge.

Williamson, T. (1994). *Vagueness*. London: Routledge.

Interpersonal Metadiscourse in English and Italian University Lectures: a Cross-Cultural Analysis of Person Markers

Giuliana Diani

Abstract

It is commonly assumed that, in university lectures, it is the subject matter that is important, and the relationships between the participants are only secondary (Rounds 1987). Even though this may be true for most lectures, it does not mean that these relationships do not exist or that the subject matter is presented from a purely objective point of view. Recent research shows that university lecturers very often express an attitude towards the topic explained or discussed (Fortanet-Gómez 2004; Biber 2006).
In such a genre language use is inherently dialogic and the lecturer accomplishes an exchange of meaning through a strategic manipulation of rhetorical and interactive elements (Bamford 2000, 2005) and through the organization of the lecture into meaningful patterns, in order to be understood by his/her students. This paper presents the preliminary results of a comparative analysis of the rhetorical strategies used by lecturers in their negotiation of knowledge with students in university lectures in English and Italian. Using corpus-based methods, the present analysis attempts to compare the quantitative and qualitative use of interpersonal metadiscursive devices across the two languages. Quantitative-qualitative evidence collected from the corpora in each language is analysed and discussed with special attention for divergences indicative of language-dependent variables within a generic framework reflecting the national culture.

1. Introduction

In university settings, academic lectures represent the principal genre of instruction, a crucial means to communicate to students the contents of the subject matter as well as other course-related issues (Flowerdew 1994; Thompson 1994; Young 1994; Flowerdew & Miller 1997). Similar to other academic genres (e.g. textbooks), the primary purpose of academic lectures is to convey to the audience the knowledge base of a discipline; that is, university lectures are essentially a pedagogical process genre (Thompson 1994). As such, the main functions of lectures include: introducing key theories, concepts, and research (Young 1994); integrating ideas from previous lectures and readings to the current lecture (Thompson 1994); providing information that may be unavailable in textbooks; and explaining complex constructs through various examples (Young 1994). However, recent research shows that university lecturers are more than a genre of information imparting. They are value-laden

discourses in which lecturers not only disseminate information to the audience, but they also express their stance towards the topic explained or discussed (Fortanet-Gómez 2004a; Biber 2006). This expected activity of the lecturer is emphasised by Biber's (2006: 87) claim:

> Lecturers in university registers seem more concerned with the expression of stance than with the communications of facts. […] Instructors take advantage of their positions of power to convey their own opinions and attitudes. Thus, in addition to simply conveying information, teachers shape the ways that students approach knowledge, helping them to assess how statements are to be interpreted (e.g. whether they should be adopted as fact, criticized, or understood from a particular perspective).

The aim of this paper is to contrastively analyse person markers (i.e. first person pronouns) as metadiscursive features in English and Italian university lectures. In particular, attention is given to those occurrences in which a marking of stance is identified, i.e. the lexical choice of an affective or evaluative verb accompanying the pronouns. I intend to explore whether their use is susceptible to cross-cultural variation.

While the use of first person pronouns in written academic discourse has received quite a lot scholarly attention in English (e.g. Kuo 1999; Tang & John 1999; Hyland 2002), across different disciplinary fields (Hyland 2001; Harwood 2005), across cultures (English, German, French, Russian and Bulgarian – Vassileva 2000; English, French and Norwegian – Breivega et al. 2002; English, Ukrainian and Russian – Yakhontova 2006; English and Spanish – Martín Martín 2004; Lorés Sanz 2006; Mur Dueñas 2007; English and Italian – Bondi 2007; Giannoni 2008; Molino 2010), and in texts written by native and non-native speakers of English (Hyland 2002; Martínez 2005), little attention has been given to the topic in English academic discourse, with the exception of Rounds (1985, 1987a, 1987b) and Fortanet-Gómez (2004b) for academic lectures, and, to the best of my knowledge, no cross-cultural analyses of the issue have been conducted so far in lectures. It is this intercultural perspective, which will be the focus of the analysis presented here.

My interest combines a focus on contrastive rhetoric (Mauranen 2001; Connor 2002, 2004), with an interest in linguistic resources in English and Italian. From this perspective, a comparative dimension of the analysis seems important, one which would allow for the study of linguistic variation across cultures.

In the following, a discussion of person markers as a category of interpersonal metadiscourse will be included in section 2. The comparable corpora analysed and the methodological procedures followed are described in section 3. Section 4 presents the results drawn from the analysis of the corpora under investigation. Finally, in section 5 some concluding remarks are offered in the light of the results obtained.

2. Person markers as features of interpersonal metadiscourse

Metadiscourse has recently been defined as "the cover term for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a part community" (Hyland 2005: 37). Referring to oneself by means of a person marker may contribute to shaping the writer-reader/speaker-listener relationship and it can also help writers/speakers indicate what their own contribution and ideas are. In that sense, person markers can be considered to play an important role in the construction of the writer or speaker's stance. According to the definition presented above, then, they can be seen as metadiscursive features. They constitute a category of interpersonal metadiscourse in Hyland's (1999, 2000) taxonomy. Recently, Hyland (2004, 2005) and Hyland and Tse (2004) have proposed a change in the terminology adopting Thompson's (2001) label of 'interactive' and 'interactional' as replacements for the traditional Hallidayan distinction between 'textual' and 'interpersonal' metadiscourse. Following Hyland's latest proposal, person markers are classified as interactional resources, that is, those which correspond to the interpersonal side of the traditional Hallidayan dichotomy. The interactional dimension concerns "the ways writers conduct interaction by intruding and commenting on their message" (Hyland 2005: 62). More specifically, person markers reflect the degree of the writer or speaker's presence.

Although the present study aligns with the principle that metadiscourse categories are intrinsically interpersonal and ultimately aim to persuade the writer/speaker, I prefer to continue using the functional distinction of textual and interpersonal metadiscourse markers. However, person markers have not been considered a category of their own in other taxonomies (e.g. Vande Kopple 1985, 2002; Crismore et al. 1993; Dafouz-Milne 2003). Following the most recent and influential metadiscourse taxonomies, person markers are considered an independent category of interpersonal/interactional metadiscourse, as they greatly contribute to the relationship established between writer(s)/speaker(s) and reader(s)/listener(s).

3. Materials and Methods

The analysis is based on two small comparable corpora of university lectures in different disciplines (linguistics, psychology and economics) and in different cultural contexts (English and Italian). The corpora consist of 5 lectures each. The English data are taken from *MICASE* (*Michigan Corpus of Academic Spoken English*), (Simpson et al. 1999),(totalling 46,256 words), while the Italian data are assembled by the author at the University of Modena and Reggio Emilia (23,596 words).

The methodology adopted for this study combines a discourse and a corpus perspective. Discourse analysis contributes to the definition of pragmatic

functions of person markers, whereas corpus linguistics offers ways of looking at lexical patterns: in particular, using *WordSmith Tools* (Scott 1998), I studied wordlists and concordances. These were worked out by comparing corpora to each other. The software *WordSmith Tools* was used to carry out the quantitative analysis of person markers for the interpretation of the differences in the frequency and distribution of first person subject pronouns in the two corpora. This was supplemented with qualitative analysis.

The analysis carried out focused on:

a) tokens which expressed the speaker's self (English *I* and exclusive *we* – *we* for *I*, 'authorial' *we*, Wales (1996), and their Italian counterparts, *io* and *noi*) followed by a verb conveying a stance meaning;
b) pragmatic implications/discourse function in English and Italian.

In identifying verbal stance, I have followed Biber et al.'s (1999: 972-975) classification of the semantic categories of stance markers: i) epistemic stance, that is, the speaker's comments on the status of information in a proposition: certainty or doubt, actuality, precision, limitation, source of knowledge or perspective, e.g. *Since last year I* ***think*** *they have improved*; ii) attitudinal stance, that is, the reports of personal attitudes or feelings, e.g. *I* ***wish*** *it was Friday though*; iii) style of speaking stance, that is, the speaker or writer's comments on the communication itself, e.g. *I shall* ***argue*** *that a state that accepts integrity as a political ideal has a better case for legitimacy than one that does not*.

It must be pointed out that the verbal stance markers presented in this study are those that appear directly related to the use of the first person pronouns in the corpora. Biber et al. (1999) present a much wider taxonomy of stance markers, including adverbials, adjectives, nouns or verbs with subjects other than first person pronouns, which have not been considered here.

4. Results

The results are presented in two parts according to the objectives of the study. The first section (4.1) includes the results of person markers across the language corpora, i.e. first person pronouns when used in combination of verbs expressing stance. In the second section (4.2), attention is given to the stance verbs associated with the pronouns.

4.1. Person markers in English and Italian university lectures: a cross-cultural comparison

Person markers are realized by means of self-references (i.e. singular and plural first person pronouns and possessive adjectives referring to the writer/speaker). Only personal pronouns were considered in the present study. The first step of the analysis was to quantify all instances of first person pronouns in both

corpora. Working with the software program, *WordSmith Tools*, wordlists were generated and then concordances of all occurrences of English *I* and *we*, and their Italian counterparts (*io* and *noi*) were produced. Since the aim was to identify only those tokens of the speaker's self followed by a verb conveying a stance meaning, a manual filtering process was necessary to make sure that only such occurrences were included.

As regards the English pronoun *we* and its Italian counterpart *noi*, those instances in which they made reference to both lecturers and student listeners or to the discourse community as a whole were excluded, as they are considered to be inclusive engagement markers, thus performing a different function. The context was carefully analysed to include only those instances that refer to the lecturer(s).

English and Italian differ in the way first person pronouns are encoded in grammar. In English, subject pronouns are indispensable to identify the agent/actor of a given process. In Italian, person pronouns may be used, but they are most often omitted, as the subject is always signalled through morphology in the verb-ending. For that reason, a computational search for the Italian verb-endings **amo* (present indicative)/**emo* (future indicative)/**emmo* (conditional) was carried out to obtain all tokens of verbs in the first person plural in the Italian corpus. But that was not done for the first person singular as it has different forms of verb-ending. A thorough reading of the corpus was thus necessary to find its occurrences. So, although in this particular realisation of person markers pronouns are compared with verb-endings, the comparison is still valid as they can be considered functionally equivalent.

Table 1 indicates the distribution of first person pronouns used in combination with verbs conveying stance in the two language corpora. The data are reported according to the normalized parameter of tokens per 10,000 words.

first person pronouns + stance verbs	MICASE lecture corpus (tokens per 10,000 words)	Italian lecture corpus (tokens per 10,000 words)
I/Io	40.43	4.23
Exclusive *we/noi*	11.89	3.39

Table 1: Distribution of first person pronouns + stance verbs across the language corpora

The English corpus shows that *I* followed by verbs of stance is much more quantitatively prominent than *we* (40.43 v. 11.89). The same tendency was noted in a study by Fortanet-Gómez (2004b) on the use of first and second person pronouns in English university lectures, which showed that *we* was less frequent than *I* in that register. Because these findings are in contrast with Rounds' (1987a) earlier study on the use of personal pronouns in university mathematics lectures, Fortanet-Gómez suggests that university lectures are becoming more personalized and with more focus on individuals. As will be pointed out in this

study, the wide variety of stance verbs found in combination with *I* in the lectures under investigation would seem to support this idea.

Cross-cultural comparison with the Italian corpus shows that the distribution of first person pronouns is much more limited in the Italian corpus. The overall count of first person singular pronouns is more than nine times higher in English compared to Italian (40.43 vs. 4.23). On the whole, the Italian corpus shows more limited reference to first-person involvement. This is reflected in a low frequency of both 1st person singular and plural pronouns, which are attested at a similar percentage (*io* 4.23 vs. *noi* 3.39). It therefore seems that cultural traditions and conventions in the use of self-references are less strong in the Italian classroom teaching context.

The analysis of the use of personal pronouns in the corpora was carried out in relation to the linguistic features of the verbs accompanying the pronouns. Particular attention was given to the presence of epistemic and attitudinal verbs used as stance markers according to the semantic categories proposed by Biber et al. (1999: 972-975).

4.2. Verbs of stance associated with first person pronouns across the two language corpora

My focus here is on cross-cultural variation which can be looked at both across languages and within the same language.

I start by briefly studying internal variation in English. Table 2 shows the occurrences of verbs of stance associated with *I* and exclusive *we* in the MICASE lecture corpus.

I + verbs of stance	(%)	exclusive *we* + verbs of stance	(%)
think	24.06	be gonna	32.73
will	14.97	could	23.64
wanna	11.23	will	21.82
want	11.23	can	14.54
know	9.63	be going to	7.27
can	7.49		
be gonna	7.49		
could	6.42		
be going to	3.74		
guess	3.74		

Table 2: Percentage distribution of verbs of stance most frequently used in combination with *I* and exclusive *we* in the MICASE lecture corpus

The figures show that lectures rely heavily on verbal stance in combination with first person pronouns. However, differences emerge when we examine the choices of verbs following the pronouns. The table shows that the occurrences of

verbs of stance accompanying *I* are distributed over a much wider range of lexical items than those following exclusive *we*. The range is related to the presence of lexical verbs like *think*, *know, guess*, *want*, *wanna*, and also to the use of some modal and semi-modal verbs like *can*, *could*, *will*, *be going to/be gonna.* If we consider the stance verbs accompanying exclusive *we*, choices are limited to modal and semi-modal verbs like *can*, *could*, *will*, *be going to/be gonna*.

Considering the verbs of stance most frequently used with *I* in the corpus, *think* is the most frequent verb in percentage terms (24.06%). *Think* accounts for the greater number of epistemic stance markers in all lectures under investigation. Here is an example taken from the corpus:

(1) **i think** that the window offer is a once-in-a-lifetime opportunity […]

It should be noted that not all occurrences of *think* have been considered. For example, *I'm thinking* was excluded, since it expressed the cognitive process of thinking, rather than 'giving an opinion'. However, although the most usual meaning of *think* in the corpus is 'to provide an opinion', sometimes its use seems to be nearer to a filler, with the function of giving the lecturer time to think about what s/he wants to say, as in the following excerpt:

(2) so, that's uh, the stages are helpful **i think** in giving us a kind of sense of, things […]

Although this analysis could be applied to several other occurrences of *I think* in the corpus, it is difficult to distinguish their use from the one in which *I think* expresses an opinion. For this reason, all these occurrences have been included in the analysis.

Tokens of *I think* drawn from the computational analysis were closely examined to determine their discourse function. The analysis reveals that *I think* tokens were used by lecturers mainly to comment on the content of what is being said, thus conveying personal attitudes about course content, as is illustrated in the following excerpts:

(3) [...] in many ways you can argue that evolutionary theory is the ultimate environmental explanation of behaviour, cuz it says it's the environment, that determines selection criteria that determines how things happens. um, so **i think** this is a kind of complicated uh uh statement it's it's a little easier to be breaking, natural selection down into sort of what are the assumptions that are embedded in here [...]

(4) [...] and so what **i think** that the paper clearly does is it shows that the kinds of sophistication that are in the option value or the stochastic dynamic program modals are really quite important for problems where the incentives become sharply nonlinear at one point in time as they do, obviously with an early-out window [...]

Similarly, instances of *I know/I don't know* and *I guess* are used in the corpus as personal epistemic stance markers. *I don't know* is above all frequent in the

negative form (10 occurrences). The corresponding positive form *I know* occurs four times. For example,

(5) he's probably one of the, maybe the foremost uh kind of, **i don't know** if you'd call him a personality theory or theorist of developmental theorist [...]
(6) let's not confuse letters and sounds please that's something **i know** it's just it's a uh habit that even traditional historical linguists used to make [...]
(7) Isaac Newton's impact on physical science, is of the same magnitude as Darwin_ **i guess** you'd normally say it the other way, cuz people always assume Newton's the big guy.

The frequent use of *I don't know* in the lectures under investigation is confirmed in related work on the use of multi-word sequences in classroom teaching. For example, *I don't know* is one of the most frequent sequences of words in the study by Biber et al. (2004: 384) of lecturers' use of lexical bundles in university teaching. Biber et al. (1999: 1002) discuss *I don't know* and combinations with *I don't know* as lexical bundles or "recurrent discourse building blocks", i.e. bundles of words that show a statistical tendency to co-occur" (1999: 991). The following four-word bundles were found in the MICASE lecture corpus under investigation: *I don't know what*, *I don't know how*, *I don't know if*.

Attitudinal meanings are also expressed in the corpus through the combinations *I* + the lexical verb *want* and its phonological reduction *wanna.* These stance devices are present in the corpus with the role of framing the lecturer's self-motivated desire in classroom teaching, as is illustrated in the following excerpts:

(8) these are points that either aren't, mentioned, or that **i want** to highlight more or at least comment on [...]
(9) now in talking about biological approaches to behavior, **i** really do **want** to make a distinction […]
(10) now there are a couple of things that **i wanna** point out about these, uh, uh, about these assumptions [...]

However, a closer consideration of these occurrences shows that they are often used for functions other than the expression of personal desire. For example, the following excerpts illustrate the use of *I want* and *I wanna* for topic introduction, i.e. announcing the lecturer's intention to initiate a new topic:

(11) one of the issues that **i want** to particularly concentrate on today is the issue [...]
(12) what **i want** to do is to start taking about evolutionary theory [...]
(13) now i **wanna** spend a little bit of time talking about natural selection [...]

Similarly, some modal verbs used in combination with *I* that express intention in the corpus like *will* and *be going to/be gonna* (e.g. *i'll try to tell you*; *i'm going to* make *a slight modification*), are also used to announce the proposed plan of the class session or some future action, as in:

(14) one of the things **i'm gonna** try to do in the lecture is to try to [...]
(15) that's what this is showing, same thing for capital... alright well so much for getting to the Static Neoclassical Model **i'll, i'll** do that next time, using this aggregate production function. uh, and **i'll** also talk about the Lucas Model [...]

Returning to Table 2, another frequent stance marker accompanying *I* is the modal verb *can*. The occurrences of *I can* are used in the corpus mainly to indicate the lecturer's ability to carry out a procedure (and guide students through the steps of explanations), rather than expressing core stance meanings relating to epistemic possibility or personal permission. For example:

(16) if i started at time T, it's now time T plus delta. **i can** put a T on this one […] …
(17) and the reason that i broke this up into two pieces is so **i can** tr- substract the, first piece of this second, from the first piece of the first equation […]

The modal *could* was similarly used in the corpus, as in:

(18) let's think about this for a minute. <pause while writing> If uh, E, changes smoothly, as a function of time, **i could** make a first-order approximation […]
(19) and so if i knew the four key parameters **i could** calculate the probability that each individual would retire at age T […]

One interesting aspect of *I can*/*I could* occurrences is that they concentrate mainly in the two economics lectures. The outstanding number of their instances in those lectures may be related to the discipline. The 'rhetoric of economics' can be shown to rely on mathematical reasoning and philosophical argument (Bondi 1999).The examples above show that the lecturers proceed by reasoning on the range of possibilities that emerge from a state of affairs.

Let us now consider the occurrences of exclusive *we* followed by stance verbs. While the results offer a wider variety of stance verbs accompanying *I*, a very limited range of these markers accompany exclusive *we* in the corpus, including only modal and semi-modal verbs. The semi-modal *be gonna* proves to be the most frequent choice (32.73%), followed by *could* (23.64%), *will* (21.82%), *can* (14.54%) and *be going to* (7.27%). For example:

(20) What **we're gonna** do, in, today's lecture, is [...]
(21) they're talking about biological causes is what **we could** call the proximal cause. the proximal cause just
(22) **we'll** come to expectations uh, pretty soon next week [...]
(23) that is represented in the Latin alphabet by what **we call** the letter F. and therefore it is assumed […]
(24) okay so the nonstochastic terms **we're going to** call little-G […]

If we consider verbal expression of intention in the English corpus, we can see that the significant use of *will* following both *I* and exclusive *we* finds a reasonable correspondence in a choice between use of future tense as a signal of intentionality (e.g. *i'll tell you more about it later/today we'll spend all the time*

on the so-called option value models) and choice of expressing volition (e.g. *i'll try to make that clear as we go along*/*this part of our brain the hindbrain we'll talk about that when we get to brain physiology*).

Moving on to Italian, we find a picture of greater divergence in the range of stance verbs. An analysis of the verbs of stance accompanying the 1st person singular pronoun *io* reveals that the relevant choices found were more limited. If the English corpus suggests a fair balance between choices of lexical verbs and of modals following *I*, the Italian corpus shows that references to lexical verbs following *io* are by far preferred: more than half of the occurrences of *io* are largely dominated by the verb of volition *volere* [want] (55%), whereas one-fifth is realized through the verb *sembrare* [seem] (20%); the rest is divided between the verbs *preferire* [prefer] (15%) and *credere* [believe] (10%).

Most occurrences of *volere* serve a discourse organizing function, that is, they have the function of announcing the lecturer's intentionality and purpose, as in:

(25) Quello che **vorrei** fare oggi è una introduzione…
[what i would like to do today is an introduction to...]
(26) la prossima volta **vorrei** anche parlarvi….
[i would like to talk about... next time]

The same function was noted for many instances of *want* and *wanna* in the MICASE lecture corpus. However, differences were found in the verb tense. While Anglo-American lecturers favoured the use of the present indicative *I want* and *I wanna* for statements of purpose, Italian lecturers preferred the conditional form *vorrei* [i would like] as an oblique, hedged form.

When we analyse the pragmatic use of the other lexical verbs accompanying *io* in the corpus like *sembrare* and *credere*, we find a similarity between these verbs and those found in the English corpus (*think*, *know*, *guess*) in their discourse function. Like the English verbs, the Italian *sembrare* and *credere* tend to be associated with the discourse function of commenting on the content course on the part of the lecturer. Here are some examples from the Italian corpus:

(27) **Io credo** che questo sia un esempio veramente molto bello e illuminante…
[i think this is an impressive example...]
(28) quello che **mi sembra** di poter sottolineare è che…
[it seems to me that we can emphasize that...]

However, the quantitative data show that verbal reference to the making of comments is much more limited in Italian than it is in English. One reason may be that Italian lecturers are less inclined to emphasise their subjectivity when making claims and presenting their opinions. They tend to be considerably less visible than their Anglo-American colleagues.

If we consider the verbs of stance accompanying the 1st person plural pronoun *noi*, we find a similar picture across the two language corpora. Like the English corpus, the Italian corpus offers a very limited range of stance verbs following

the 1st person plural pronoun, which is only related to the use of some modal verbs like *dovere* and *potere*.

(29) La prima cosa su cui **dobbiamo** fare una breve valutazione è […]
[The first thing we must consider is...]
(30) **noi possiamo spiegare** questa situazione facendo riferimento […]
[we can explain the situation by taking into account...]

5. Concluding remarks

The findings of this study, while conditioned by the limited size of the corpora, offer evidence of cross-cultural variation in the use of person markers and the verbs associated with them expressing stance. The results indicate that the use of 1st person singular pronoun followed by verbs of stance is a common resource used by both Anglo-American and Italian lecturers in their university classroom. However, the frequency of first singular person pronouns has been found to be much higher in the English corpus than in the Italian one (40.43vs. 4.23). Interestingly, the frequency difference between the 1st person singular and plural pronoun *io* and *noi* in the Italian corpus is not outstanding (*io* 4.23 vs. *noi* 3.39), as it is in the English corpus (*I* 40.43 vs. *we* 11.89). The lower number of first person pronouns in the Italian corpus than in the English one could perhaps be explained in terms of the possible different conventional, traditional views on self-representation pertaining to the different national cultures (Atkinson 2004), in which the lectures are delivered. Italian culture may run counter to the use of features which emphasise one's authorial persona, whereas Anglo-American culture could prompt the use of self-mentions to a greater extent. Thus, it may be argued that what contributes to establishing a credible ethos and, consequently, what constitutes an appropriate self-representation of the speaker may differ across different national cultures; these differences, in turn, may influence the frequency of use of first person pronouns in oral discourse in general and in lectures in particular.

Besides the greater frequency of use of first person pronouns in the English corpus, the contrastive analysis has yielded significant differences in the range of verbs of stance accompanying the pronouns throughout the lectures in English and in Italian. The English corpus offers a much wider variety of these markers than the Italian one, which is related to the presence of some evaluative and affective verbs (*think*, *know*, *guess*, *want*, *wanna*,) and also to the use of some modal and semi-modal verbs (*can*, *could*, *will*, *be going to/be gonna*). In the Italian corpus, on the other hand, the lexical choice of stance verbs is more limited (*volere, preferire*, *sembrare, credere, dovere, potere*).

It should be pointed out, however, that similarity has been observed between the two corpora in the discourse functions of person markers. In both English and Italian, first person pronouns tended to be associated with verbs signalling lecturers' intentionality or purpose (English – *want*, *wanna*, *will*, *be going to*/*be*

gonna; Italian – *volere*, *preferire*) and verbs expressing personal opinions (English – *think*, *know*, *guess*; Italian *sembrare*, *credere*).

The specificity and small size of the corpora analysed here calls for more analyses that confirm the results obtained. Nevertheless, it can be tentatively concluded that the differences in frequency and distribution of use of person markers through which a lecturer is portrayed as an authorial self in a lecture could depend on the specific national culture. It would be interesting to investigate if this could also depend on the academic discipline.

Giuliana Diani
Dipartimento di Studi Linguistici sulla Testualità e la Traduzione
Università di Modena e Reggio Emilia
Italy
giuliana.diani@unimore.it

References

Atkinson, D. (2004). Contrasting Rhetorics/Contrasting Cultures: Why Contrastive Rhetoric Needs a Better Conceptualization of Culture. *Journal of English for Academic Purposes* 3(4), 277-289.

Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Biber, D., S. Conrad & V. Cortes (2004). If you Look at...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics* 25(3), 371-405.

Bondi, M. (1999). *English across Genres: Language Variation in the Discourse of Economics*. Modena: Il Fiorino.

Bondi, M. (2007). Historical Research Articles in English and Italian: A Cross-Cultural Analysis of Self-Reference in Openings. In: M. Bertuccelli Papi, G. Cappelli & S. Masi (eds.), *Lexical Complexity: Theoretical Assessment and Translational Perspectives*. Pisa: Pisa University Press, 65-83.

Breivega, K. R., T. Dahl, & K. Fløttum (2002). Traces of Self and Others in Research Articles. A Comparative Pilot Study of English, French and Norwegian Research Articles in Medicine, Economics and Linguistics. *International Journal of Applied Linguistics* 12(2), 218-239.

Connor, U. (2002). New directions in contrastive rhetoric. *TESOL Quarterly* 36(4), 493-510.

Connor, U. (2004). Intercultural Rhetoric Research: beyond Texts. *Journal of English for Academic Purposes* 3(4), 291-304.

Crismore, A., R. Markkanen & M. S. Steffensen (1993). Metadiscourse in Persuasive Writing: a Study of Texts Written by American and Finnish University Students. *Written Communication* 10(1), 39-71.

Dafouz-Milne, E. (2003). Metadiscourse Revisited: A Contrastive Study of Persuasive Writing in Professional Discourse. *Estudios Ingleses de la Universidad Complutense* 11, 29-52.

Flowerdew, J. (1994). Research of Relevance to L2 Lecture Comprehension: an Overview. In: J. Flowerdew (ed.), *Academic Listening: Research Perspectives*. Cambridge: Cambridge University Press, 7-29.

Flowerdew, J. & L. Miller (1997). The Teaching of Academic Listening Comprehension and the Question of Authenticity. *English for Specific Purposes* 16(1), 27-46.

Fortanet-Gómez, I. (2004a). Verbal Stance in Spoken Academic Discourse. In: G. Del Lungo Camiciotti & E. Tognini Bonelli (eds.), *Academic Discourse – New Insights into Evaluation*. Bern: Peter Lang, 99-119.

Fortanet-Gómez, I. (2004b). The Use of 'We' in University Lectures: Reference and Function. *English for Specific Purposes* 23(1), 45-66.

Giannoni, D. S. (2008). Medical Writing at the Periphery: The Case of Italian Journal Editorials. *Journal of English for Academic Purposes* 7(2), 97-107.

Harwood, N. (2005). 'Nowhere has Anyone Attempted ... In This Article I Am to Do Just That'. A Corpus-Based Study of Self-Promotional I and We in Academic Writing across Four Disciplines. *Journal of Pragmatics* 37(8), 1207-1231.

Hyland, K. (1999). Disciplinary Discourses: Writer Stance in Research Articles. In: C. N. Candlin, & K. Hyland (eds.), *Writing: Texts, Processes and Practices*. London: Longman, 99-121.

Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing*. London: Longman.

Hyland, K. (2001). Humble Servants of the Discipline? Self-Mention in Research Articles. *English for Specific Purposes* 20(3), 207-226.

Hyland, K. (2002). Authority and Invisibility: Authorial Identity in Academic Writing. *Journal of Pragmatics* 34(8), 1091-1112.

Hyland, K. (2004). Disciplinary Interactions: Metadiscourse in L2 Postgraduate Writing. *Journal of Second Language Writing* 13(2), 133-151.

Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.

Hyland, K., & P. Tse (2004). Metadiscourse in Academic Writing: a Reappraisal. *Applied Linguistics* 25(2), 156-177.

Kuo, C. H. (1999). The Use of Personal Pronouns: Role Relationships in Scientific Journal Articles. *English for Specific Purposes* 18(2), 121-138.

Lorés Sanz, R. (2006). 'I Will Argue That': First Person Pronouns as Metadiscoursal Devices in Research Article Abstracts in English and Spanish. *ESP across Cultures* 3, 23-40.

Martín, P. (2004). Personal Attribution in English and Spanish Scientific Texts. *BELLS: Barcelona English Language and Literature Studies* 12. http://www.publicaciones.ub.es/revistes/bells12/PDF/art09.pdf [20/07/2010].

Martínez, I. A. (2005). Native and Non-Native Writer's Use of First Person Pronouns in the Different Sections of Biology Research Articles in English. *Journal of Second Language Writing* 14(3), 174-190.

Mauranen, A. (2001). Descriptions or Explanations? Some Methodological Issues in Contrastive Rhetoric. In: M. Hewings (ed.), *Academic Writing in Context*. Birmingham: University of Birmingham Press, 43-54.

Molino, A. (2010). Personal and Impersonal Authorial References: A Contrastive Study of English and Italian Linguistics Research Articles. *Journal of English for Academic Purposes* 9(2), 86-101.

Mur Dueñas, P. (2007). 'I/We Focus on': A Cross-Cultural Analysis of Self-Mentions in Business Management Research Articles. *Journal of English for Academic Purposes* 6(2), 143-162.

Rounds, P. (1985). *Talking the Mathematics Thorough: Disciplinary Transaction and Socio-Educational Interaction*. Unpublished PhD dissertation. Ann Arbor: University of Michigan.

Rounds, P. (1987a). Multifunctional Personal Pronoun Use in Educational Setting. *English for Specific Purposes* 6(1), 13-29.

Rounds, P. (1987b). Characterizing Successful Classroom Discourse for NNS Teaching Assistant Training. *TESOL Quarterly* 21(4), 643-671.

Scott, M. (1998). *WordSmith Tools*. Oxford: Oxford University Press.

Simpson, R., S. L. Briggs, J. Ovens & J. Swales (1999). *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan. http://www.lsa.umich.edu/eli/micase/micase.htm.

Tang, R. & S. John (1999). The 'I' in Identity: Exploring Writer Identity in Student Academic Writing through the First Person Pronoun. *English for Specific Purposes* 18 (Suppl.), S23-S39.

Thompson, G. (2001). Interaction in Academic Writing: Learning to Argue with the Reader. *Applied Linguistics* 22(1), 58-78.

Thompson, S. (1994). Frameworks and Contexts: A Genre-Based Approach to Analysing Lecture Introductions. *English for Specific Purposes* 13(2), 171-186.

Vande Kopple, W. J. (1985). Some Exploratory Discourse on Metadiscourse. *College Composition and Communication* 36(1), 82-93.

Vande Kopple, W. J. (2002). Metadiscourse, Discourse and Issues in Composition and Rhetoric. In: E. Barton & G. Stygall (eds.), *Discourse Studies in Composition*. New Jersey: Hampton Press, 91-113.

Vassileva, I. (2000). *Who Is the Author*? *A Contrastive Analysis of Authorial Presence in English, German, French, Russian and Bulgarian Academic Discourse*. Sankt Augustin: Asgard.

Wales, K. (1996). *Personal Pronouns in Present-Day English*. Cambridge: Cambridge University Press.

Yakhontova, T. (2006). Cultural and Disciplinary Variation in Academic Discourse: The Issue of Influencing Factors. *Journal of English for Academic Purposes* 5(2), 153-167.

Young, L. (1994). University Lectures: Macro-Structure and Micro-Features. In: J. Flowerdew (ed.), *Academic Listening: Research Perspectives*. Cambridge: Cambridge University Press, 159-176.

Content and Wording of Academic Course Descriptions

Sara Gesuato

Abstract

Academic course descriptions (ACDs) are Informational-regulatory texts: they orientate students and impose rules on them. The analysis of 100 ACDs representative of 10 disciplines (10,304 words) shows that their communicative goals are realized through the choice of topics dealt with, and through the representation of processes and the entities involved in them. Thus, the ACDs outline the content, method and background of courses (referential function), but also their logistics/requirements and goals/outcomes (conative function). The focus of the ACDs is seemingly more on what courses are about (referential function) than who they are for or by (conative function): the courses, their content and activities are more frequently mentioned than the students and the teachers, and these participants are hardly ever represented as direct interlocutors. However, the ubiquitous simple present and *will*-future, which authoritatively represent events as non-negotiable arrangements, serves the same purpose as deontic modality. Overall, the ACDs present courses not as the product of the teachers' and students' joint efforts, but as self-determined entities responsible for their policies and content, which mediate between the teachers and the students by avoiding the confrontation implicit in the imposition of rules by the former onto the latter.

1. Introduction

Academic discourse studies mainly comprise three research domains: analysis of how scholars represent, validate and disseminate their research findings (e.g. Swales 1990, Bhatia 1993, Myers 2000, Bondi 2004, Diani 2004, Römer 2005, Soler 2007, D'Angelo 2009, Grazzi 2009), examination of how novices are socialized into the interactional practices of academic communication (e.g. Swales & Feak 1994/2004, 2000, Tognini-Bonelli & Del Lungo Camiciotti 2005, Hyland & Bondi 2006, Flottum, Dahl & Kinn 2006, Bruce 2008, Gotti 2009, Hyland 2009, Hyland & Diani 2009), and also description of how academics develop communicative practices in handling professional and social relationships, and in carrying out administrative tasks (e.g. Swales 1988, Fairclough 1993, Räisänen 1999, Trix & Psenka 2003, Giannoni 2006, Gesuato 2008, Gea Valor & Inigo Ros 2009, Bernardini, Ferraresi & Gaspari 2009).

This paper is meant to contribute to the third above-mentioned line of research by examining academic course descriptions (ACDs). These texts qualify as a hybrid institutional genre, partly housekeeping and partly gatekeeping (Srikant Sarangi, personal communication 2009). They serve two communicative purposes: a logistical-informational one (i.e. conveying information to prospective attendees

about the contents and structure of courses) and a pedagogical-regulatory one (i.e. imposing requirements that regulate access to and participation in courses). The analysis focuses on how the content, structure and wording of ACDs are influenced by their twofold goal, partly orientational and partly directive.

2. Data

The data analysed[1] comes from 100 ACDs, representative of 10 disciplines, collected on the Internet with the following search queries: "course description" AND "biology/ geography/ history/ journalism/ law/ literature/ music/ physics/ psychology/ statistics". The original texts comprise 63 short, one-paragraph long ACDs and 37 longer, multi-paragraph ones. The former outline the main topics and goals of courses and optionally refer to their policies. The latter – divided into such sections as format, grading, contact details and assignments –[2] provide detailed information about the content, structure, policies and logistics of courses; e.g.:

(1) **LIT 210(3-0-3)**
Children's Literature

This course defines the nature and function of children's literature by locating an examination of its history, genres, trends, and controversies in both an understanding of children's cognitive and imaginative response to reading and an exploration of culturally constructed images of and for children. The course offers methodologies for critical reading of a variety of children's texts and for selecting literature appropriate for a number of child-oriented programs. The course offers opportunities for observation and participation in story hours and other literature-based activities in locations such as child care facilities and public libraries.

PR: ENG 123 (Lit-01)

(2) **Introduction to Russian Literature**

Daily Assignments for the course - Fall, 2002 (including links to individual authors and works)

General Information about the Course:

LTRN 101 "Introduction to Russian Literature." An introduction to the works of major 19th and 20th century Russian writers (Pushkin, Gogol, Dostoevsky, Tolstoy, Gorky, Solzhenitsyn and others), including a discussion of how Russian literature is translated into other languages and other media (such as film, music and art). All readings are in ENGLISH.This course has been designed especially for first-year college students, but it is open to all students. The course may be used towards fulfilling the *Humanities Distribution Requirement*. This course also may be used towards fulfilling the requirements for the *European Studies Minor*.

[1] This work is an expanded revision of Gesuato (forthcoming), based on a larger corpus.
[2] Hist-02 consists of multiple typographic paragraphs, and is thus a long ACD, but is not divided into sections.

> The purpose of the course is twofold: to give students a basic introduction to Russian literature and to develop in them an appreciation of good literature. To meet these goals we will talk about literature in general and different ways of approaching a literary work, we will also analyze an array of Russian literature from both the 19th and 20th centuries. We will read not only prose, but poetry and drama as well; we will also read some literary criticism. The course will introduce students to several special topics, such as 1) problems of literary translation, 2) adapting literary works to radio, TV and film, 3) the current literary scene in Russia.
>
> **I. Required reading**
>
> A.S.Pushkin. *Eugene Onegin*. Penguin, 1979 [novel in verse]
>
> N.V.Gogol, et al. *The Government Inspector and Other Russian Plays*. Penguin Classics, 1991. […]We will also read some poems of various poets; excerpts from Dostoevsky's *Crime and Punishment*; excerpts from Tolstoy's *Anna Karenina* and *War and Peace* […] The cost of these Xeroxed materials is approximately $7. This collection of readings will be available for purchase by October 1st in the Department of Modern Languages and Literatures office, top floor Carnegie. Other handouts (class notes, study questions, etc.)--free! You should purchase a large three ring binder for these materials.
>
> **II. Media Presentations:**
>
> Performance "Eugene Onegin" (video of the opera),
> movie excerpts: "The Inspector General" (US and Russian versions) […]
> videos on the lives and works of various authors and other things...as available.
>
> **III. Lectures:**
>
> Brief lectures on literary theory--based mainly on the text by R. Wellek and A. Warren and that by R. Seldon (listed below). Also several brief lectures on special topics such as: Russian literary tradition; general characteristics of prose and poetry analysis; problems of translation; literature and the media; Socialist Realism; the current literary scene in Russia.
>
> **IV. Suggested Readings:**
>
> (Check the library or the bookstore for copies)
> R. Wellek and A. Warren. *Theory of Literature*.
> R. Selden. A Reader's Guide to Contemporary Literary Theory […]. (Lit-04)

For this study I considered the complete texts of the short ACDs and the main sections of the longer ones[3](identifiable by the title *Course description*, or if that was missing, alternative titles such as *Aims* or *Objectives, Course structure* or *Content*).[4] As Table 1 shows, the corpus consists of 10,304 words, the biggest subcorpus being Law, and the smallest Psychology. Short ACDs outnumber

[3] In the case of Hist-02, which consists of one long section with no overtly labelled internal sections (see previous footnote), I considered the whole text.

[4] Partial exceptions to the selection procedure apply to Journ-02 and Phys-04. Journ-02 consists of several sections, of which one called *Course Description* and one *Course Structure*. The former begins with three quotations – one from a journal, one from a report, and one from a Bob Dylan song – and then goes on to specify what the course in question is *not* about or for. The latter, instead, provides a description of the course. Because of its content, therefore, rather than its labelling, I considered the latter portion of the ACD. Phys-04 consists of several paragraphs. The only one that does not come with a title succinctly describes the course in general terms, and is thus the one I chose for examination.

longer ones in seven disciplines; the exceptions are Literature, Law (with an equal number of short and long ACDs) and Physics. The ACDs can consist of as few as 26 words up to as many as 1,521, the average length of an ACD being about 130 words.

My mainly qualitative analysis of the ACDs comprised: (a) a classification of their content (i.e. recurrent units of information were revealed by repeated readings of the texts); (b) a description of the semantic-syntactic encoding of the events represented, and the participants mentioned (i.e. the transitivity of predicates, and the syntactic functions and semantic roles of their arguments were manually labelled); and (c) an outline of the tone of the texts (i.e. the strength of assertions and impositions was assessed on the basis of recurrent tense and modality choices). The research goal was to outline the authors' conceptualization of courses and their orientation towards their readership.

Discipline	*short ACDs*	*long ACDs*	*global words*	*shortest ACD*	*longest ACD*	*average words per ACD*
Biology	9	1	990	32	154	99
Geography	8	2	1,033	26	404	103.3
History	6	4	1,181	37	284	118.1
Journalism	8	2	932	36	160	93.2
Law	5	5	1,521	52	328	152.1
Literature	3	7	1,253	50	223	125.3
Music	7	3	1,217	53	277	121.7
Physics	3	7	743	32	119	74.3
Psychology	6	4	714	35	198	71.4
Statistics	8	2	720	41	168	72
total	**63**	**37**	**10,304**	**--**	**--**	**128.8**

Table 1: Distribution of long and short ACDs across subcorpora, length of subcorpora in number of words, longest and shortest ACDs, and average length of ACDs in number of words

3. Data Analysis Picture

The ACDs convey various combinations of five types of information: about the content of courses, their goals/outcomes, logistics, methods and/or their disciplinary backgrounds. The content-based identification criteria of these information units, as well as their encoding options, are outlined below, together with illustrative examples.

Content comprises the topics to discuss and the issues to address in class.[5] It is either encoded as the target of others' (e.g. the teacher's, students' or, more often, the course's) action, or alternatively included in presentative constructions as the complement of a relational verb; e.g.:

(1) This class examines the relationship of numbers to news stories6 (Journ-07)
(2) A study of the structural organization and processes of eukaryotic cells. Topics of discussion will include regulation of transcription, gene product processing and transport, organelle biogenesis and function […] and cell interactions (Bio-06)
(3) Social environment, political, and religious institutions, and the main intellectual currents of the Latin West studied through primary sources and modern historical writings (Hist-01)

Goal/Oucome includes both educational accomplishments laid down for courses or their participants (i.e. purposes to be achieved) and expected experiential end-results of events, such as understanding or learning.[7] Reference to the former is often preceded by such phrases as *so as to, is designed to*. The latter, instead, are represented either as independent events (e.g. *understanding, learning to do X*) or as consequences of actions performed by courses for the benefit of students, and thus encoded as objects of predicates (e.g. *deepen an appreciation of X, offer an opportunity to do X*); e.g.:

(4) This course <u>will bring an understanding</u> of what it means to be "urban" […] (Geo-02)
(5) Students […] will be expected to discuss, analyze, and solve these problems (Law-02)
(6) Through two main projects […] students become fluent in the language, workflow and rigorous demands of Internet publishing (Journ-03)
(7) More than that, <u>I hope you will learn</u> how to write a complete solution to a problem, formulating your arguments clearly and concisely. (Phys-04)

Logistics comprises the organizational and technical aspects of courses (e.g. location of class meetings), and their policies (e.g. prerequisites).[8] It specifies what courses do or do not involve, and also what they offer to or impose on students. Instantiations of this information unit are characterized by reference to technical characteristics of courses (e.g. *there is no textbook; X credits; assignments will be graded pass/fail; offered in the Spring semester*) and by the use modal expressions, or equivalents, which introduce what can or must (not) be done (e.g. *to be required; requirement; can be taken as X; available*); e.g.:

---

[5] Text segments instantiating the information unit *content* may include reference to specific (sub)topics (e.g. "pharmacological treatments") or generic ones (e.g. "concepts").
[6] Here and elsewhere, unless otherwise specified, underlining is added to highlight the text segment conveying a given information unit.
[7] I counted the indication of what students will learn to do in, or thanks to, courses as an instantiation of the information unit *goal/outcome*, while I regarded the indication of what students study or master as notions (i.e. 'know what' as opposed to 'know how') – e.g. the object of study – as part of the information unit *content*.
[8] I did not count the titles of courses as instances of the information unit *logistics*.

(8) Course is repeatable as topic changes (Lit-02)
(9) Students are strongly advised to read as many as possible of the following works in advance: Nabokov *Despair; Lolita;* Bunin *The Dark Avenues* […] (Lit-05)
(10) This technology-oriented course is an equivalent prerequisite to the complementary course PHYS860 (Applied Science Topics in Micro/Nano-Technologies) […] (Phys-05)

Method refers to the approach adopted by a teacher to conduct lessons and/or the type of participation envisaged of students. This information thus specifies how the object of courses will be studied, and is meant to help readers understand the direction courses will take. It may include reference to procedural knowledge or selective attention paid to a particular aspect or part of a general topic;[9] e.g.:

(11) Students examine a variety of musical genres, develop critical listening habits and apply concepts presented in class through keyboard demonstrations (Mus-03)
(12) The emphasis in this course is on developing practical expertise in using computers for music research […] Humdrum provides tools for extracting, transforming, linking, classifying, contextualizing, comparing, and analysing music-related information (Mus-05)
(13) The course […] utilizes basic computerized tools for historical methodology (Hist-05)

Background information provides a context for the topic of courses and/or their disciplines – it may highlight the relevance of courses to the readership's interests, outline the scientific domains to be explored in class and/or introduce key concepts relevant to courses; e.g.:

(14) Fifty percent of the global population and 80% of Americans live in cities. Urban geography is the social science that investigates the integration of built forms, human interactions and the environmental aspects of places (Geo-02)
(15) Our environment is dynamic and ever-changing, constantly modified by natural processes and human activities (Geo-07)
(16) Our modern mathematics comes to us along a path that twists and winds through many centuries, cultures, and parts of the world, ever growing and expanding along the way to become one of the great and massive intellectual achievements of the human race. […] A genuine appreciation of the historical development of mathematics involves some degree of understanding of the mathematics that was done as well as the study of the surrounding history and culture and the lives of those who discovered it. (Hist-02)

Table 2 shows how often the above-identified content categories are exemplified in the corpus. In line with the genre nomenclature, most ACDs (98%) refer to the content of courses. Many also provide information relevant to the running of courses, namely about the logistics and methods involved (72% and 63%,

[9] If what is technically an indication of method is encoded as part of the object of study, I counted it as an instance of the information unit *content* (e.g. "HIV/AIDS awareness and prevention via artistic expressions will also be examined" Bio-02). On the other hand, if a text segment conveys information interpretable both as logistical and methodological, I classified it as an instance of the information unit *method.*

respectively). Envisaged goals or outcomes, which raise expectations in the readership, are mentioned in less than half of the corpus (44%), while background information, which sets the context for courses, is found only in 22% of the ACDs.

Discipline	*No. of ACDs referring to*				
	Content	*Goal/outcome*	*Logistics*	*Method*	*Background*
Biology	10	6	8	5	2
Geography	10	3	7	6	3
History	10	5	5	10	2
Journalism	9	3	7	6	3
Law	9	5	6	8	5
Literature	10	4	7	9	5
Music	10	8	9	5	1
Physics	10	3	6	4	1
Psychology	10	3	10	5	9
Statistics	10	4	7	5	0
Total	**98**	**44**	**72**	**63**	**22**

Table 2: Frequency and distribution of content categories across ACDs

4. Representation of entities

Four types of entities are recurrently mentioned in an ACD: the course, its content, its participants and the participants' planned activities.[10]

Courses are referred in two ways: either as curriculum components or as events to be run. In the former case, they are represented as DOERS (i.e. AGENTS responsible for educational contents and goals) and are syntactically encoded as thematic subject NPs of action verbs in the active voice (65%); e.g.:

(17) This course examines the sociocultural context of human behavior […] (Psych-05)
(18) Introduces psychology bases of instructional systems (Psych-07)
(19) The course will survey major mathematical developments […] (Hist-02)
(20) Evening laboratory periods will emphasize observation […] (Phys-07)

In the latter case, they are represented as entities affected by others' actions (i.e. as PATIENTS) – showing up as object NPs of active action verbs or subject NPs of passive action verbs (20%) – or alternatively as CARRIERS OF PROPERTIES (i.e. EXPERIENCERS) – realized as subject NPs of relational verbs (31%); e.g.:

[10] Additional entities, processes and topics are occasionally mentioned in the ACDs, especially in the *background* and *method* sections (e.g. humanity at large, people's attitude toward a discipline, documents examined). Reference to these is not taken into consideration in this study.

(21) This is the first course in a two-semester sequence on data analysis (Psych-04)
(22) In addition to covering libel law and the Federal Communications Commission (FCC), course encompasses issues related to radio employment contracts, trademarks […] (Law-08)
(23) If demand arises a crash course may be run during vacation […] (Stat-08)
(24) This course has been designed especially for first-year college students […] (Lit-04)

Alternative encoding options account for 18% of the data (see Table 3); e.g.:

(25) The major themes described in this course are *Waves and Optics, Electricity and Magnetism, Quantum and Atomic Physics.* (Phys-02; original italics)
(26) There is a build-your-own-motor contest as part of this course (see Calendar). (Phys-08)
(27) The main objective of the course is to understand that Latin American countries have some common characteristics (Hist-09)

Discipline	*Course as*			
	Doer: ACDs (tokens)	*Experiencer: ACDs (tokens)*	*Patient: ACDs (tokens)*	*Other: ACDs (tokens)*
Biology	8 (13)	6 (10)	5 (9)	1 (2)
Geography	7 (15)	2 (4)	2 (3)	0 (0)
History	9 (17)	4 (4)	1 (1)	2 (2)
Journalism	7 (9)	1 (1)	1 (1)	1 (1)
Law	6 (9)	2 (2)	1 (1)	3 (3)
Literature	8 (14)	1 (1)	1 (3)	1 (1)
Music	8 (15)	3 (6)	2 (2)	3 (3)
Physics	3 (7)	6 (10)	2 (3)	5 (7)
Psychology	5 (10)	1 (1)	2 (2)	0 (0)
Statistics	4 (5)	4 (7)	3 (6)	2 (2)
Total	**65 (114)**	**31 (47)**	**20 (31)**	**18 (21)**

Table 3: Semantic representation of courses[11]

Similarly, the **course content** can be represented as a patient affected by others' action – syntactically rendered as the object NP of an action verb in the active voice, the subject NP of an action verb in the passive voice, or the object of a preposition – or, alternatively, as an EXPERIENCER carrying some property – syntactically encoded as the subject or object NP of a relational verb; e.g.:

(28) We will survey a wide range of musical examples, Western and non-Western, ancient and contemporary […] (Mus-09)
(29) This course provides an introduction to landscape-forming processes and landforms […] (Geo-08)
(30) TOPICS INCLUDE FREQUENCY DISTRIBUTIONS, GRAPHS, STEM-AND-LEAF DISPLAYS, BOXPLOTS, SCATTER DIAGRAMS […] (Stat-09; original capitalization)

[11] In this and the following tables, the first figure in each cell identifies the number of texts in which a given feature is found, while the number in parentheses identifies the total occurrences of that feature.

(31) This course is a systematic study of the elements of the physical environment (e.g. weather, climate, landforms, water, soil and vegetation […] (Geo-10)
(32) A comparative analysis of these circumstances will be discussed in detail. (Hist-09)
(33) In addition to the basic concepts of Electromagnetism, a vast variety of interesting topics are covered in this course: Lightning, Pacemakers, Electric Shock Treatment, Electrocardiograms, Metal Detectors, Musical Instruments, Magnetic Levitation, Bullet Trains, Electric Motors, Radios, TV […] (Phys-08)

Table 4 shows that the two most frequent semantic encoding options for the course content are PATIENT, relevant to 87% of the ACDs (with 247 tokens), and EXPERIENCER, relevant to 44% of the ACDs (with 55 tokens).

Discipline	*Course content as*			
	Doer: ACDs	*Experiencer: ACDs*	*Patient: ACDs*	*Other: ACDs*
Biology	0 (0)	6 (6)	9 (20)	0 (0)
Geography	0 (0)	1 (1)	9 (21)	0 (0)
History	1 (1)	4 (5)	10 (36)	2 (2)
Journalism	0 (0)	3 (3)	8 (22)	1 (1)
Law	1 (1)	4 (4)	9 (25)	0 (0)
Literature	1 (2)	2 (3)	9 (38)	0 (0)
Music	0 (0)	6 (9)	10 (30)	1 (1)
Physics	0 (0)	6 (9)	6 (18)	0 (0)
Psychology	0 (0)	5 (7)	9 (25)	0 (0)
Statistics	0 (0)	7 (8)	8 (12)	0 (0)
Total	**3 (4)**	**44 (55)**	**87 (247)**	**4 (4)**

Table 4: Representation of the course content

Students can be cast in various semantic roles: as DOERS, RECIPIENTS, EXPERIENCERS and PATIENTS (see Table 5). When represented as DOERS (25%), students appear, however, to be following others' directions, rather than acting of their own free will. In such cases, they are encoded as subject NPs of active action verbs or as object NPs of the preposition *by*; e.g.:

(34) Students examine a variety of musical genres […] (Mus-03)
(35) […] you will use a personal computer […] (Mus-08)
(36) Students should be ready to engage in discussions of mode and invention, of genre and structure, and intent and execution (Lit-06)
(37) Students explore history of early childhood and elementary education […] (Hist-03)
(38) […] laboratory experiments allowing you to actively explore these principles, (Phys-10)

As RECIPIENTS (21%), students are presented as BENEFICIARIES of courses or target addressees. Syntactically, they are thus encoded as object NPs of the preposition *for*, as object NPs of active action verbs or as subject NPs of passive action verbs or of active relational verbs; e.g.:

(39) This sequence is designed for graduate students [sic] Psychology. Graduate students from other disciplines are welcome […] (Psych-04)
(40) The module introduces students to Social Psychology (Psych-09)
(41) The course provides students with an opportunity to examine aspects of, and synergies between, regional, national and international history […] (Hist-06)

Discipline	*Students as*				
	Doers: ACDs	*Recipients: ACDs*	*Experiencers: ACDs*	*Patients: ACDs*	*Other: ACDs*
Biology	5 (6)	3 (3)	6 (7)	0 (0)	0 (0)
Geography	1 (2)	1 (1)	0 (0)	0 (0)	0 (0)
History	4 (8)	2 (3)	3 (3)	1 (2)	1 (2)
Journalism	0 (0)	4 (4)	4 (6)	1 (2)	0 (0)
Law	3 (5)	3 (6)	4 (5)	1 (1)	0 (0)
Literature	1 (1)	2 (5)	1 (1)	0 (0)	0 (0)
Music	4 (8)	2 (4)	4 (5)	2 (4)	1 (1)
Physics	4 (5)	1 (1)	2 (2)	1 (1)	0 (0)
Psychology	1 (1)	2 (4)	1 (1)	0 (0)	0 (0)
Statistics	2 (4)	1 (1)	4 (6)	1 (1)	0 (0)
Total	**25 (40)**	**21 (32)**	**29 (36)**	**7 (11)**	**2 (3)**

Table 5: Semantic representation of students

As EXPERIENCERS (29%), students are either cast in the role of participants involved in the achievement of learning goals – in which case they are encoded as subject NPs of verbs expressing conscious, involuntary processes – or alternatively represented as the target of directions and expectations – in which case they show up as subject NPs of verbs expressing relations or involuntary processes; e.g.:

(42) Students […] [u]nderstand the […] methods of statistics inference […] (Stat-02)
(43) […] you need to be comfortable with math at the level of high-school […] (Stat-10)
(44) Students are responsible for all lecture and multimedia material presented in class […] (Hist-08)

Finally, students are represented as PATIENTS when the logistics of courses is being discussed (7%). In such cases, they are syntactically encoded as object NPs of active action verbs or as subject NPs of passive action verbs; e.g.:

(45) I will assign team members based on an analysis of skill […] factors (Journ-02)
(46) Students […] will be graded by a combination of final examination, class participation, class debate and a field research project (Law-09)

Discipline	*Teachers as*			
	Doers: ACDs	*Experiencers: ACDs*	*Patients: ACDs*	*Other: ACDs*
Biology	0 (0)	0 (0)	1 (1)	0 (0)
Geography	0 (0)	0 (0)	0 (0)	0 (0)
History	0 (0)	1 (1)	0 (0)	0 (0)
Journalism	1 (1)	0 (0)	0 (0)	0 (0)
Law	0 (0)	0 (0)	1 (3)	1 (1)
Literature	0 (0)	0 (0)	0 (0)	0 (0)
Music	0 (0)	1 (2)	0 (0)	0 (0)
Physics	0 (0)	2 (3)	0 (0)	1 (1)
Psychology	1 (2)	0 (0)	0 (0)	1 (1)
Statistics	0 (0)	0 (0)	0 (0)	0 (0)
Total	**2 (3)**	**4 (6)**	**3 (5)**	**3 (3)**

Table 6: Semantic representation of teachers

Teachers are hardly ever mentioned on their own (see Table 6), and rarely encoded in the first person; e.g.:

(47) I present the general linear model with particular emphasis on exploratory data analysis, contrast analysis, residual analysis, and structural models (Psych-04)
(48) Also, some guest lecturers can be invited to enrich the content of the course on specific topics (Law-03)
(49) Of course, I expect students enrolled in this course to uphold the UMBC Code of Student Conduct for Academic Integrity. (Hist-08; original underlining of the second text segment)
(50) interaction with instructors in discussion sections to provide one-on-one help with concepts and problem solving […] (Phys-10)

Students and teachers may be mentioned together. This applies to a minority of the ACDs (see Table 7). Students and teachers are usually represented as volitional participants engaged in joint deliberate acts – thus showing up as subject NPs of active action verbs – but also as conscious participants involved in common experiences – in association with active verbs of cognition. Finally, they are occasionally encoded as possessive adjectives before nouns expressing the activity they are engaged in or the experience they are involved in; e.g.:

(51) We will begin with a survey of global electoral geography before turning to the geo-history of voting in the United States (Geo-03)
(52) Mathematics is the language of Physics, and in this course we shall learn some of that language. (Phys-04)
(53) Taken together, Physical and Environmental Geography help us to understand how natural and human processes drive constant changes in the environments we live in (Geo-07)
(54) Our proximate goal will be to examine some of the more important theoretical frameworks that have been proposed as avenues toward understanding cultural evolution (Mus-04)

Discipline	*Teachers and students as*		
	Doers: ACDs (tokens)	*Experiencers: ACDs (tokens)*	*Other: ACDs (tokens)*
Biology	0 (0)	0 (0)	0 (0)
Geography	1 (1)	1 (1)	0 (0)
History	0 (0)	0 (0)	0 (0)
Journalism	0 (0)	0 (0)	0 (0)
Law	3 (6)	0 (0)	1 (1)
Literature	4 (11)	0 (0)	0 (0)
Music	2 (2)	0 (0)	1 (2)
Physics	0 (0)	1 (1)	0 (0)
Psychology	0 (0)	0 (0)	0 (0)
Statistics	0 (0)	0 (0)	0 (0)
Total	**10 (20)**	**2 (2)**	**2 (3)**

Table 7: Semantic representation of teachers and students together

The **study activity** carried out by the students and/or teacher may also be mentioned in the ACD. When encoded as a deverbal noun, it is most often represented either as an EXPERIENCER (25%) – showing up as the complement of a relational verb, the object of a preposition, the object of a relational verb or as part of a text segment with no predicate – or alternatively as a PATIENT (25%) – showing up as the subject NP of a passive action verb or as the direct object NP of an active action verb (see Table 8); e.g.:

(55) We will begin with a survey of global electoral geography (Geo-03)
(56) This course provides an introduction to landscape-forming processes […] (Geo-08)
(57) This course is a systematic study of the elements of the physical environment (Geo-10)
(58) Finally there will be more reading than in a typical mathematics course. (Hist-02)
(59) A general overview of major theories […] and principles in psychology (Psych-08)
(60) Close textual analysis will be complemented by consideration of the broader historical and cultural context (Lit-05)

Discipline	*Study activity*			
	Doer ACDs	*Experiencer: ACDs*	*Patient: ACDs*	*Other: ACDs*
Biology	0 (0)	3 (3)	6 (7)	1 (1)
Geography	0 (0)	3 (4)	4 (6)	0 (0)
History	0 (0)	4 (7)	2 (3)	2 (2)
Journalism	0 (0)	1 (2)	1 (1)	1 (1)
Law	0 (0)	1 (1)	3 (3)	2 (3)
Literature	0 (0)	2 (4)	2 (2)	3 (6)
Music	0 (0)	1 (2)	1 (1)	0 (0)
Physics	1 (2)	3 (8)	3 (5)	1 (1)
Psychology	0 (0)	5 (8)	0 (0)	0 (0)
Statistics	0 (0)	2 (5)	3 (3)	0 (0)
Total	**1 (2)**	**25 (44)**	**25 (31)**	**10 (14)**

Table 8: Semantic representation of the study activity

5. Tone

The projected events mentioned in the ACDs, including the conditions students are supposed to abide by, are represented with the certainty of scheduled arrangements. This sense of predetermination is mostly conveyed by the widespread use of the simple present tense and *will*-future, and the scarcity of epistemic modal expressions, which are found in 22 ACDs; e.g.:

(61) This course is not a substitute for BIOL 110, 220 or 221 or a prerequisite for other science courses (Bio-07)
(62) FACULTY AND/OR STUDENT WILL SELECT A CURRENT EDITION (Journ-10)
(63) The course profiles significant theories and people who have shaped early childhood education, focusing on recent centuries. Students consider what and who determines the goals of education, who defines and articulates the problems of education, and how that determines what solutions are created (Hist-03)
(64) […] students can easily consult the didactical materials […], which makes [sic] possible both for students and staff members to take it from distance […] The best essays […] can be published in the web site […] the best students could benefit of a field trip […] a [sic] MP of the Province of Bozen/Bolzano can be invited […] (Law-03)

Such recurrent choices lend authoritativeness and credibility to the voices behind the texts. On the one hand, the certainty with which information is conveyed makes it clear that no room for negotiation is envisaged: students are supposed to accept the information they receive as valid and binding. On the other hand, that certainty inspires confidence, because it suggests that courses are well-organized and not left to chance, and that the instructors will deliver what they promise.

Deontic modality is not very common, either, being found in 16 ACDs; e.g.:

(65) students may not receive credit for both biology 4622 and Biology 6622 (Bio-06)
(66) Students interested in enrolling in a 298 seminar must first obtain a PTE (Petition to Enroll) number (Bio-10)
(67) There will be an attempt to keep the mathematical pre-requisites to a minimum, but there should be a good grasp of mathematics at least through the level of calculus. […] Mathematics did not arise in a vacuum, and students should obtain some overview of the world history into which mathematical development was embedded (Hist-02)

Requirements, tasks, assignments and expectations are therefore more often expressed impersonally or encoded as predictive statements; e.g.:

(68) Includes laboratory […]. Prerequisite: high school algebra and trigonometry (Phys-07)
(69) However, the precalculus topics will be presented at a more sophisticated level, commensurate with a senior level math course. (Hist-02)

The impression conveyed is that what is required of the addressees comprises inevitable, predictable events, which are established by some external source, and thus are not under the ACD authors' control and are not being deliberately imposed. This form of encoding imposition appears to safeguard the negative face of the readers – who are not represented in a one-down position – and the positive face of the ACD authors – who are not represented as authoritarian.

6. Discussion

The present analysis has addressed two descriptive issues: what information units make up the ACDs considered, and how the entities and processes referred to in them are represented.

The frequency with which the various types of content categories recur in the ACDs positively correlates with the twofold informational-regulatory function of the texts. An outline of the content of courses is relevant to most of the texts (98%), and is in line with the name of the genre (i.e. course descriptions), which identifies the topic and main goal of the ACDs. At the same time, the complex nature of the events being described – run by academic authorities, involving students' active participation, and affected by administrative and logistical constrains – requires that reference be made to organizational matters, which helps students "find their academic bearings" (72%). Also, because courses play an educational role for their addressees, reference is frequently made to the approaches adopted by teachers (63%); this puts students in the picture about how to get the most out of courses. However, envisaged goals or outcomes, which determine the teachers' design of courses and inform students about what they are likely to learn, are harder to predict: they are not completely under the course designers' control and may be subject to updates; as a result, they are less frequently mentioned (44%). Finally, information on the disciplinary backgrounds of courses – relevant more to the actual practice of instruction per se than to the description of courses – is found in a minority of the texts (22%).

The entities mentioned in the ACDs are syntactically-semantically represented in accordance with the roles they play for the interactants, and their representation indicates how the ACD writers think of them. As something decided on by the teachers, the content of courses is mostly represented either as the OBJECT OF SOMEBODY'S ACTION (87%) or as the CARRIER OF SOME PROPERTY that describes it (44%). The study activity – the course of action that students and teachers participate and engage in – is represented mainly as a PATIENT (25%) or an EXPERIENCER (25%). Students are cast in a variety of roles, which reflects the variety of events they take part in: as EXPERIENCERS of learning goals (29%), RECIPIENTS of offers (21%), AGENTS performing assigned tasks (25%) and PATIENTS directed by others (7%). Teachers – although responsible for the content and structure of courses – are rarely mentioned on their own, and not always accurately represented: as PATIENTS (3%) or DOERS (2%). Teachers and students are occasionally mentioned together, and mostly represented as DOERS (10%).

Of the entities represented in the ACDs, courses are both the most frequent and the most prominent. The former characteristic is in no way surprising – courses are the topic of ACDs. The latter is due to the fact that they are often represented as self-determined and volitional entities, and are attributed agent-like properties, as if they were directly and intentionally in charge of educational contents, goals and policies. By contrast, students and teachers are marginalized: they are infrequently mentioned in the ACDs – even if they are responsible for co-constructing courses with their complementary forms of participation – and are rarely represented as direct interlocutors (i.e. with *I* and *you*), even if they are the communication participants that produce and receive the ACDs.

Such patterns may partly be due to the official nature of the texts – which present academic institutions' educational offers to the public at large[12] – and partly to the nature of the corpus – whose long ACDs are more likely to include deontic modality and forms of direct address in the sections specifying policies and requirements. However, these communicative choices also constitute negative politeness strategies for downtoning the directive force of the texts. The information conveyed through the ACDs is represented as "facts", that is, as definite arrangements single-handedly decided on by teachers, not susceptible to change and unconditionally valid. This is to be expected, given the role-relationship holding between teachers and students – the former being in an official position to establish policies for the latter to respect – and the role of courses – which are meant to educate and assess learners. But it counts as a form of imposition all the same. This may represent a threat to the addresses' negative face – whose freedom of action is restricted by course policies – and to the addressers' positive face – who may be disliked for coming across as authoritarian. Imposition, however, is mitigated in the texts through two communicative choices. On the one hand, the representation of courses as external, personified participants responsible for the assignment of roles and tasks turns them into mediators, or buffers, between teachers and students. On the other, the representation of teachers and students as remote third parties obscures the fact that these are direct interactants, the former imposing requirements on the latter.

Unlike the comparable, institutional genre of academic degree program descriptions (Fairclough 1993), the ACDs examined do not appear to have a promotional orientation, that is, to be produced for the purpose of attracting, or competing for, audience. They contain no direct appeal to the readership, no offer of incentives, no bait advertising what is in for prospective students – except for envisaged outcomes – and more generally no explicit attempt to please readers through positive politeness strategies. The focus is on communicative effectiveness, i.e. on providing accurate directions to self-motivated addressees, who know they are expected to meet requirements. The confident and determined precision with which information is conveyed, however, is softened by the negative politeness strategy of indirectness, which prevents addressees from feeling imposed upon.

[12] On the other hand, on the first day of class, teachers may use a lot of *I*'s*, you*'sand *we*'s when presenting their courses, so as to create rapport with the students.

7. Conclusion

The ACDs are oriented toward two communicative-interactional goals: clarity and effectiveness. Being subject to rigid space constraints, they focus on what is relevant and useful for the running of courses: information on the content, structure and logistics of courses is meant to prevent misunderstandings, while information on course policies prepares students to deal with courses by raising expectations to be met.

The former type of content appears to be conveyed with confidence. This impression is due to the widespread use of the simple present tense and *will*-future, and the limited use of epistemic modality, which lend credibility to the text authors. These sound like authoritative, responsible people who have thought out their courses of action and "mean business".[13]

At the same time, the encoding of the latter type of content reveals the authors' concern for safeguarding the interlocutors' face. The teacher-student relationship is carefully handled by means of two mitigating strategies: on the one hand, the participants are represented impersonally as third parties – rather than as addressers and addressees – and on the other, explicit deontic modality is scarcely used. This way, the risk of confrontation implicit in the laying out of rules is avoided: requirements are mentioned, but do not appear to be imposed on the addressee (Fairclough 1993: 157). This suggests that students' cooperative participation is expected – or taken for granted – through the performance of representative acts, rather than required through the performance of directive ones.

However, more direct involvement with the readership, as well as a clearer indication of what teachers want students to do or not do may be found in the sections on policies/requirements included in long ACDs. This suggests that future studies of the ACD genre could analyse short ACDs as distinct from long ones.

Finally, additionally insights into the genre can be gained by comparing ACDs across disciplines. As Table 2 suggests, different scientific domains may favour different information units in ACDs (e.g. *goal/outcome* in Music; *logistics* in Biology, Music and Psychology; *method* in History, Law and Literature; and *background* in Psychology), which are likely to reveal distinctive communicative practices in different academic communities.

Sara Gesuato
Department of Anglo-Germanic and Slavic Languages and Literatures
University of Padua
Italy
sara.gesuato@unipd.it

[13] What is mentioned in ACDs is, of course, likely to be subject to contextual, institutional constraints such as: university directives on the formal, typographic/textual structure of ACDs (e.g. division into pre-labelled sections), their phraseology (e.g. the use of specific terms or formulas), and their degree of thoroughness and specificity (depending on whether ACDs have the value of binding contracts or not).

References

Bernardini, S., A. Ferraresi & F. Gaspari (2009). Institutional English on the Websites of Italian vs UK/Irish Universities: A Preliminary Corpus-Based Analysis of Degree Programme Descriptions. Paper presented at the Conference *Genres on the Move: Hybridization and Discourse Change in Specialized Communication*, Naples, University of Naples Federico II, 9-11 December 2009.

Bhatia, V. K. (1993). *Analysing Genre*. London/New York: Longman.

Bondi, M. (2004). The Discourse Function of Contrastive Connectors in Academic Abstracts. In: K. Aijmer & A. Stenström (eds.), *Discourse Patterns in Spoken and Written Corpora*. Amsterdam/Philadelphia: Benjamins, 139-156.

Bruce, I. (2008). *Academic Writing and Genre: A Systematic Analysis*. London: Continuum.

D'Angelo, L. (2009). Academic Poster Presentation: A Marginalized Genre? Paper presented at the Conference *Genres on the Move: Hybridization and Discourse Change in Specialized Communication*, Naples, University of Naples Federico II, 9-11 December 2009.

Diani, G. (2004). A Genre-Based Approach to Analysing Academic Review Articles. In: M. Bondi, L. Gavioli & M. Silver (eds.), *Academic Discourse, Genre and Small Corpora*. Roma: Officina Edizioni, 105-126.

Fairclough, N. (1993). Critical Discourse Analysis and the Marketization of Public Discourse: The Universities. *Discourse & Society* 4(2), 133-168.

Flottum, K., T. Dahl & T. Kinn (2006). *Academic Voices: across Languages and Disciplines*. Amsterdam: Benjamins.

Gea Valor, M. L. & M. Inigo Ros (2009). On the Dynamic Nature of Genre: A Diachronic Study of Blurbs. In: K. Hyland & G. Diani (eds.), *Academic Evaluation: Review Genres in University Settings*. Basingstoke: Palgrave MacMillan, 199-216.

Gesuato, S. (2008). Calls for Abstracts: A Genre Shaping Professional Practices. In: P. Sánchez Hernández, P. Pérez-Paredes, P. Aguado Jiménez & R. Criado Sánchez (eds.), *Researching and Teaching Specialized Languages: New Contexts, New Challenges*. Universidad de Murcia: Edit.um, 60-71.

Gesuato, S. fc. Course Descriptions: Communicative Practices of an Institutional Genre. In: G. V. Polese et al. (eds.), *Genre(s) on the Move. Hybridization and Discourse Change in Specialized Communication*. Napoli: Edizioni Scientifiche Italiane.

Giannoni, D. S. (2006). Book Acknowledgements across Disciplines and Texts. In: K. Hyland & M. Bondi (eds.), *Academic Discourse across Disciplines*. Bern/Berlin: Peter Lang, 151-175.

Gotti, M. (2009). *Commonality and Individuality in Academic Discourse*. Bern/Berlin: Peter Lang.

Grazzi, E. (2009). A Corpus-Based Analysis of Modality in Four English Textbooks for the First Certificate Exam. In: D. Torretta, M. Dossena & A. Sportelli (eds.), *Forms of Migration. Migration of Forms. Proceedings of the 23rd AIA Conference*. Bari: Progedit, 260-274.

Hyland, K. (2009). *Academic Discourse: English in a Global Context*. London: Continuum.

Hyland, K. & M. Bondi (2006). *Academic Discourse across Disciplines*. Bern/Berlin: Peter Lang.

Hyland, K. & G. Diani (eds. 2009). *Academic Evaluation: Review Genres in University Settings*. Basingstoke: Palgrave MacMillan.

Myers, G. (2000). Powerpoints: Technology, Lectures, and Changing Genres. In: A. Trosborg (ed.), *Analysing Professional Genres*. Amsterdam/Philadelphia: Benjamins, 177-191.

Räisänen, C. (1999). *The Conference Forum as a System of Genres. A Sociocultural Study of Academic Conference Practices in Automotive Crash-Safety Engineering*. Göteborg: Acta Universitatis Gothoburgensis.

Römer, U. (2005). *Progressive, Patterns, Pedagogy. A Corpus-Driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam/Philadelphia: Benjamins.

Soler, V. (2007). Writing Titles in Science: An Exploratory Study. *English for Specific Purposes* 26, 90-102.
Swales, J. (1988). Language and Scientific Communication. The Case of the Reprint Request. *Scientometrics* 13, 93-101.
Swales, J. (1990). *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
Swales, J. & C. B. Feak (1994/2004). *Academic Writing for Graduate Students. Essential Skills and Tasks*. Ann Arbour: University of Michigan Press.
Swales, J. & C. B. Feak (2000). *English in Today's Research World: A Writing Guide*. Ann Arbour: University of Michigan Press.
Tognini-Bonelli, E. & E. Del Lungo Camiciotti (2005). *Strategies in Academic Discourse*. Amsterdam: Benjamins.
Trix, F. & C. Psenka (2003). Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty. *Discourse & Society* 14(2), 191-220.

Identity Markers in the Academic Discourse in English on Western–Islamic Finance: Two Case Studies

Mariarosaria Provenzano

Abstract

This paper introduces a socio-cognitive enquiry into the analysis of a corpus of academic textbooks on Islamic Finance in English addressing Western university students. The focus is on the ongoing process of *hybridization* between conflicting Western and Islamic generic norms (cf. Swales 1990; Hyland 2000) taking shape in the academic discourse of Islamic Finance and marking the identity of this discipline which, in the Western tradition, does not encompass any of the religious precepts that, instead, permeate the Islam-influenced norms of such specialized discourse. A comparative critical analysis (cf. Halliday 1985; Fairclough 1995, van Dijk 2008) will be therefore carried out on the structure of this emerging academic genre with the objective to identify (a) specific cross-cultural 'identity differences' in terms of specialized discourse schemata (Gotti 2003) and (b) degrees of 'hybridization' in terms of novel semantic and textual features aimed at making unfamiliar concepts of the Islamic Finance *accessible* and *acceptable* (de Beaugrande & Dressler 1981) to Western students in the perspective of broader intercultural transactions. To this purpose, the Appraisal Framework (Martin and White 2005) will be applied to two case studies focused on some Islamic specialized concepts (i.e., *murabahah* and *riba*) involved in the hybridisation process in order to recognize the writers' 'illocutionary intent'.

1. Introduction: setting the scene

The present study develops from the awareness that new concepts in the specialized domain of Islamic Finance are emerging and that they are expected to have an impact upon the disciplinary 'schemata' of Finance that have been so far globally recognized on the basis of Western/US norms and parameters. This emerging process seems then to acquire a new importance in the implementation of ELF courses for university students of Finance and Economics precisely due to the need to face this intercultural challenge and make the new concepts 'accessible' (cf. Widdowson 1996) to them. Hence the present study proposes a socio-cognitive analysis of a corpus of university textbooks produced by Islamic scholars addressing Western students. The focus is on the ongoing process of *hybridization* shaping these textbooks, insofar as *hybridization* is here meant to define the process of meaning negotiation between the conflicting principles and norms of the Islamic Finance academic genre, on the one hand, and those of the Western/US Finance genre (cf. Swales 1990; Hyland 2000; Gotti 2005), which may ultimately determine (a) the construction of new 'hybrid' identities and (b)

affect the recognition of the once-for-all established norms of the Economics genre. The study deals with the analysis of two relevant specialized concepts from the domains of Islamic and Western Finance: the 'Riba', conventionally rendered as 'interest', and 'murabahah', perceived and rendered in Western 'contract' terms as 'sale and profit'. The following sections are therefore aimed at exploring the nature and scope of the 'hybridization' process by focusing on the conceptual and textual strategies applied to make Islamic concepts 'accessible' to Western students as well as on the 'intentionality' level as this is encoded in the specific textual choices.

2. Research Rationale, Hypothesis, Objective

As mentioned in the 'Introduction', 'hybridization' is meant to define an ongoing process aiming at simplifying, whenever this is required by the readers' own divergent expectations. Such awareness leads to an overall process of text 'reformulation' (Provenzano 2008), aimed at adapting texts to the receivers' own socio-cultural constraints. In the case at hand, the reformulation process achieved through 'hybridization' aims at making specialized Islamic concepts 'accessible' (cf. Widdowson 1996) to Western students by providing a consistent explanation and clarification of such concepts by resorting to Western discourse culture in its terms and structures and, thus, favouring the construction of 'new' hybrid identity markers. The hypothesis of the study is that precisely this process of 'hybridization' is necessary in the construction of these emerging textbooks, insofar as it helps construct 'accessible' meanings. The objective is, therefore, to explore the cross-cultural "identity markers" signalling degrees of the conceptual and textual hybridization of this novel Western - Islamic academic genre, also by focusing on the extent to which this process of hybridization implies considering the construction of a text on the basis of 'new' intercultural parameters. With this purpose in mind, the next section will be aimed at focusing on the theoretical background, i.e. the relevant 'standards of textuality' (de Beaugrande & Dressler 1981), which are explored in a cross- cultural perspective.

3. Theoretical Background

The present section is aimed at identifying the 'standards of textuality' (ibid.), whose analysis is considered relevant in constructing the 'new' hybrid texts on the academic genre of the Islamic Finance, in such a way as to make them 'accessible' to Western students. In particular, two of them are considered relevant to the analysis: 'informativity' (ibid.: 139), and 'intertextuality' (ibid.: 182). 'Informativity' is here meant to indicate the reader's assumed unfamiliar knowledge conveyed to Western students, and 'intertextuality' is realized in the texts through the 'hybrid' texts' consistent references to correlated Islamic contractual forms, also rendered by means of Western culture terms and

discourse structures, whose ultimate effect is to render the specialized Islamic texts as clear and accessible as this is required for Western students. More specifically, focusing on 'informativity' will entail (a) the analysis of the structural/ textual strategies by which these texts are made 'informative' to Western students, and (b) the focus on the way(s) 'informativity' is realized within this intercultural perspective, thus finally considering and affecting the consolidated Western norms on the specialized 'genre' of the academic textbooks on Finance. Hence, it is here worth mentioning just some of the strategies activated to make the specialized Islamic texts 'accessible' to Western students, thus helping identifying hybrid identity markers in the construction of the new academic genre of Finance. As an example, the 'cross- cultural paraphrasis' represent some of these strategies, since they would enable providing 'new' information to the students, thereby acquiring a facilitating, highly didactic value in the overall process of text 'reformulation', as it will be shown in details in the next section.

Moreover, another relevant theoretical reference is represented by the Hallidayan models (1985; 1978), whose application is aimed at analyzing the 'identity markers' in the hybrid corpus from a functional, socio- semiotic perspective, i.e. by focusing on their value in the textual construction of dialogic relationships between the Islamic scholars on the one hand, and the prospective Western receivers on the other hand. In such a perspective, the 'register analysis', with particular concern for the 'Tenor' parameter (Halliday 1978), will also be applied with the aim of focusing on the linguistic strategies applied to create favourable conditions of approach to the topics discussed on the part of the Western students.

4. Method

The method accounts for a comparative critical analysis of the hybrid corpus with two reference corpora, i.e. (a) original Islamic academic texts on Finance, (b) Western/US texts on the same topic(s). The analysis will develop on the basis of the CDA (Fairclough 1995), aimed at focusing on the conceptual and textual strategies activated in the hybrid corpus to influence the reader's perception. The objective of the next lines is, therefore, to provide a description, and an interpretation of the salient extracts from the hybrid corpus, showing the extent to which the choices made aim at influencing the reader's perception.

It is necessary to explain here also the practical aim of the textbooks, i.e. they aim not only to provide students with the required theoretical equipment, but also explain and guide them on how to make profits with the Islamic world. Hence, the two objectives of 'hybridization' or 'reformulation' may be summed up as consisting of two levels: (a) the conceptual one, based on the simplification and explanation of unavailable Islamic concepts through Western terms and discourse structures (see, for example, the texts on the 'Riba' case where interpretation is grounded on the actualization of forms/concepts through (a) new informativity

patterns, and (b) 'new coherence patterns'); and (b) the textual and pragmatic one. Let us explore, hence, the salient texts from the hybrid corpus, by pointing out the relevant analytical structures.

4.1. 'Analysis'

'New informativity and coherence patterns'

As anticipated in the previous sections, the necessity to make Islamic specialized concepts clear to Western students has laid the basis for the creation of new textual patterns, aimed at providing the required information precisely in the way(s) this is needed to fit the cognitive and socio-cultural schemata of the text receivers. In our hybrid corpus, 'new' examples reflect this specific need through the actualization of hybrid, informative patterns, that are shown below. In this perspective, the two case studies that are taken into account consider two specialized concepts from the specialized domains of Islamic and Western Finance. The first of them is the Islamic 'Riba', since this concept has often been taken as controversial to the Western parallel, but not similarly connoted concept of 'profit'. The 'Riba', among other things, is negatively connoted, as it appears in the extracts from the original Islamic corpus that are quoted below and that contrast with the positive connotation associated to the parallel Western concept of 'interest'.

(1) Whoever *pays more* or *takes more* has *indulged* in *riba*. (from the *Surah al-Baqarah Fourth Revelation*, verses 275-281)
(2) Some of the *more popular* mortgage derivative products have been *interest-only* and *principal-only strips*. (from the US/Western corpus Bodie et al. 2008)

Hence, the need to 'reformulate' it in simpler terms, i.e. in a way that it could comply with the cognitive schemata and specialized discourse requirements of the Western receivers. Therefore, the novelty appears in the hybrid corpus through the use of hybrid terms and structures aiming to finally simplify the 'original' concept, as it is evident in the following quotation(s) from the corpus. The extracts have been selected precisely on the basis of their relevance to the identification of these 'hybrid identity markers'.

(3) *The structure of Islamic finance* revolves around *the prohibition of any return* derived on *a loan/debt* (*Riba*) and *the legality of profit*. (from Ajub 2007: 73)

Analysis is based on the *hybridization* of terms and reconciliation of opposite concepts belonging to the two domains of Islamic and Western Finance, i.e. "loan/debt", "return", "profit" associated to the domain of Western Finance, and 'Riba' as the Islamic parallel concept. The concern for simplification comes to be actualized also by means of extended modal forms (cf. Guido 2004: 280) in the occurrence "the prohibition" aimed at (a) suggesting the implication of 'perentory obligation' associated to the action of 'prohibiting return' and (b) guiding the students towards the identification of the expected social behaviour

within this specific domain. Equally relevant is the use of the nominalization form (cf. Gotti 2005: 77) "the legality", which introduces a Western discourse structure in the textbook, thus identifying new informativity patterns and reconciling two divergent economic views on Finance. The line of reasoning followed by the Islamic scholars is also made clearer through the use of specific deontic modals, such as "must" conveying the sense of religious obligation as applied to the economic domain. Below is the quotation from Ajub (2007):

(4) This implies that all financial transactions *must* be representative of *real transactions* or the *sale* of *goods*, *services* and *benefits*.

Hybridization is here realized by providing students with the necessary knowledge, based on both the modal, and also on the use of specialized Western terms 'real transactions', 'sale', 'goods', 'services', 'benefits'.

4.2. Cross-cultural paraphrasis and 'intertextuality'

Another relevant structure in the hybridization process and in the informativity patterns is the use of cross-cultural paraphrasis, that appear as a device for achieving conceptual clarity and provide the Western receivers with the linguistic equipment that explains concepts they are not familiar with. Exemplifications are provided by the 'Riba' corpus, which is rendered in Western terms through cross-cultural definitions, such as the one indicated below:

(5) *Riba – commonly known as interest –* is *an increase* taken *as a premium* from the debtor. (from Ajub 2007)

Other examples of the achievement of the explanatory function are provided below:

(6) 'Riba al- nasi'ah': The term *nasi'ah* means *to postpone or to wait* and it refers to the time period that is allowed for the borrower to repay the *loan* in return for the addition of the *premium*. *Hence* it refers to the *interest* on *loans*. (from Zamir 2007)

The focus is on the specialized Western terminology 'loans', 'interest', 'loan', 'premium', but also on the synonymic rendering of legal action through (a) doublets and (b) function words. Also consider the use of cohesive devices such as 'hence', here aimed at providing a complete definition of the specialized concept. Within this clarifying perspective, we shall also consider 'intertextuality' that allows to make the text comprehensible by providing explanations of the specialized Islamic financial concepts at hand (i.e. 'Riba', but also 'Murabahah'), through consistent intertextual references to parallel Islamic contractual forms, that are also actualized through Western discourse structures, as it is made visible in the following quotation from Usmani (2002):

(7) But *one should not ignore* the fact that the most important requirement for validity of *murabahah* [related concept] is that *it is a genuine sale* with all its ingredients and *necessary* consequences.

4.3. 'Hedging'

Besides the conceptual level, hybridization also occurs at the textual level: at this stage, we consider the 'hybrid markers' used to construct social relationships between the Islamic scholars, on the one hand, and the audience of Western students, on the other hand. Among these markers, 'hedges' are salient textual devices aimed at downgrading emphasis to make controversial issues, i.e. 'interest rate' or 'halal ('holy') profit', 'acceptable'.

(8) *No doubt, the use of the rate of interest* for determining a *halal profit cannot* be considered *desirable*. (from Usmani 2002)

'Hedging' (cf. Hyland 2000) is realized through the adverb "no doubt", but also tentative textual devices are aimed at achieving a similar effect. Below are indicated the most salient examples, pointing out the discursive implications correlated to their specific use.

(9) *It* ['the use of the rate of interest'] *certainly* makes the transaction *resemble* an *interest-based financing*, at least *in appearance*, […] even this *apparent resemblance should be avoided* as far as possible.
(10) [But *one should not ignore*] (from Usmani 2002)

The use of modal structures and other 'hedging' devices ('should'; 'certainly'; 'in appearance', 'apparent') is particularly relevant in the construction of the 'new' identity of the Islamic academic genre of Finance, since it helps identify the scholar's own viewpoint, specifically guiding the audience towards a clear line of behavior. This is realized by the modal 'should', both covering the senses of 'directiveness' (cf. Guido 2004: 282) but also 'social obligation', thus acting as 'facilitator' of the 'illocutionary meaning'. Identity is also constructed through the impersonal form 'one', that conveys a formal tone to the construction of the discourse.

A final remark concerns a synthesis of the discourse strategies aimed at explaining specialized concepts through the target-culture discourse genre, whose main traits are: (a) function words opening up the interpretative space; (b) 'metaphors', 'hedging' attributes and deontic modality, this latter used to express permission rather than ability; (c) 'intertextual markers' providing references to the interlocutor's familiar specialized concepts.

Below are practical applications:

(11) a. Most derivatives were originally created as 'hedging' devices, *or ways of controlling or reducing risks* generated by fluctuating interest rates or currencies. *By another definition*, derivatives are bets on interest rates, currencies and commodities that result in real cash obligations or rewards.
b. Also, financial innovation *can* be a *good* thing for Islamic finance, since by being '*sliced and diced'*, financial products *can* be tailored to the needs of *religious*-minded investors. *Objectionable* features *can* then be removed from a product.
c. The 'principal-only' component, *just like* zero-coupon bonds, *can* satisfy clients who do not want to deal directly with interest. (...)

5. Conclusions

The analysis of the selected corpora involving the two case studies concerning two specialized Islamic financial concepts of 'Riba' and 'murabahah' pointed out an ongoing process of textual and discursive hybridization, characterized by specific identity markers. This process is aimed at constructing a favourable approach to the 'new' academic genre of Islamic Finance as it addresses Western/International students with diverging theoretical and practical interpretations of the specialized concepts. The work opens up the necessity of ethno-methodological investigations to be submitted to Western students, aiming to probe its actual accessibility.

Mariarosaria Provenzano
Department of Foreign Languages and Literatures
University of Salento (Lecce)
Italy
mrosariaprovenzano@ateneo.unile.it

Corpus references

Ayub, M. (2007). *Understanding Islamic Finance.* Wiley Finance: Chichester.
Bodie, Z., A. Kane, & Marcus, A. (2008). *Investments*. McGrawHill: Boston.
Usmani, M. T. (2002). The Use of the Interest-Rate as a Benchmark, In: idem, *An Introduction to Islamic Finance* (*Series – Arab and Islamic Laws Series*), Leiden: Kluwer Law International, 48-49.
Zamir, Z. (2007). Prohibition of Interest (Riba) in Islam – The Social, Moral and Economic Rationale. http://www.nzibo.com/riba.

References

de Beaugrande, R. & W. Dressler (1981). *Introduction to Text Linguistics*. London: Longman.
Fairclough, N. (1995). *Critical Discourse Analysis*. Edinburgh: Pearson Education.
Gotti, M. (2005). *Investigating Specialized Discourse*. Bern: Peter Lang.
Guido, M. G. (2004). *Mediating Cultures. A Cognitive Approach to English Discourse for the Social Sciences*. Milan: LED.
Halliday, M.A.K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
Halliday, M.A.K. (1985/1994). *An Introduction to Functional Grammar.* London: Edward Arnold.
Hyland, K. (2000). *Disciplinary Discourses. Social Interaction in Academic Writing.* Edinburgh: Pearson Education.
Provenzano, M. (2008). Towards a Hybrid Academic Discourse of Western-Islamic Finance. A Comparative Enquiry. *Lingue e Linguaggi* 2, Dipartimento di Lingue e Letterature Straniere, Università del Salento. Lecce: Pensa Editore, 93-110.
Swales, J. (1990). *Genre Analysis*: *English in Academic and Research Settings.* Cambridge: Cambridge University Press.
Widdowson, H.G. (1996). *Linguistics.* Oxford: Oxford University Press.

Disciplinary Cultures in Academic Posters

Larissa D'Angelo

Abstract

Within the English for Specific Purposes (ESP) discipline, much attention has been dedicated to the "open genre network" of academic writing. Conference poster presentations instead have received little attention from researchers. Furthermore, there are no systematically collected corpora of conference poster presentations belonging to different academic disciplines, allowing for a comprehensible textual and semiotic analysis of this marginalized genre. A corpus comprising 60 posters belonging to 6 different academic disciplines will be devised and the material collected will be analysed to understand how different academic cultures influence move structures, underlying patterns, text-audience relations, and communicative purposes of academic posters.

1. Introduction

In the past three decades, within the English for Specific Purposes (ESP) discipline, much attention has been dedicated to the "open genre network" of academic writing (Swales & Feak 2000: 8), comprising the most open, public and visible research genres. In particular, research on academic writing has examined a variety of genres such as journal articles, abstracts, dissertations, and conference proposals; other genres instead, such as conference poster presentations have received little, if no attention at all from researchers. This is probably due to the fact that, as Swales (2004) recognized, not all research genres have equal value in the eyes of disciplinary participants and these values seem to shift as we move from one part of the research world to another. Also, as Hyland (2000) pointed out, different disciplines value different kinds of arguments and set different writing tasks: if Power Point conference presentations, for example, have become fashionable and increasingly common in almost every discipline, it is in the hard sciences that conference posters are mostly used and valued.

Although poster sessions are sometimes marginalized and are generally considered less prestigious than paper sessions, they play an important part in scientific conferences and constitute a valid and interesting alternative to paper presentations. In fact, by facilitating informal discussions between presenters and their audience, poster sessions provide a more intimate forum for exchange than do regular paper presentations. Also, compared with genres with more rigid structures, these academic works differ in the lack of prescriptive guidelines, the allowance for creativity and individuality, the space for narratives and stories, and the goal both to inform and to persuade.

Nevertheless, the academic poster has evolved from traditional genres, such as the research article, conference visuals and handouts (Swales 2004: 16). For example, like research articles, to achieve both coverage and clarity, posters need to follow a precise format and content organization: in addition to a title/author label and abstract, most successful posters, in fact, provide brief statements of introduction, method, subjects, procedure, results, and conclusions (Alley 2003). At the same time, as for conference Power Point presentations and handouts, visual presentation and graphics become vital to outline a piece of work in a form that is easily assimilated and stimulates interest and discussion (Matthews 1990; Tufte 1990).

A further useful concept, when considering this genre is that of the *genre set* (Devitt 1991; Swales 2004), which sees the use of conference presentations limited to senior scholars and conference posters mostly used by or assigned to junior scholars, working their way up a commonly perceived and accepted hierarchy of genres. Probably for this reason, although poster presentations can be ideally considered as a genre in academic communication activity, akin to conferencepapers, journal articles, and grant proposals (Bazerman 1988; Berkenkotter & Huckin 1993; Hyland 1998), they are treated as "a poor country cousin" to oral presentations (Swales & Feak 2000: 81), with an uncertain reputation:

> The poster as a genre has been struggling to find a niche for itself as a viable alternative to the traditional conference presentation because in most fields it retains a second-class status. Although the APA now offers substantial prizes for the best student posters at its conferences, it is significant, I believe, that these are restricted to JR graduate student work (Swales 2004: 64).

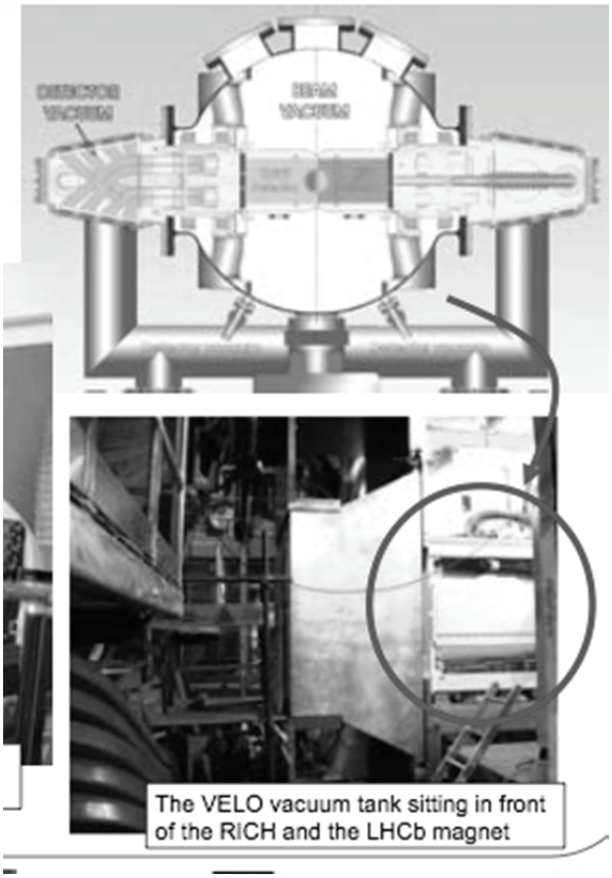
Thus, conference posters have so far been marginalized within less prestigious "constellations of genres" (Swales 2004: 7), such as book reviews, scientific research/laboratory reports and editorials and they are often considered as second class compared to oral presentations (Swales & Feak 2000). Despite this, there is quite an impressive amount of material available online on academic posters, ranging from how-to tips and techniques (Block 1996; Woolsey 1989) to discussions of how to use posters as a tool for professional development in the workplace (Miracle 2003). Posters have also been studied as a situated practice (Brown & Duguid 1991; Gherardi 2000, 2001; Wenger 2000) and social action (Miller 1984; Paré & Smart 1994) and some articles describe the use of posters as a teaching and evaluation device within university courses (Bracher, Cantrell & Wilkie 1998; Denzine 1999; Hay & Thomas 1999; Howenstine et al. 1988; van Naerssen 1984).

On the other hand, very few linguistic/semiotic analyses have been carried out on academic posters. The few available are pioneering explorations of the genre, taking into consideration forms, norms and values (MacIntosh & Murray 2007), how poster exhibitions are systematically and prospectively accepted and their perceived value (Salzl et al. 2008) or briefly considering grammar, rhetoric, graphic design and visual perception in general (Matthews 1990).

Despite these tentative studies, vision-language (and vision-gesture) interaction has, indeed, been neglected in the multimodal resources community, though the needs for related multimodal corpora are becoming increasingly demanding. Currently there are no systematically collected and annotated corpora of conference poster presentations belonging to different academic disciplines, allowing for a comprehensible textual and semiotic analysis of this 'marginalized' genre.

Thus, the aim of this paper is to investigate how discourse is presented through posters and which textual and visual reader-oriented strategies are commonly employed in different academic disciplines. An academic poster can in fact be defined as a multimodal communicative genre, with text, graphics, colour, speech, and even gesture used to convey meaning (Kress & van Leeuwen 2001). Also the fact that the content is displayed as a "visual unit," all on a single view plane, distinguishes it from other academic genres.

Because of its unique characteristics, the poster genre necessarily needs to be analysed differently from other genres, applying more than one framework of analysis to fully understand how its different parts work together to successfully convey meaning.



Poster 1: Poster detail displaying an example of vectors

In this sense, following Kress & van Leeuwen's observation (2001: 44), it is important to understand that the semiotic code of language and the semiotic code of pictures each have their own particular means of realizing what in the end are perhaps quite similar semantic relations. For example, what in language is realised by words of the category 'action verbs' is in pictures realised by elements that can be formally defined as 'vectors'. For example, Poster 1 is a visual detail of a Physics poster, which could be perhaps verbally translated in the following words: "Look here for a schematic representation of the VELO vacuum tank".

The arrow and circle invite the reader to "look" more attentively at a certain picture and help explain what is being represented. Likewise, what in language is realised by locative prepositions is realised in pictures by formal characteristics that create the contrast between foreground and background, as in the following example.

Poster 2: Poster detail displaying a background/foreground contrast

In Poster 2, displaying a detail of a law poster, we see an incarcerated man holding his head, probably thinking of the situation he is in. The man is shown in the background, and in the foreground we read the words "Problem" and "Solution" written in yellow, followed by an explanation. Because of the background/foreground detail, we can translate in words what the author of the

poster probably wanted to tell us: "there you see a man clearly worrying over a problem; here you can read the solution to the problem."

However, one should note that sometimes, what is expressed through language cannot be expressed through images and vice-versa, making the relationship between images and text complementing but in no way dependent on one another (Kress & van Leeuwen 2001: 17). Having said this, to carry out an exhaustive analysis, one should take into consideration the written text together with the visual elements as they both work together to achieve unity, clarity and most of all, salience.

To better comprehend the genre knowledge and the skills required of researchers in the academic poster genre, a multimodal corpus of 60 conference posters belonging to the discipline of Law and Physics has been devised, constructed and subsequently analysed. A number of authors and academic staff working in the two disciplines have also been interviewed in order to better comprehend how the communicative strategies are learnt and employed and how this multimodal genre is perceived and valued in different disciplines.

Although the genre of academic posters has also a spoken component, because presenters verbally describe the poster summarize it, point out the most salient elements and interact with viewers, such verbal component was not be taken into consideration. The analyses have been carried out only on the visual and textual elements present in the text, in order to comprehend if and how much these elements are used to interact with the reader and are able to replace verbal interaction.

2. Methodology

The text present in the academic posters collected is analysed using Hyland's (2000) theoretical approach to metadiscourse interpretation. Using Thompson's (2001) terminology, the academic metadiscourse found in the poster is distinguished between interactive resources (transitions, frame markers, endophoric markers, evidentials, code glosses) and interactional resources (hedges, boosters, attitude markers, engagement markers, self mentions). The full list of interactive and interactional markers searched is drawn from Hyland (2000).

If written text can be analysed in terms of interactive and interactional forms, the same can be done for the visual components found in poster presentations. A new approach to reading visual images came with the publication of Gunther Kress and Theo van Leeuwen's work (1996). Taking as their starting point the idea that visual images can be read as 'text', the metaphor of 'grammar' can be applied to the study of visuals. In this sense 'grammar' is not a set of rules for the correct use of language but rather a set of socially constructed resources for the construction of meaning.

Kress & van Leeuwen's work is revolutionary in the sense that it provides a key to reading images as if they were a text. In particular they raise the status of

the visual component in multimodal texts by considering it as "an independently organized and structured message – connected with the verbal text, but in no way dependent on it: and similarly the other way around" (1998: 17). Consequently, they take the view that language and visual communication both realize the same fundamental and far-reaching systems of meaning that constitute our cultures, but that each does so by means of its own specific forms, and independently, although not everything that can be realized in language can also be realized by means of images, or vice versa (1998). This approach becomes then fundamental for those seeking a descriptive framework of analysis to be applied to multimodal texts.

interactive resources	achieved through	interactional resources	achieved through
information value	left- right top-bottom centre-margin triptych	salience	contextualization foreground/background contrast in colour (saturation, differentiation, modulation)
framing	frame lines discontinuities of colour discontinuity of shape empty space between elements		use of pictures (demand/offer) use of schematic analytical pictures (one- or two-, three- dimensional graphs) representation depth illumination brightness font size font type font colour disalignment
connective elements	vectors (size, colour, shape, attenuated/amplified (density/frequency)) repetition of shapes repetition of colour allignment	size of frame	medium shot: human figure from knees up medium close shot: human figure from waist up close up: head and shoulders extreme close up: anything less than head and shoulders, or an isolated body part is used for dramatic visual impact.
graphs	conversion process taxonomies (covert/overt) flowcharts networks tables	perspective	frontal vertical
fonts	size colour type		

Table 1: Framework of analysis for visual elements

In order to analyse the visual component of academic posters, a framework of analysis has been devised, drawing from Kress & van Leeuwen's (1996; 2001) semiotic work. In particular, a number of visual elements have been categorized as interactive or interactional depending on their communicative function.

As Table 1 shows, interactive elements, organizing information and guiding the viewer in the comprehension of the multimodal text, are constituted by the following interrelated systems: Information value, Framing, Connective Elements, Conversion Processes, Taxonomies, Flowcharts and Networks.

Information Value is created through the placement of elements, which endows them with specific informational values attached to the various 'zones' of the image: left and right, top and bottom, centre and margin (see Figure 1).

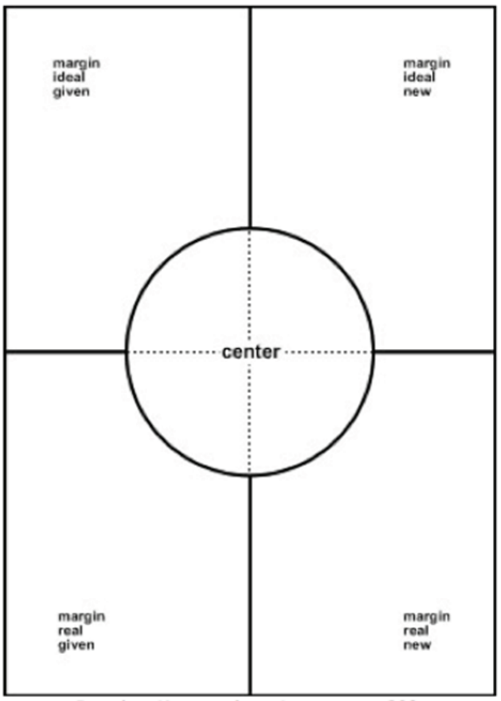


Figure 1: The dimensions of visual space (Kress & van Leeuwen 1996: 208)

The presence or absence of framing devices (realised by elements which create dividing lines, or by actual frame lines) instead disconnects or connects elements of the image, signifying that they belong or do not belong together in some sense. Other interactive devices could be connective elements such as vectors (used in different sizes, shapes, colour or used in an attenuated or amplified way, denoting density or frequency), the repetition of shapes and colours and the alignment of elements within the poster. Information could be visualized through conversion processes, taxonomies (covert or overt), flowcharts, networks and tables.

Interactional elements, used to attract, involve and engage viewers, evolve around the concept of Salience, through which elements are made to attract the viewer's attention to different degrees, as realized by such factors as Contextualization (a scale running from the absence of background to the most

fully articulated and detailed background), Representation (a scale running from maximum abstraction to maximum representation of pictorial detail), placement in the foreground or background, contrasts in tonal value or colour, use of pictures of real people and/or objects instead of schematic analytical pictures, depth in pictures and images, illumination, brightness and use of font (size, type and colour). Interactional devices are also the Size of Frame (ranging from 'extreme close-ups', which attract and interest the viewer more, to 'very long shots', which are much less salient), the Perspective used (frontal, engaging the viewer explicitly or vertical, conferring more or less power to the viewer). Another differentiation must be made between those images depicting people directly looking at the viewer (defined as 'demand pictures'; Kress & van Leeuwen 1996: 126), and those in which the subject of the picture looks away. In the first case, just like Engagement devices in text, the picture engages and demands the attention of the viewer. Just like a direct question forces the reader to answer and engage in a dialogue with the author, the image of a person looking directly into the camera (and therefore at the spectator), demands the attention of the viewer and establishes a relationship with him/her. If the person depicted in the picture does not look at the camera and is involved in some sort of activity, the viewer is simply a spectator ('offer picture'; 1996: 124). The person depicted is 'offered' to the viewer, who can see what the person does but no relationship is established between the two. Also important are the types of graphs used in posters, as one-dimensional graphs are more objective and less salient than two- or three- dimensional graphs (1996: 104).

The visual components of posters are analysed qualitatively. Visual elements are categorized either as Interactional or Interactive and single instances are counted in order to determine the different levels of modality (salience) of posters.

The visual components of posters have been necessarily analysed qualitatively. Visual elements have been categorized either as Interactional or Interactive and single instances have been counted in order to determine the different levels of modality (salience) of posters and the most frequent interactive/interactional devices used. The quantitative and qualitative analyses are based on both automatic and manual searches. For the computer-based counts, Wordsmith Tool 4.0 (Scott 2004) and Portable Document Format (PDF program) search options have been used. The results were followed by manual correction to rule out any non-relevant cases.

3. Physics subcorpus

3.1. Analysis of the texts

The texts of the posters making up the Particle Physics corpus are made up in total of 26709 words and the average word count per poster is 809. This first data reveals immediately that within the discipline of Particle Physics authors tend to

be wordy and prefer 'crowding' the poster with text. This tendency goes against the most recurring advice found in online poster guidelines and best-practice tips, specific to this discipline but found also in other academic areas, which says that the number of words in a poster should not exceed 500.

What in a research article can be described, explained and debated in several pages, in a poster must be condensed in few words and a very limited amount of space. Most conferences issue an "Instructions for Posters" or poster guidelines with spatial limitations of a 4-ft x 6-ft or 4-ft x 8-ft poster area. Although no mention is usually made to the maximum amount of words allowed per poster, because of spatial limitations, poster text and graphics need to be as concise and condensed as possible, relating only the most important facts and key points. Such conciseness and brevity is sometimes very difficult to achieve, especially if an entire article has already been written before the poster is presented. For of this reason, several guidelines available online and through Departments and Writing Centres advise to pay much attention to the written part of the poster, as it often ends up being the hardest to master. Authors are advised, for example, to consider sentence length, as short or medium-length sentences are visually more effective than long sentences.

Along with narrowing down the poster's substantive content, presenters are advised to use compressed language (Swales & Feak 2000). To achieve this, guidelines suggest using bullet points and telegraphed wording to avoid full sentences, which contribute to a text-dense appearance. Although much attention is generally given to the amount of text found in posters, the fact that posters are often not adequately abbreviated is underscored by an emphatic statement from Stoss (2010): "The poster is NOT the pasting of a scholarly article on poster board or foam-core and standing by to defend results reproduced in miniature on the 'poster'." The poster may be closer to "an illustrated abstract" (Hess, Tosney & Liegel 2010) written large and put on display.

The analysis conducted on the corpus clearly indicates the importance of metadiscourse in the genre of posters, as it has found 529 instances of metadiscoursive devices, with an average of 18 per poster.

As Table 3 shows, poster authors in the field Particle Physics tend to use far more interactive than interactional features, the former being more than twice as frequent as the latter. Interestingly, the same can be said for other genres and other disciplines, such as book reviews appearing in Applied Linguistics, Economics, Law and Medicine (D'Angelo 2008; 2010). This is probably due to the fact that interactional devices are much more face-threatening than their interactive counterparts and are therefore used less frequently.

interactive resources	occurrences	interactional resources	occurrences
transitions	240	hedges	26
frame markers	69	boosters	36
endophoric markers	10	attitude markers	23
evidentials	11	engagement markers	60
code glosses	37	self mentions	17
total	**367**	**total**	**162**

Table 3: Metadiscourse occurrences in the text

For what concerns interactive discourse, transition markers, frame markers and code glosses seem to be the most frequent devices, whereas in the case of interactional resources, engagement markers and boosters are the most frequently used.

As can be noted in Table 3, the transitional devices are the most frequently used in discourse in general and they serve an important function, as they act as bridges between parts of the text, which help the reader interpret ideas in the way the writer wants him/her to understand them. The frequent use of code glosses and frame markers indicates instead that in this discipline authors demonstrate their expertise by constructing arguments clearly, and highlight the unfolding text in a less personal or challenging way. The following are examples respectively of code glosses and frame markers found in the text of Physics posters:

(1) Most problems are fixed by replacing faulty components (e.g. electronics board, patch panel, cable).
(2) At LHCb, the VELO has three vital roles: 1. Trigger on a B decay of interest (the VELO is part of the online software trigger); 2. Suppress multiple interactions (A pile-up veto in the hardware trigger); 3. Track reconstruction: used to seed the tracking in the rest of LHCb.

The high frequency of engagement markers is an important indication of the involvement of the reader. Engagement markers, in fact, contribute not only by bringing the reader into the text and establishing solidarity among scholars but also by working towards the creation of a shared evaluative context. The way Physicists employ engagement markers, such as questions marks, can be seen in Example 3:

(3) How can we study the magnetic distortion effect ?

The most interesting result is that within this discipline, boosters are more numerous than hedges. This suggests the idea that, within this discipline, authors of posters have more liberty to make bolder statements, draw conclusions or argue for controversial positions. A frequent use of boosters is usually found in

the works of established scholars, rather than novice writers. Kirsch (1993: 52) confirms this tendency when she writes

> Faculty members recognized the importance of establishing their authority by adhering to the disciplinary conventions of their fields, such as making references to previous scholarship, situating their work in current discussions, highlighting their contributions to ongoing research, and following standard methodologies of their disciplines.

The fact that boosters are frequently used in the genre of academic posters, where authors are mostly postgraduate students or young scholars with limited academic experience, makes this result even more peculiar. Two examples of this bold, upfront writing style are the following, found in a poster by a postgraduate student:

(4) LHCb will precisely measure CP violation in the decays of hadrons containing b-quarks.
(5) One can clearly see that the signal from beam 1 MIB is on-time with the proton-proton signal, while beam 2 MIB is separated by about 10 ns.

3.2. Analysis of the visuals

As Table 4 shows, also for what concerns the visual component of posters, interactive resources are more recurrent than interactional ones, suggesting that authors are more concerned with making the poster comprehensible and logic, rather than aesthetically alluring.

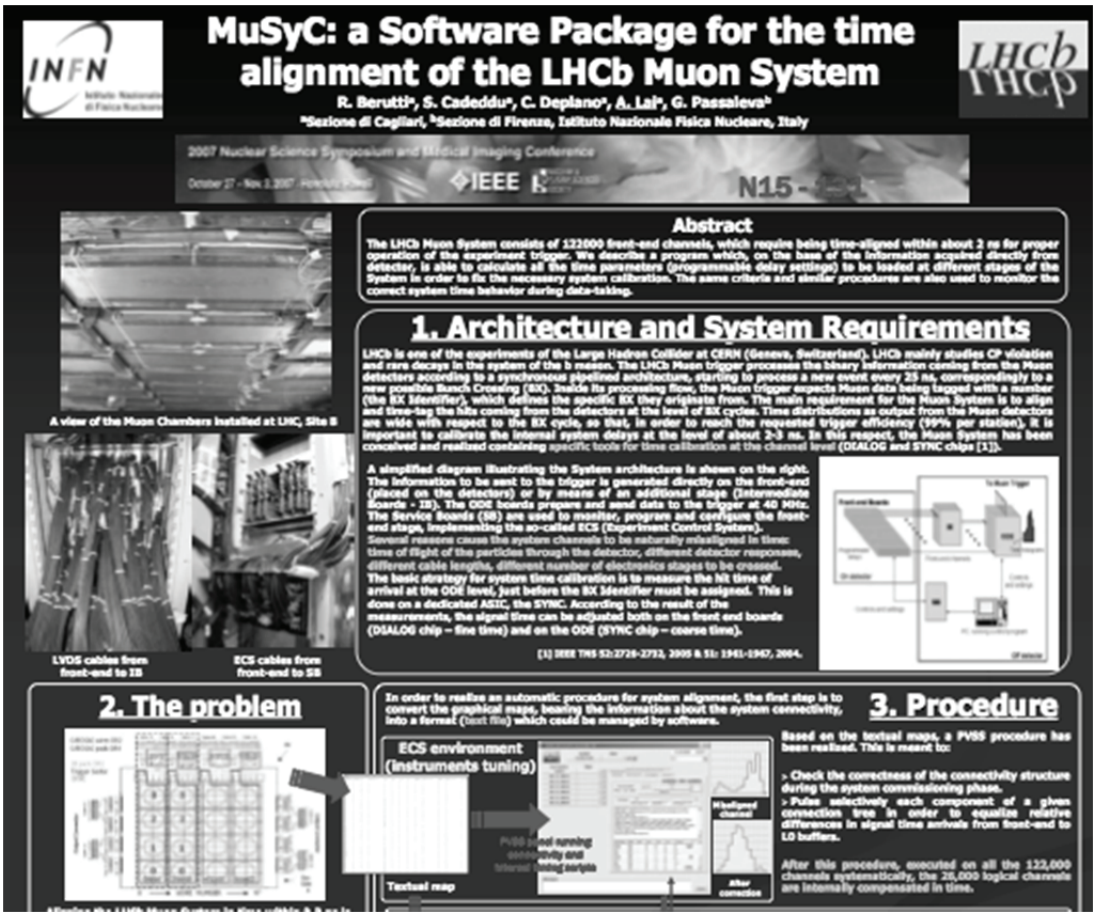
interactive resources	occurrences	interactional resources	occurrences
information value	12	salience	103
framing	41	size of frame	3
connective elements	40	perspective	13
graphs	57		
fonts	46		
total	**196**	**total**	**119**

Table 4: Metadiscourse occurrences in visuals

The most numerous interactive resources found are graphs in general, and in particular, charts and tables. This is probably peculiar to this discipline, where experimental data are predominant. What is interesting is that more than half the authors who choose to display a table or chart, they do so three-dimensionally, a detail that makes the table more appealing to the viewer and therefore augments posters' salience.

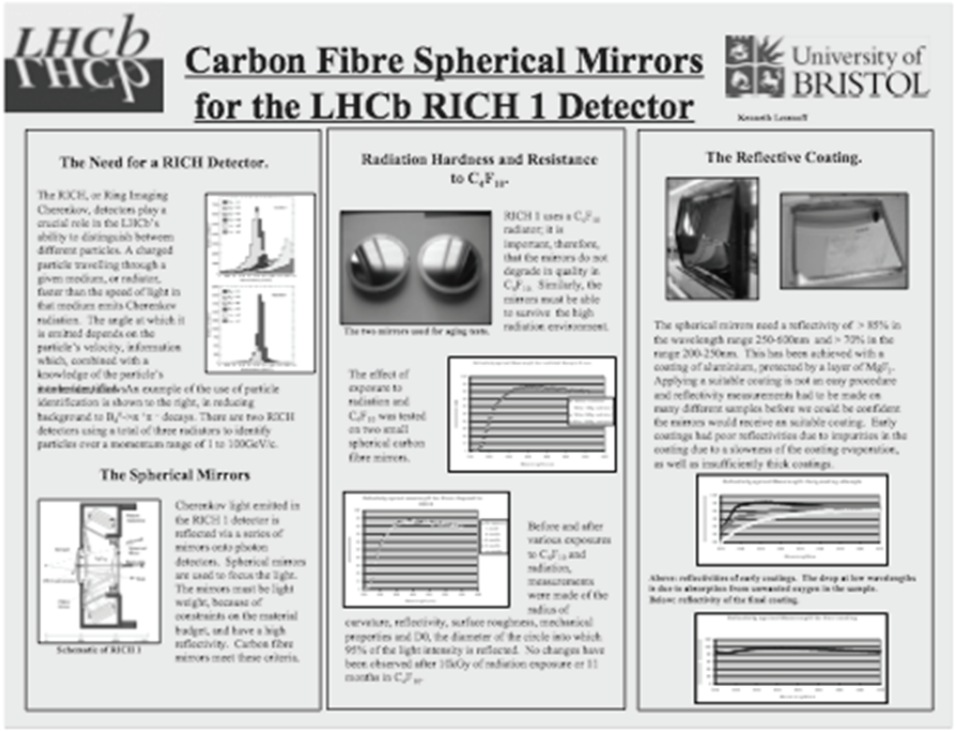
Other frequently used interactive devices are fonts, whose type, colour and/or size are used to guide the reader through the text, making it more comprehensible, perhaps highlighting which parts are connected or underlining the most important textual elements. An example of font type/size and colour

used interactively, is Poster 3, where we see a repeated font size and type to distinguish the titles from the rest of the text and the colour orange (here indicated by the slightly darker grey) to highlight parts of text, considered more important.



Poster 3: Font size, type and colour used as interactive device

Although frames are here widely used to separate or connect blocks of text, units of data or images, posters within this disciple rarely display a clear format, which would help the reader follow and anticipate the flow of information, as it happens with other genres such as the research article, where the traditional IMRAD format establishes that an Introduction is followed by a Methodological section, a section with Results and a Discussion. Only 12 out of 30 posters use the interactive resource called 'information value', a layout that organizes information coherently, and the majority of this batch, align the flow of information vertically. Poster 4 for example is divided in three main sections by three vertical columns. The reader is thus invited to follow the stream of information going from top to bottoms as suggested by the vertical columns and the titles inserted in the text.



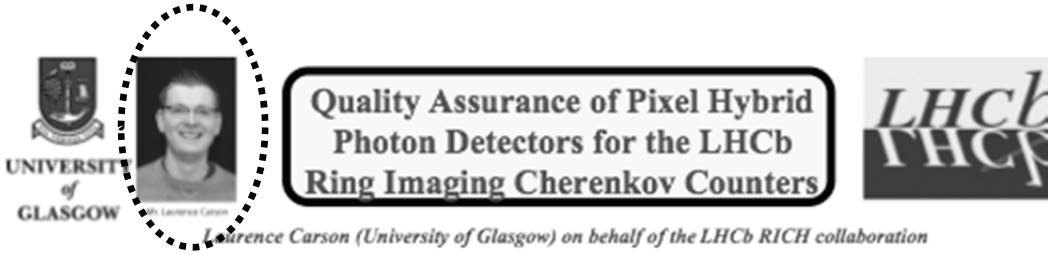
Poster 4: Poster with text and images aligned vertically

The use of graphs, especially if two- or three-dimensional, can also count as an interactional device, in that a viewer tends to be attracted more to a poster displaying charts, graphs and schematic analytical pictures, than by a poster displaying text only. However, an excessive number of graphs should be avoided as they 'crowd' the poster with too much information for the viewer to be understood in a short amount of time. A poster with too many graphs easily obtains the opposite effect and therefore, for the purpose of this study, only posters displaying between 1 and 4 charts have been selected as having this interactional device.

Along with charts, also pictures augment considerably a poster's salience. More than half the posters in this discipline display at least one picture, although only half of these use big enough pictures to be clearly seen from at least 6 feet, the typical distance set between a poster and its viewer. On the other hand, the majority of the pictures displayed a frontal or vertical perspective, two devices that help establish a relationship with the viewer. Just like engagement markers, a frontal perspective perhaps displaying a person looking directly at the camera seeks above all to bring about an imaginary relation between the represented person and the viewer. As Kress & van Leeuwen (1996: 122) have noted,

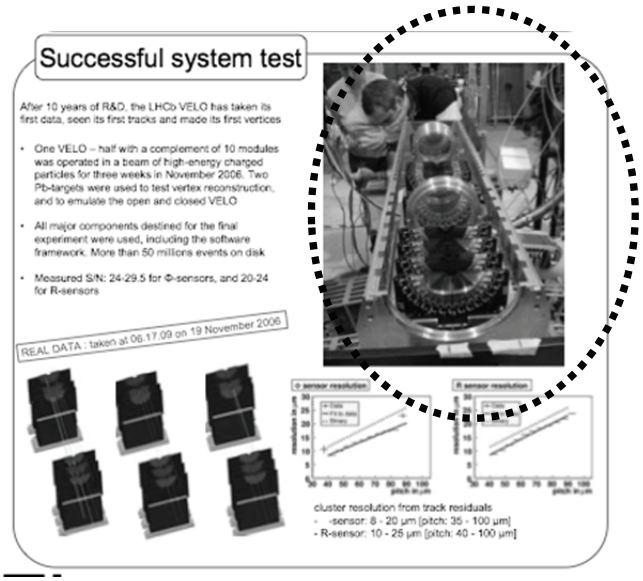
> [t]here is […] a fundamental difference between pictures from which represented participants look directly at the viewer's eyes, and pictures in which this is not the case. When represented participants look at the viewer, vectors, formed by participants' eyelines, connect the participants with the viewer. Contact is established, even if it is only on an imaginary level. In addition there may be a further vector, formed by a gesture in the same direction […].

These kinds of pictures, also defined as 'demand pictures', acknowledge the viewer explicitly, addressing them with a visual 'you' (1996: 122) and demand that he/she enters into some kind of imaginary relation with the person shown. In the case of Poster 5, where in a detail of a poster we see a frontal picture of the author of the poster, the author is trying to establish a relationship with his audience and establish a positive, intimate contact.



Poster 5: Poster detail displaying a demand picture

A vertical perspective is also highly salient, although the viewer is invited to observe from above the situation and is put into a powerful position (Poster 6) as an omniscient spectator.



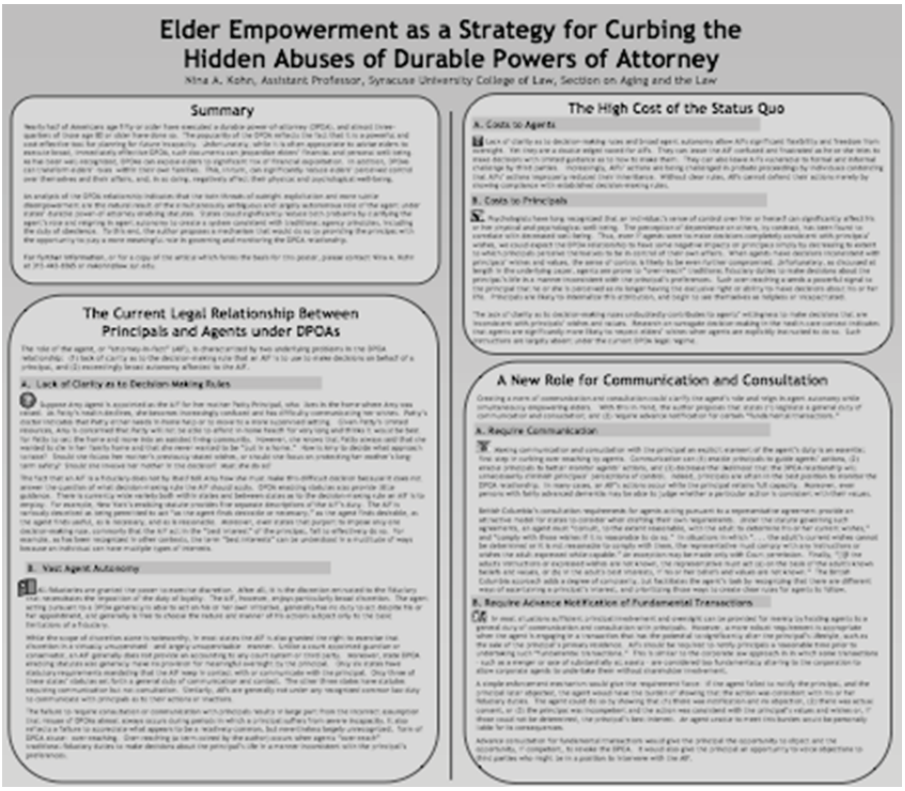
Poster 6: Poster detail displaying a picture with a vertical perspective

4. Law corpus

4.1. Analysis of the texts

The total number of words contained in the law corpus is 20550, an average of 685 words per poster. The amount of text contained in law posters is less than the one found in physics posters, a characteristic that is not found in other genres used by lawyers, such as the research article and the book review. In these genres lawyers have been found to produce longer texts and generally, longer, more articulated

sentences (D'Angelo 2010). It is thus surprising to find such brevity in academic posters, which suggests the idea that, within this discipline, this genre is used differently, giving more importance to visual elements rather than text. However, in this corpus 10% of posters display only text and no visual elements. In these cases, information appears dense, sometimes divided in different sections, very much recalling the layout of a research article (see Poster 7).

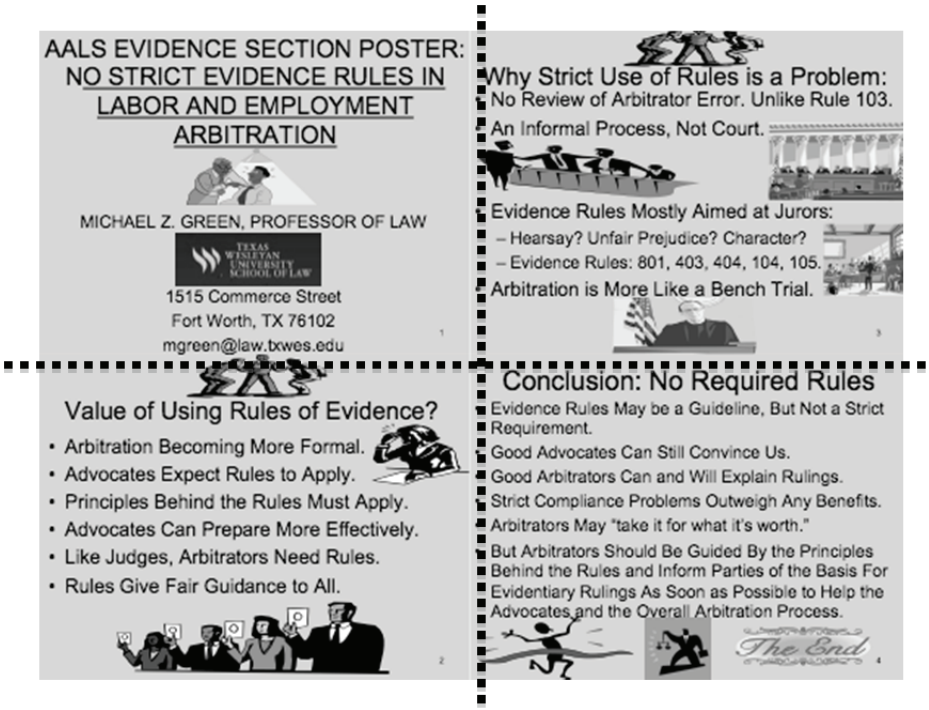


Poster 7: Law poster displaying text only

In an interview conducted it was mentioned that authors often display great amounts of text not because the discipline lacks experimental data or laboratories and instruments to be shown, but because authors construct their posters from one or more research articles they have already written and, in many cases, lready published. The great amount of text already written before the poster is created seems to influence the way the poster is presented, allowing authors to simply copy and paste sections of text from their papers to their posters. This practice goes against one of the most frequent advices found in best-practice examples and poster tips, which says that a balance between text and images should be obtained, in order to reach 40% of blank space on the view plane, which renders the poster more legible and less 'crowded'.

The choice of posting text only on a poster has been explained by an interviewed post doc fellow, as an example of how the poster genre is still in its initial stage in the discipline of Law. Researchers and practitioners working in this field have only recently begun to see poster sessions at conferences and still few of them venture in actually making and presenting a poster. The lack of

examples in the field and a limited knowledge of the genre as used in other disciplines probably explain the diversity found in this corpus. A number of authors rely heavily on the format and presentation of more traditional genres such as the research article; others collate PPT slides (Poster 8) and still others use very limited amount of words per poster (in one case just 50 words) and rely entirely on the visual impact of the pictures depicted.



Poster 8: Law poster displaying collated PPT slides

Focusing on metadiscourse, in other genres lawyers have been found to use interactive and interactional markers more frequently than applied linguists, doctors and economists (D'Angelo 2010, Sala & D'Angelo 2009). As Table 5 shows, this does not seem to happen in the genre of academic posters, as they have been found to contain far less modality markers, both interactive and interactional, than posters produced by physicists.

interactive res.	occurrences	interactional res.	occurrences
transitions	85	hedges	23
frame markers	46	boosters	1
endophoric markers	4	attitude markers	10
evidentials	6	engagement markers	22
code glosses	23	self mentions	12
total	**164**	**total**	**68**

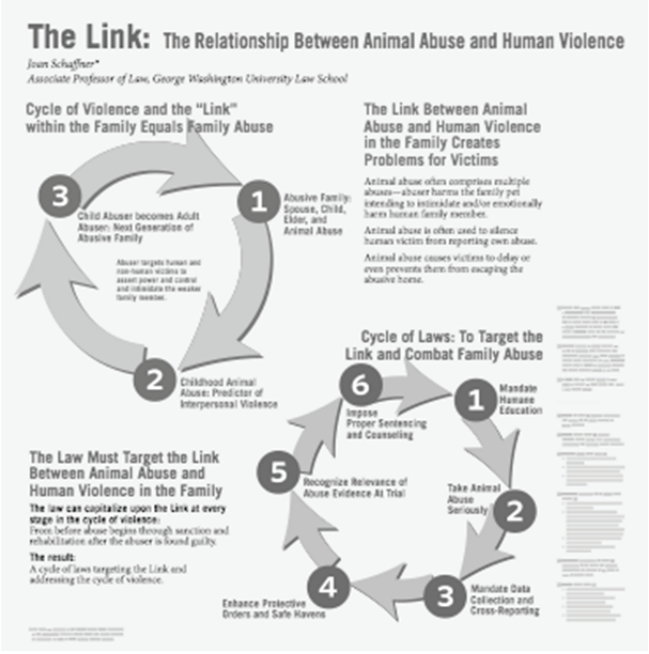
Table 5: Metadiscourse occurrences in the text

In particular, the frequency count found only one occurrence for what concerns boosters, and less than half the number of attitude markers found in the Physics corpus, a clear sign that lawyers avoid strong statements and the act of presenting arguments with absolute conviction. Also, three times less engagement markers are found in the Law corpus than in the Physics one, which suggests that in this genre, lawyers prefer not to explicitly refer to or build a relationship with readers. As will be further analysed, the act of attracting and engaging the reader will be done mainly trough visual elements.

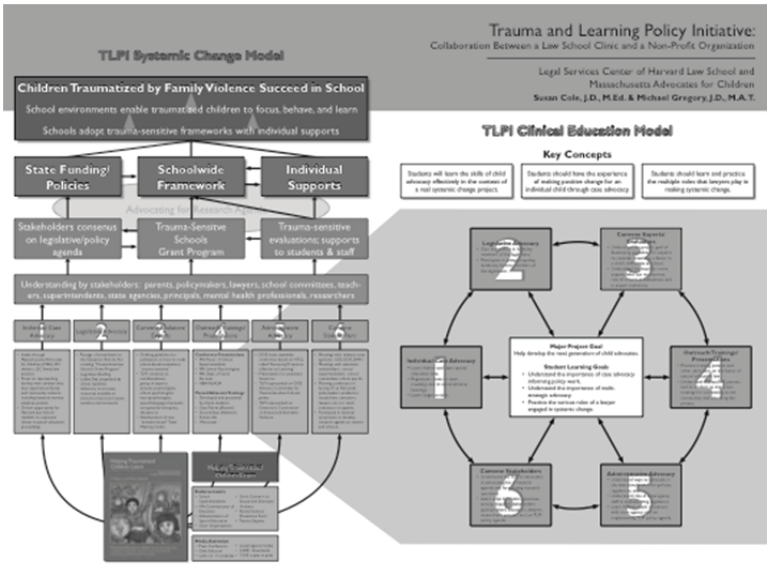
The most recurring resource has been found to be, as in the Physics corpus, transition markers, followed by frame markers. Frame markers are probably numerous because of the frequency with which bullet and numbered lists are used in the text. The condensed message of a list, in fact, appeals to the reader's need to gather information efficiently: the information is laid out, unadorned, for quick comprehension, as in Poster 9, where we see a poster simply made up by three lists.

Poster 9: The repeated use of lists in the text of a Law poster

Another peculiarity is that numerous posters display their text in the form of conversion charts (Poster 10), networks or taxonomies (Poster 11). The text contained in these charts necessarily has been condensed, chunked and fragmented. The result of this process is a text that, compared to the text produced in Physics, is easily and quickly readable but lacks in interactive resources.



Poster 10: Text presented in the form of a conversion chart



Poster 11: Text presented in the form of a taxonomy and a network

In these cases the text is organized in such a way that without the aid of visuals it would not be possible to fully comprehend the content.

4.2. Analysis of the visuals

The Law corpus gathers posters with different designs, layouts and most of all, posters that make a different use of visual elements such as pictures, schematic analytical pictures and graphs. In certain cases, the use of visuals is so prominent

that the text is very limited or has marginal importance. The entire concept evolving around the poster is presented through one or more pictures or schematic analytical figures, or even through a visual/verbal metaphor (Poster 10). In this example the brief text reads

> Coco Chanel visually taught us the wisdom of the little black dress and pearls. You already teach pearls of legal wisdom, now address the visuals. Are they effectively articulating your message? Are they engaging your audience? Are they readable and visually legible? Take a full-length look at your visual aids: do they aid long-term memory?

The metaphor between the pearls depicted in the visual and the 'pearls of wisdom' mentioned in the text is an example of how strongly images can be used to convey meaning and, in a certain sense, entertain the viewer. Such poster in fact, is not easily forgotten. Because the role played by visuals is so important in certain disciplines, Dubois (1985) suggests that the commercial display booths prominent in biomedical conferences may have inspired poster presentations.

Poster 12: Visual/verbal metaphor in a Law poster

At the same time, once the attention of the passerby has been obtained thanks to the visual and textual data displayed on the poster, it is vital to keep the viewer's interest live as long as possible, which means mastering the one-to-one interaction sufficiently well to deliver the message of the poster. Tham (1997) emphasizes that a poster must be understandable on its own, in the absence of the presenter:

> A poster is simply a static, visual medium (usually of the paper and board variety) that you use to communicate ideas and messages. The difference between poster and oral presentations is that you should let your poster do most of the "talking"; that is, the material presented should convey the essence of your message.

Poster guidelines found mainly in the hard sciences say that ideally, a well-constructed poster is self-explanatory and frees the presenter from answering obvious questions so that he/she can supplement and discuss particular points of interest. In the case of Law poster this obviously does not happen, because of the lack of text present. The poster is used here mainly to attract the attention of viewers and the communicative message is left to the oral component of poster presentations. It is the author himself who stands by his poster and completes the information displayed on the board, answers questions and debates the concepts presented.

As Table 6 shows, if we apply the framework of analysis to the visual components, we find that Law posters in general, make use of slightly less interactive and interactional features than Physics posters. However, much more attention is given to the information value resource, which assures that the poster is organized coherently and that the reader knows which information to process first. Most posters use a top/bottom format, as in the Physics corpus, left/right organization of the text, which makes the poster readable from left to right.

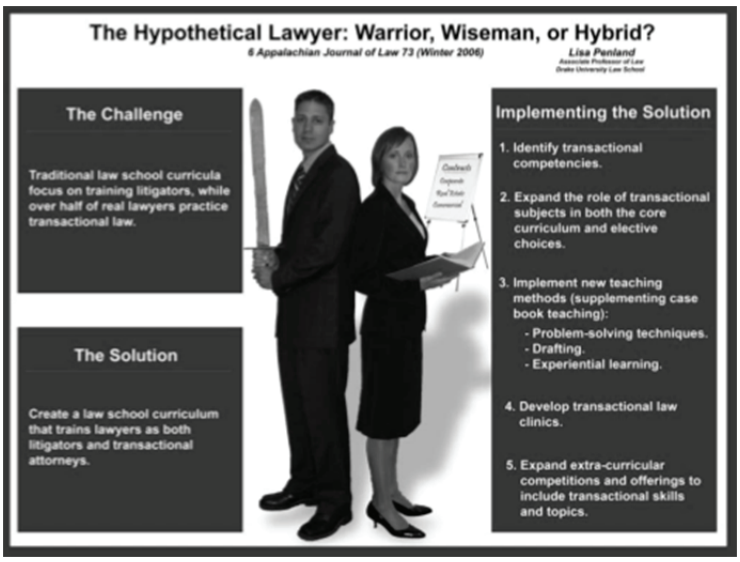
Also noticeable is the use of font size, colour and type to guide the reader in the comprehension of the poster. As it happens in the Physics corpus, authors rely on colour, size and graphic diversity to highlight separate sections, differentiate content and even reinforce the visual message as in Poster 10, where the font type used recalls the elegance of the pearls displayed.

Although the number of posters displaying pictures is the same for both corpora, lawyers tend to use the size of frame resource more frequently than physicists, a characteristic that is explained by the fact that in this discipline, as it happens in advertising, authors like displaying one or two enlarged images, occupying most of the space available (Poster 1). This procedure works very well in attracting viewers, a fundamental aspect of poster presentations, which, unfortunately, is often forgotten. Such prominent visuals remind of Wittich & Schuller's (1973: 125) description of a poster as "a visual combination of bold design, colour, and message intended to catch and hold the attention of the passerby long enough to implant a significant idea in the mind."

interactive res.	occurrences	interactional res.	occurrences
information value	21	salience	93
framing	38	size of frame	7
connective elements	25	perspective	11
graphs	15		
fonts	47		
total	**146**	**total**	**111**

Table 6: Metadiscourse occurrences in visuals

The use of pictures displaying people differs significantly between the two disciplines: Law posters, unlike Physics ones, make a wide use of pictures representing people (36% against 16%). Also important is the fact that, like in advertising, these posters do not use real people but actors. Pictures therefore are 'staged' and actors pose in front of the camera. The result are images that are visually more appealing and that serve more easily the message of the author, as in Poster 13, where we see two actors playing the role of a warrior and a wiseman. The connection between the picture and the title is immediate and the text found on the left and right side of the poster further explains the concept.



Poster 13: People displayed in Law posters

The present paper wants to be an exploratory research in the field, aiming at drawing attention to the interesting and hybrid genre of academic posters, a genre that makes use of different, sometimes fascinating communicative strategies and is therefore difficult to do well. Although a corpus of 60 posters is not enough to draw definite results, it is nonetheless interesting to note that the majority of the posters collected follow a number of conventions, best-practice notions and advices currently widely circulating on Internet through University websites, personal weblogs, conference sites, University Writing Centres, poster websites and Online Poster Journals. This widespread circulation of information and prescriptive guidelines indicate that poster sessions are becoming an increasingly important part of scientific conferences and constitute a valid and interesting alternative to paper presentations at conferences.

However, the lack of linguistic analyses currently available on academic posters or even corpora gathering examples of poster in different disciplines is a clear sign of the marginal importance this genre still retains in the Academia. Furthermore, although several guidelines exist indicating how a poster should be visually devised and presented, an analysis taking into consideration the semiotic code of language and the semiotic code of images has never been carried out. By applying the frameworks of analysis described above, academic posters can be

analysed in their entirety, taking into consideration the visual as well as the textual elements they display. As demonstrated, an interdisciplinary analysis reveals interesting patterns in the interactive and interactional discourse, underlining more or less effective communicative strategies employed by poster presenters, depending on their discipline, authority and experience.

Although the present study has been carried out on only two disciplines, Physics and Law, and on a limited number of posters, already significant differences and similarities have emerged. Even if Lawyers and Physicists belong to very different fields of studies and that the poster genre has only been recently used by lawyers, a number of formats as well as interactive and interactional elements present in texts and visuals are the same. This suggests the idea that within this genre a number of cross-disciplinary conventions exist and that the rules and formats used in one discipline often influence other disciplines.

Several differences between the two corpora have also been noted, such as the central rather than marginal role of the speaker in presenting the poster, the predominant use of pictures and images in Law posters, the recurrent use of bi- and tri-dimensional graphs and the abundance of boosters in Physics posters.

These differences and similarities are expected to become even more evident when the analysis is extended to other disciplines and the number of posters is augmented, shedding light on the different (or similar) academic conventions surrounding the poster genre.

Larissa D'Angelo
Department of Comparative Language, Literature and Culture
University of Bergamo
Italy
larissa.dangelo@unibg.it

References

Alley, M. (2003). *The Craft of Scientific Presentations.* New York: Springer.

Block, S. M. (1996). Do's and Don'ts of Poster Presentations. *Biophysical Journal* 71, 3527-3529.

Bracher, L., J. Cantrell & K. Wilkie (1998). The Process of Poster Presentation: A Valuable Experience. *Medical Teacher* 20, 552-557.

Brown, J. S & P. Duguid (1991). Organizational Learning and Communities-of-Practice: toward a Unified View of Working, Learning and Innovation. *Organization Science* 2, 40-57.

D'Angelo, L. (2008). Gender Identity and Authority in Academic Book Reviews: a Metadiscourse Analysis Across Disciplines. *Linguistica e Filologia* 27, 205-221.

D'Angelo, L. (2010). Exploring Gender Identity in Academic Book Reviews. In: G. Garzone & J. Archibald (eds), *Discourse, Identities and Roles in Specialized Communication*, Vol. 125. Bern: Peter Lang, 363-382.

Denzine, G. M. (1999). An Example of Innovative Teaching: Preparing Graduate Students for Poster Presentations. *Journal of College Student Development* 40, 91-93.

Devitt, A. J. (1991). Intertextuality in Tax Accounting: Generic, Referential, and Functional. In: C. Bazerman & J. Paradis (eds), *Textual Dynamics of the Professions.* Madison, WI: University of Wisconsin Press, 336-355.

Gherardi, S. (2000). Practice-Based Theorizing on Learning and Knowing in Organizations. *Organization* 7, 211-223.
Gherardi, S. (2001). From Organizational Learning to Practice-Based Knowing. *Human Relations* 54, 131-139.
Hay, I., & S. M. Thomas (1999). Making Sense with Posters in Biological Science Education. *Journal of Biological Education* 33, 209-214.
Hess, G. R., K. Tosney & L. Liegel (2010). Creating Effective Poster Presentations. http://www.ncsu.edu/project/posters. 05.05.2011.
Howenstine, E., I. Hay, E. Delaney, J. Bell, F. Norris, A. Whelan, M. Pirani, T. Chow & A. Ross (1988). Using a Poster Exercise in an Introductory Geography Course. *Journal of Geography in Higher Education* 12, 139-147.
Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow: Longman.
Kress, G., & T. van Leeuwen (1996). *Reading Images: The Grammar of Visual Design.* London: Routledge.
Kress, G. & T. van Leeuwen (2001). *Multimodal Discourse: The Modes and Media of Contemporary Communication.* London: Arnold.
Matthews, L. D. (1990). The Scientific Poster: Guidelines for Effective Visual Communication. *Technical Communication* 37, 225-232.
Miller, C. R. (1984). Genre as Social Action. *Quarterly Journal of Speech* 70, 151-167.
Miracle,V. A. (2003). How to Do an Effective Poster Presentation in the Workplace. *Dimensions of Critical Care Nursing* 22, 171-172.
Paré, A., & G. Smart (1994). Observing Genres in Action: Towards a Research Methodology. In: A. Freedman & P. Medway (eds), *Genre and the New Rhetoric.* London: Taylor & Francis, 146-154.
Salzl, G., S. Gölder, A. Timmer, J. Marienhagen, J. Schölmerich & J. Grossmann (2008). Poster Exhibitions at National Conferences: Education or Farce? *Deutsches Ärzteblatt International* 105, 78-83.
Scott, M. (2004). *WordSmith Tools - Version 4*. Oxford: Oxford University Press.
Stoss, F. (2010). Poster Presentations. Designing Effective Posters. http://library.buffalo.edu/libraries/asl/guides/bio/posters.html. 05.05.2011.
Swales, J. (2004). *Research Genres: Exploration and Applications*. Cambridge: Cambridge University Press.
Swales, J. & Feak, C. (2000). *English in Today's Research World: A Writer's Guide*. Michigan: Michigan University Press.
Tham, M. T. (1997). Poster Presentation of Research Work. http://lorien.ncl.ac.uk/ming/dept/tips/present/posters.htm. 03.05.2011.
Thompson, G. (2001). Interaction in Academic Writing: Learning to Argue with the Reader. *Applied Linguistics* 22, 58-78.
Tufte, R. E. (1990). *Envisioning Information*. Cheshire, Connecticut: Graphics Press.
van Naerssen, M. (1984). Science Conference Poster Presentations in an ESP Program. *The ESP Journal* 3, 47-52.
Wenger, E. (2000). Communities of Practice and Social Learning Systems. *Organization* 7, 225-246.
Woolsey, J. D. (1989). Combating Poster Fatigue: How to Use Visual Grammar and Analysis to Effect Better Visual Communication. *Transactions in Neurosciences (TINS)* 12, 325-332.

Conjunctive Adverbials in Academic Written Discourse: a Corpus Analysis Based on a Sociolinguistic Approach

Irena Hůlková

Abstract

The present paper looks into one of many features of academic English and that is the presence of conjunctive adverbials which contribute to the overall lucid stratification of a text and thus to its coherence. Based on a sociolinguistic approach, the paper attempts to investigate possible differences in men's and women's use of conjunctive adverbials in terms of their frequency of occurrence (of different semantic categories as well as individual representatives) and it also attempts to trace interdisciplinary diversities. The analysis, which is based on a research article corpus of 50 published papers from five different academic disciplines totalling 350,000 words, provides a brief excursion to written academic English represented by the subregister of research articles.

1. Introduction

It cannot be doubted that English as a lingua franca occupies a dominant role in an academic environment; it is used by an international readership and is of paramount importance to all those involved in tertiary education, such as students, teachers and researchers. When compared to other registers, English academic prose – with its highly specialist nature – is rather distant from the general-use means of communication of all native speakers, for most of them are not involved in this register on a regular basis. Yet it can also be looked at as 'global' (Biber et al. 1999: 16), in that it is aimed at "an international audience with relatively little influence from the national dialect of the author" (ibid.). Regarding this, the register of academic prose should, in my opinion, be debated in depth, as it offers a plethora of issues to be investigated.

2. The register of academic prose

The chief function of academic prose is referential, which means that the overall impact of any piece of discourse produced within the register in question should be explicit, unambiguous and logical. To add, as Knittlová (1990) points out, the primary communicative purpose of the register of academic prose is to convey ideas from different scientific disciplines precisely, cogently and thoroughly (ibid.: 27). In order to achieve this purpose successfully, academic prose uses various linguistic means (i.e. lexical and grammatical items) typical of this register, among which probably the most obvious feature is its vocabulary (Crystal & Davy 1969: 251). Knittlová (1990) states that the more 'scientific' a

particular discipline is, the more specialized terminology it requires, which suggests that the lexical repertoire of individual academic disciplines is rather modest, and, in fact, stereotypical (ibid.: 27). However, this should not be perceived as a negative aspect, for it in fact corresponds with the main functions of the register of academic prose (e.g. in that it provides an unambiguous and precise understanding of a scientific text) (ibid.: 27-28).

With regard to the grammatical attributes, three word classes are particularly predominant in academic prose: nouns, adjectives, and prepositions, where nouns take up approximately 60 per cent of all content words (Biber 2006: 14-15). In terms of the frequency of occurrence of verbs, they are much less frequent in academic prose than in other registers; nevertheless, there are certain verb categories that are typical of this register, such as the copulas *be* and *become*, 'existence verbs' like *contain, include, indicate, involve, represent* (Biber et al. 1999: 364) and also derived verbs (primarily those formed with *re-* and *-ize*) (ibid.: 16). Another distinctive feature in connection with the use of verbs is the higher frequency of occurrence of the passive voice in academic prose as compared to the other registers. As Biber and Conrad (2009) declare, about 25 per cent of all finite verbs are used in the passive (ibid.: 116-117).

As for sentence structure, academic prose has standard syntax with complete sentences that are organized in a logical order and the relationship of which is typically indicated unambiguously and clearly. One effective means to achieve this is the use of conjunctive adverbials, which is the subject matter of the present paper. Generally speaking, their occurrence is investigated in the register of academic prose; however, at a more specific level it is only the subregister of research articles that is represented here and discussed in the following section.

2.1. The subregister of research articles

What some linguists (e.g. Martin 1985, Couture 1986, Swales 1990, 2004) label 'genre', others (such as Halliday in Halliday & Hasan 1989, Biber et al. 1999, Biber 2006, Biber & Conrad 2009) call 'subregister' or 'register'. Biber & Conrad (2009) suggest that register, genre, and style should be looked at as three different perspectives on text varieties rather than different kinds of texts. This means that the same text can be analyzed from all these perspectives or approaches (ibid.: 15). While in the register perspective the analysis focuses on the pervasive or typical linguistic characteristics such as frequent words or grammatical features that are functional, and it can thus be based only on text excerpts, the genre perspective, by contrast, focuses on language features that can occur only once in the whole text, in most cases at the beginning or ending boundary (e.g. an abstract at the beginning of a research article and 'Conclusions' at its end), are typically conventional (rather than functional) and can only be analyzed in complete texts (ibid.: 6-7). Taking this into account, the type of the present analysis, which focuses on words (i.e. conjunctive adverbials) that are pervasive and frequent, corresponds to Biber & Conrad's register perspective, and the term 'subregister' is preferred in this paper to denote the text variety of research articles.

The general purpose of the subregister of research articles (RAs) is of course akin to that of the register of academic prose, which means that it also conveys information; however, at a more specific level, RAs also have to "contribute new knowledge to the field and convince other experts that this knowledge has scientific merit" (Biber & Conrad 2009: 126). Similarly, Swales (1990) also points out that a research article is not in fact complete until it is published and thus "made available to the wider research community" (ibid.: 94). Moreover, due to the need for publishing, "the research article is a gargantuan genre" (ibid.: 95) and "has become the standard product of the knowledge-manufacturing industries" (Knorr-Cetina 1981 as cited in Swales 1990: 95).

When defining the research article in terms of its format, Swales (1990) describes it as a piece of written text – often including non-verbal elements – that is usually no longer than a few thousand words (ibid.: 93). He then discusses the structure of the RA as a whole as well as of its constituent parts known as IMRD pattern, i.e. Introduction, Methods, Results and Discussion (ibid.: 11). Biber & Conrad (2009) further explain that these four components can in fact be viewed as very specific subregisters within the subregister of RAs, when the basic difference among them lies in their communicative purpose (ibid.: 45).

Another way of analyzing research articles, which is predominant nowadays, is with respect to the so-called rhetorical moves, which Swales first introduced in 1981 and later (in 1990 as well as 2004) developed and reworked. Swales (2004) defines 'move' as "a discoursal or rhetorical unit that performs a coherent communicative function in a written or spoken discourse" (ibid.: 228), and uses it to analyze research article introductions that consist of three moves representing the revised *Create a Research Space* (CARS) model (Swales 1990: 141); however, analogically and with slight adjustments, this process could also be used for the analysis of the Methods, Results and Discussion sections. In terms of the textual structure of the research article, it can be summed up as follows:

- the average length of articles is about 8,000 words and there are significant interdisciplinary differences ranging from some 5,000 words up to 10,000 or 11,000 per paper (based on my own corpus findings)
- references are distributed throughout the whole RA and thus show that every stage of the study relates to the work by other authors (Bazerman 1984 as cited in Swales 1990: 115) – Biber & Conrad (2009) confirm that the density of article citations is very high and similarly to Bazerman (1984) state that this became frequent only in the twentieth century (Biber & Conrad: 164)
- other linguistic features (many of which can also be mentioned at a more general level, i.e. in connection with the register of academic prose) include, for example, no first person references (instead, agentless passives and inanimate subjects are used), the use of the passive voice, limited use of verbs and, by contrast, frequent occurrence of nouns and complex noun phrases, and common use of noun-noun sequences (i.e. nouns modifying other nouns) (Biber & Conrad 2009: 162-165).

3. Conjunctive adverbials

As already mentioned, only one particular group of words occurring in the subregister of research articles is looked into in the present study, and it is the category of conjunctive adverbials. Since this term is not unified in all linguistic studies and there are many other labels used to refer to a group of adverbials that serve a connective function, let me briefly mention the most widespread and significant ones.

Greenbaum (1969), Quirk et al. (1985), and Greenbaum & Quirk (1990) use the label 'conjuncts'. They regard conjuncts to be one of the grammatical functions of adverbials, and the only difference is that Greenbaum (1969: 230) distinguishes three functions – adjuncts, conjuncts, and disjuncts, whereas Quirk et al. (1985: 417) and Greenbaum & Quirk (1990: 162) list four, i.e. adjuncts, conjuncts, disjuncts, and subjuncts. Biber et al. apply the term 'linking adverbials' (ibid.: 761). Likewise, Leech & Svartvik (2002) also use a two-word term with the noun 'adverbial', but they pre-modify it with 'sentence', thus providing the label 'sentence adverbials' (ibid.: 187). The label 'connective adjuncts' is used by Huddleston & Pullum (2002) in *The Cambridge Grammar of the English Language* (ibid.: 775).

As can be observed, the terms listed above are numerous and display a certain variety. Still, they have something in common: very often, they comprise adjectives such as 'conjunctive', 'connective', 'connecting', or 'linking', and nouns like 'adjuncts', 'adverbs', or 'adverbials'. It is thus more than obvious what their main function in a piece of discourse is and which form they usually take: they play an important cohesive role not only in academic prose, but also in other registers, and the majority of them are realized by single adverbs, although in academic prose "prepositional phrases are also relatively common" (Biber et al. 1999: 884).

In the present paper the term 'conjunctive adverbials' (CAs) is used. Its aptness resides in the combination of the two words 'conjunctive' and 'adverbial', which point out the fundamental attributes of this category: its connective function and the fact that it behaves as an adverbial in a sentence. Coincidentally, it is a term used by Celce-Murcia & Larsen-Freeman in *The Grammar Book* (1999).

4. Research objectives

The primary objectives of the present research can be postulated as follows:

- the frequency of occurrence of all CAs in the research article corpus as a whole,
- the frequency of occurrence of the most common CAs in five different academic disciplines (namely, adult education, management, politics, psychology, and sociology),

- the use of CAs by men versus women authors – this objective is motivated by the fact that generally speaking men and women do not communicate in the same way. Holmes (1995), for example, believes that women "interact more cooperatively and focus on relative closeness" (ibid.: 7), and Fjelkestam-Nilsson (1983) claims that women's language tends to be more emotive and uncertain (ibid.: 120). The present analysis, therefore, attempts to find out whether there are also any differences in the use of CAs,
- the frequency of occurrence of six semantic categories of CAs (a) in the RA corpus as a whole, (b) with respect to the author's gender (i.e. in male versus female part of the corpus), and (c) in individual academic disciplines.

4.1. Corpus

The present study is based on a corpus consisting of 50 published research articles and totalling approximately 350,000 words. The articles represent five different academic disciplines, in particular adult education (AdE), management (Mng), politics (Pol), psychology (Psy), and sociology (Soc), and each discipline comprises ten RAs (five written by men and five by women authors). The average length of the article, which has already been modified, is about 6,500 to 7,000 words. The modification includes deletion of all the material that is not the author's coherent writing such as citations, examples, or the list of references. All the articles were downloaded in pdf format from 'SAGE Journals Online', converted into MS Word format, modified and then searched for CAs.

4.2. The frequency of occurrence of all CAs in the whole RA corpus

The total number of CAs is 2,941, which means that the frequency count per 10,000 words is 85 CAs (the exact size of the whole corpus is 346,231 words). Altogether, there were 90 different CAs used as follows (and listed in order of the frequency of occurrence):

> *however, e.g., for example, thus, i.e., first, (a)-(b)-(c)…, therefore, furthermore, finally, then, moreover, second, yet, in addition, that is, specifically, in other words, similarly, hence, rather, (1)-(2)-(3)…, instead, on the other hand, for instance, at the same time, in contrast, of course, third, also, in particular, nevertheless, consequently, in this way, as a result, further, conversely, again, in sum, likewise, so, nonetheless, on the one hand, subsequently, alternatively, by contrast, namely, next, overall, additionally, fourth, too, contrarily, for this reason, last, still, more specifically, fifth-sixth…, on the contrary, in summary, that said, taken together, after all, though, once again, to begin with, now, otherwise, accordingly, as a consequence, on the other, to summari(z,s)e, meanwhile, having said that, lastly, by comparison, initially, alternately, in conclusion, all in all, secondly, in this case, admittedly, in a more specific tone, in similar fashion, in a similar manner, eventually, correspondingly, in the same way, in a similar vein*

All instances were examined in context to ensure that they functioned as conjunctive adverbials in the texts and the above list of 90 potentially productive search items was assembled on the grounds of those listed in Greenbaum (1969), Quirk et al. (1985) and Biber et al. (1999), and also new ones that were – based

on Greenbaum's (1969) and Quirk et al.'s (1985) classification criteria – assessed as CAs (i.e. conjuncts in Greenbaum's and Quirk et al.'s terminology). It should, however, be pointed out that all the CAs following *in conclusion* (in the above list) occurred only once in the whole corpus, which means there was only one particular author (out of fifty) who used it. Similarly, if we considered only the CAs the occurrence of which represents at least one per cent of the total (i.e. of 2,941 CAs), our list would comprise the first twenty-five items starting with *however* and closing with *for instance*; these represent 83 per cent of the total (for more detail, see Table 1 below).

CA	**frequency**	**% of the total 2,941 CAs**
however	438	14.9
e.g.	284	9.7
for example	228	7.8
thus	186	6.3
i.e.	145	4.9
first	98	3.3
(a), (b), (c), ...	96	3.3
therefore	91	3.1
furthermore	86	2.9
finally	82	2.8
then	82	2.8
moreover	75	2.6
second	70	2.4
yet	67	2.3
in addition	61	2.1
that is	41	1.4
specifically	41	1.4
in other words	39	1.3
similarly	39	1.3
hence	37	1.3
rather	35	1.2
(1), (2), (3), ...	34	1.2
instead	33	1.1
on the other hand	29	1.0
for instance	28	1.0
total	**2,445**	**83**

Table 1: Twenty-five most frequent CAs in the RA corpus as a whole

As can be seen, the first position of *however* confirms Biber et al.'s findings (1999) that "to mark contrast, *however* is uniformly preferred" (ibid.: 889). However, we can see that its very frequent use substantially outnumbers not only contrastive/concessive CAs, but all CAs in general, irrespective of the semantic role they perform. *However* as such comprises almost 15 per cent of all CAs found in the whole corpus.

Another two CAs occurring with high frequency are *e.g.* (almost 10% of the total) and *for example* (almost 8% of all CAs). This is worth a note, because unlike Biber et al. (1999), who claim that *e.g.* is slightly more restricted in use than *for example* and *for instance* (for it is not normally used in initial position) and it is more commonly used in textbooks (rather than research articles) (ibid.: 890), the present findings show that *e.g.* is used in research articles very frequently and it outnumbers *for instance* (which only represents 1% of the total) as well as *for example*. It is true, though, that despite its frequent occurrence, *e.g.* usually provides background information rather than the principal ideas and thus most examples indicated with *e.g.* are within parentheses, confirming Biber et al. (ibid.).

In terms of the resultive/inferential CAs such as *therefore, thus* and *hence*, these are in most cases interchangeable (Biber et al. 1999: 889); however, my findings show that it is *thus* that is used most often (6.3% as compared to *therefore* comprising 3.1% and *hence* 1.3%). As for *hence*, I can confirm Biber et al.'s findings that if it is used by a particular author, then he or she will use it repeatedly throughout the whole text even though other conjunctive adverbials could have been used (ibid.: 890).

4.3. The frequency of occurrence of CAs by discipline

The present analysis shows that the use of CAs varies significantly according to discipline. The disciplinary variations can be observed in spite of the fact that the academic disciplines examined here all represent soft knowledge fields (i.e. the humanities and social sciences; Hyland 2007). Table 2 displays differences in the frequency of occurrence by discipline.

discipline	No. of CAs per 10,000 words
management	121
politics	86
sociology	74
adult education	68
psychology	58

Table 2: Frequency of occurrence of CAs by discipline

The academic discipline management shows by far the highest frequency of occurrence of CAs, and, in contrast, psychology the lowest. This disparity confirms the fact that individual academic disciplines do not mediate reality in

the same way. They differ from one another in vocabulary, terminology as well as overall layout, which also influences the way in which conjunctive adverbials are used. We can also notice that although the use of CAs is idiosyncratic and individual authors may demonstrate a preference for different CAs, the use of CAs by individual authors within a particular discipline does not usually differ to such an extent as when comparing authors from different academic disciplines. Thus, for instance, management (in contrast to the other four disciplines analyzed here) uses a massive amount of the CA *e.g.*, which vastly outnumbers even the most common *however*. Hyland (2007) states that "examples in soft knowledge fields represent a heavier rhetorical investment in contextualisation, perhaps even a need to persuade the reader that the phenomenon actually exists" (ibid.: 272) and explains that giving examples is an efficient way in which the writer attracts the reader's attention and encourages him or her "to recognise phenomenon through recoverable experiences and to become involved in the unfolding text (ibid.). Table 3 provides an overview of five most frequent CAs by discipline.

discipline	CA 1	CA 2	CA 3	CA 4	CA 5
management	e.g. (18.5)	however (12.2)	i.e. (9.0)	for example (8.4)	thus (7.4)
politics	however (15.1)	thus (6.2)	first (5.7)	finally (5.4)	therefore (4.1)
sociology	however (18.7)	thus (7.2)	for example (6.1)	yet (5.0)	moreover (4.4)
adult education	however (15.3)	for example (10.3)	e.g. (9.2)	furthermore (6.1)	(a), (b), (c), ... (5.9)
psychology	however (17.6)	for example (11.9)	thus (7.4)	therefore (6.4)	i.e. (6.1)

Table 3: Five most frequent CAs by discipline (% of total)

Except for management, in all the other disciplines *however* is the most frequent conjunctive adverbial, which of course confirms the fact that it is overall the most common CA in academic prose. Otherwise, besides *however*, no other CA occurs among the first five most frequent cases in all the analyzed disciplines; however, certain similar tendencies can be observed. Excluding adult education *thus* belongs to the five most frequent CAs in all the other disciplines. In the same way, *for example* is represented among the most common CAs in all the disciplines except for politics. The rest of the most frequent conjunctive adverbials in different disciplines vary to a certain extent: politics, for example, favours the listing (enumerative) items *first* and *finally*; sociology uses the concessive *yet* and listing (additive) *moreover* quite frequently, and adult education, by contrast, shows a preference for the listing (additive) *furthermore* and (enumerative) *(a), (b), (c), ...* .

In sum, although all disciplines under investigation have points of resemblance to one another in terms of the general use of conjunctive adverbials, at a more specific level we can conclude that the present corpus analysis shows

disciplinary variations concerning the frequency of occurrence as well as the choice of particular representatives.

4.4. The use of CAs by men versus women authors

From a sociolinguistic point of view, the present analysis also attempts to investigate possible differences in men's versus women's use of conjunctive adverbials; however, no dramatic variation in connection with the author's gender have been observed. Table 4 displays the top ten conjunctive adverbials used by men and women authors in the present corpus.

MEN		WOMEN	
CA	% of total	CA	% of total
however	15.1	*however*	14.7
e.g.	10.3	*e.g.*	9.0
for example	7.0	*for example*	8.5
thus	6.9	*thus*	5.7
i.e.	5.8	*(a), (b), (c), ...*	4.3
first	3.8	*i.e.*	4.0
then	3.2	*therefore*	3.7
finally	2.8	*furthermore*	3.2
second	2.7	*in addition*	2.9
furthermore	2.7	*first*	2.8

Table 4: Ten most frequent CAs as used by men and women authors

When we look at Table 4 comparing men's and women's use of CAs, we can notice that the first four CAs are identical (*however, e.g., for example* and *thus*) in terms of the general preference (to other CAs) and very similar in terms of the frequency of occurrence. The rest of the most common CAs slightly differ in individual representatives; thus, for example, women favour the listing (enumerative) *(a)-(b)-(c)…*, while men give preference to *first-second-then-finally*. Both men and women authors use *furthermore* for additional ideas and women also use the listing (additive) *in addition* quite frequently. Table 5 compares the use of CAs by men and women in individual disciplines.

Again, we can see that the difference in the average frequency of occurrence of CAs as used by men and women is insignificant, i.e. 84 to 86 tokens per 10,000 words. In a similar manner, the distribution of CAs by men and women in a particular discipline is more or less similar. In some disciplines, men use more CAs than women (e.g. sociology and psychology), the divergence in number, however, is marginal.

discipline	MEN CAs 10,000 words		WOMEN CAs 10,000 words
management	117	<	125
politics	79	<	94
sociology	78	>	71
adult education	66	<	69
psychology	62	>	54
total	84	<	86

Table 5: CAs as used by men and women in individual disciplines

4.5. The frequency of occurrence of six semantic categories of CAs

Based on Greenbaum's (1969), Quirk et al.'s (1985), and Biber et al.'s (1999) classification, I suggest six semantic categories (and some sub-categories) of CAs as follows and provide them here with a few examples:

(1) appositional:
  (a) exemplification: *e.g., for example, for instance*
  (b) reformulation: *i.e.*, *that is*, *in other words*, *namely*
(2) listing:
  (a) enumerative: *first*, *second*,…, *(a), (b), (c), …*, *next*, *last*
  (b) additive: *furthermore*, *moreover*, *in addition*, *similarly*
(3) contrastive/concessive: *however, yet, nevertheless, instead, conversely*
(4) resultive/inferential: *thus, therefore, hence, consequently, as a consequence*
(5) summative: *in sum, overall, all in all*
(6) transitional:
  (a) discoursal: *now*
  (b) temporal: *subsequently*, *meanwhile*

Table 6 shows the frequency of occurrence of the six major semantic categories as presented above.

semantic category of CAs	% of total
appositional	27.9
listing	27.6
contrastive/concesssive	27.5
resultive/inferential	15.3
summative	1.1
transitional	0.6

Table 6: The frequency of occurrence of CAs according to semantic categories in corpus as a whole

Unlike Biber et al. (1999), who state that the commonest semantic group in academic prose is that of result/inference, which "marks the conclusions that the writer expects the reader to draw" (ibid.: 881), my analysis shows that there are other semantic groups that are more frequent. In fact, three categories occur in almost identical frequency, these are appositional (27.9%), listing (27.6%), and contrastive/concessive (27.5%) conjunctive adverbials. The category of result/inference also occurs quite frequently (15.3% of the total); however, in comparison to the first three categories, its occurrence is significantly lower. It is not possible, though, to compare Biber et al.'s and the present findings entirely, for I only look into the subregister of research articles, while Biber et al. investigate the very broad register of academic prose.

semantic category of CAs	MEN	WOMEN
appositional	24	23
listing	21	26
contrastive/concessive	24	23
resultative/inferential	13	13
summative	1	1
transitional	1	0

Table 7: The frequency of occurrence of different semantic categories of CAs per 10,000 words – men and women compared

Except for the category of listing CAs, where women use slightly more often (26 to 21 tokens per 10,000 words), there are no other significant differences. This corresponds to the fact that in general the use of CAs is not determined or considerably influenced by the author's gender.

The third aspect looked into in connection with semantic categories is their frequency of occurrence in the individual academic disciplines included in the present study (see Table 8 for more detail).

Except for management, in which the category of appositional CAs is by far most common (39.8%), in the other four disciplines it is either listing (politics 34.9% and adult education 32.3%) or contrastive/concessive CAs (sociology 37.5% and psychology 32.1%) that occur most frequently. The high frequency of occurrence of appositional CAs in management is caused by its representative *e.g.*, which is extensively used in this discipline (see Table 3 above). Similarly, the appositional CAs *i.e.* and *for example* also occur to a considerable extent here. As mentioned earlier, this may be caused by the specific need of the discipline of management to provide enough examples that will convince the reader of the real existence of the phenomenon under scrutiny, because exemplification makes the author's ideas accessible and persuasive (Hyland 2007: 270). Reformulation (introduced, e.g., by the above *i.e.*) then helps to facilitate comprehension (ibid.: 269).

discipline	management		politics		sociology		adult education		psychology	
		%		%		%		%		%
appositional	432	**39.8**	100	15.9	87	19.0	119	26.0	82	26.3
listing	260	24.0	219	**34.9**	113	24.6	148	**32.3**	74	23.7
contrast /concession	215	19.8	188	29.9	172	**37.5**	130	28.4	100	**32.1**
result/inference	158	14.6	104	16.6	85	18.5	51	11.1	53	17.0
summative	14	1.3	12	1.9	1	0.2	4	0.9	2	0.6
transitional	5	0.5	5	0.8	1	0.2	6	1.3	1	0.3
total	**1084**		**628**		**459**		**458**		**312**	

Table 8: The frequency of six semantic categories of CAs in five different academic disciplines (highest frequency in bold)

Another significant disciplinary variation can be observed in connection with management again and the category of contrast/concession. While this category is very frequent in the remaining disciplines (sociology 37.5%, psychology 32.1%, politics 29.9% and adult education 28.4%), in the case of management it is represented only by 19.8 per cent (even so, it still represents the third most common semantic category in this discipline).

The other semantic categories such as resultive/inferential, summative and transitional do not show vast disciplinary differences in the frequency of occurrence. The latter two are very rare in all five disciplines under investigation.

5. Conclusions

The present paper looked into a group of cohesive means represented by conjunctive adverbials (CAs) and their distribution in the subregister of research articles. Generally speaking certain CAs are used and favoured by the majority of authors (e.g. *however, e.g., for example, thus, i.e.*) regardless of their gender and expertise. Nevertheless, at a more specific level, we can see that the use of CAs varies according to discipline to some extent (even though all the disciplines in the present corpus represent the so-called soft knowledge field, i.e. the humanities and social sciences) and is idiosyncratic (i.e. dependent upon the individual's preferences).

The disciplinary variations mirror different needs, goals and ways in which the writers represent themselves, their research and findings, and in which they want the readers to interpret their message. The highest number of CAs was used in management (121 tokens per 10,000 words), where the appositional *e.g.* vastly outnumbered all the other CAs (in the other four disciplines it was *however* that occurred most frequently), the lowest in psychology (58 cases per 10,000 words).

From a sociolinguistic point of view, the present findings indicate that CAs represent a linguistic phenomenon which is not influenced by the author's gender, which means that men as well as women use CAs in a comparable way

(both in terms of the frequency of occurrence and the choice of individual semantic categories). This may be caused by the fact that CAs as such belong to the typical cohesive means in academic prose and are also expected to be used for reasons of good style in this register, and thus all authors – irrespective of gender – share the same goal, i.e. to produce a coherent, conventional and formal piece of writing.

Irena Hůlková
Faculty of Education
Masaryk University
Brno
irena.hulkova@email.cz

References

Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers.* Amsterdam / Philadelphia: John Benjamins.

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman Grammar of Spoken and Written English.* London: Longman.

Biber, D. & S. Conrad (2009). *Register, Genre, and Style.* Cambridge: Cambridge University Press.

Celce-Murcia, M. & D. Larsen-Freeman (1999). *The Grammar Book.* Boston: Heinle & Heinle.

Couture, B. (1986). Effective Ideation in Written Text: A Functional Approach to Clarity and Exigence. In: B. Couture (ed.), *Functional Approaches to Writing: Research Perspectives*. Norwood, NJ: Ablex, 69-92.

Crystal, D. & D. Davy (1969). *Investigating English Style.* London: Longman.

Fjelkestam-Nilsson, B. (1983). *ALSO and TOO: A Corpus-Based Study of Their Frequency and Use in Modern English.* Stockholm: Almqvist & Wiksell International.

Greenbaum, S. (1969). *Studies in English Adverbial Usage.* London: Longman.

Greenbaum, S. & R. Quirk (1990). *A Student's Grammar of the English Language.* London: Longman.

Halliday, M. A. K. & R. Hasan (1989). *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective* (2nd ed.). Oxford: Oxford University Press.

Holmes, J. (1995). *Women, Men and Politeness.* London: Longman.

Huddleston, R. & G.K. Pullum (2002). *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Hyland, K. (2007). Applying a Gloss: Exemplifying and Reformulating in Academic Discourse. *Applied Linguistics* 28(2), 266-285.

Knittlová, D. (1990). *Funkční Styly v Angličtině a Češtině.* Olomouc: Univerzita Palackého, Filozofická fakulta.

Leech, G. & J. Svartvik (2002). *A Communicative Grammar of English* (3rd ed.). London: Longman.

Martin, J. R. (1985). Process and Text: Two Aspects of Human Semiosis. In: J. D. Benson & W. S. Greaves (eds.), *Systemic Perspectives on Discourse 1*. Norwood, NJ: Ablex, 248-274.

SAGE Journals Online. http://online.sagepub.com/

Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings.* Cambridge: Cambridge University Press.

Swales, J. M. (2004). *Research Genres: Exploration and Applications.* Cambridge: Cambridge University Press.

Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Hedging in Research Articles: Humanities and Social Sciences Compared

Martina Malášková

Abstract

It has been sufficiently proven that hedges are an indispensable part of academic writing. Their importance lies in the fact that these often quite complex elements can be considered polypragmatic devices that fulfil a variety of functions in discourse. In academic prose style, and research articles in particular, hedges play a crucial role in negotiating and also evaluating the meaning communicated between the writer and the reader. This paper presents the results of a small scale comparative investigation carried out on a corpus comprising research articles from two well-established journals (*Applied Linguistics* and *Essays in Criticism*). The aim of the study was to detect possible similarities and differences in the occurrence of hedging expressions in research articles used by two disciplinary communities from the social sciences and humanities. The investigation, which uses both quantitative and qualitative analyses, is based on a classification of hedges drawing on Hyland (1998). The results of the study show that there seem to be formal and semantic differences in hedging expressions used in linguistics and literary criticism research articles.

1. Introduction

The term 'hedge' was first used by Lakoff in 1972, and the underlying concept of hedging has received a considerable amount of attention since then. Numerous approaches towards this phenomenon have been adopted by different authors. Hedges have been studied in both speech and writing, from the semantic point of view, pragmatic point of view and from the point of view of cultural and gender differences (e.g. Burrough-Boenish 2002). No less attention has been paid to comparative studies trying to identify the peculiarities of different disciplines and their comparison (e.g. Salager-Meyer 1994, Skelton 1997). As a result, there are various accounts of hedging expressions to be found. Most recent approaches emphasise the complex character of hedging expressions with regard to the pragmatic motivations for their use. Hyland (2010) sees hedges as expressions whose role is to mark statements as provisional, enabling the writer to express his/her views and involving the readers in the ratification of the propositional content contained in those statements. Hyland's definition exemplifies the dual character of hedging expressions – they relate to both the propositional information and the recipient of the text. The interpersonal character of hedging has been addressed by many authors (Brown & Levinson 1978, Hübler 1983, Myers 1989). The relationship between hedging and the proposition of the

statement is, however, perceived as slightly more complicated. Some authors consider hedges as being part of the proposition (e.g. Ifantidou 2005), others do not (e.g. Crompton 1997). Generally, hedges are seen as part of commentative language (Skelton 1988) or metadiscourse (Hyland 2005).

The multifunctional character of hedging expressions is especially important in academic writing – and research articles (RAs) in particular – where the writer needs to present the results of his/her scientific work. Using hedges allows the writer not only to invest a certain degree of commitment into the truth value of the proposition, but also to establish and maintain contact with the readers. The relationship between the writer and the reader is essential here since it is in the interaction between the participants of the discourse where meaning is negotiated.

As follows from what has been mentioned above, hedging is a relatively well researched phenomenon. However, there are still areas worth further exploration. One of them is the comparison of the ways in which hedges are used by different discourse communities within the field of soft science. The present paper aims at casting light at possible differences in the types of hedging expressions and their frequency of occurrence in linguistics (L) and literary criticism (LC) RAs .

2. The semantics of hedging

The complex nature of hedging has resulted in numerous semantic classifications. The present study draws mostly on the classification by Hyland (1998), which is currently one of the most comprehensive frameworks in the field. As has already been outlined, hedging expressions fulfil a wide range of functions. These functions are inextricably linked to the three components of the discursive situation – the writer, the reader and the proposition. To get a more comprehensive understanding of the way hedges work in discourse, it is useful to relate the functions of hedges to Halliday's (1978) model of language macro-functions. In his theory Halliday delimits three main functions of language – the ideational, interpersonal and textual.

In simplified terms, the ideational function represents the speaker's/writer's attempt to describe his experience with both the external world and his internal world (Halliday 1978: 45). The interpersonal function has to do with the relationship between the participants of the communicative situation, while the third function – textual – refers to the organization of sentences into a coherent text (ibid.: 46).

Contrary to some authors who focus on hedges from the purely interpersonal point of view (e.g. Ventola 1997, Crismore & Vande Kopple 1997), the present work also takes into consideration the ideational aspect of hedging expressions. This aspect is closely connected to the propositional content itself and reflects the writer's need to achieve the highest possible accuracy in presenting the results of his scientific work. The writer does so by marking the extent to which the information in the proposition is true or applicable to real life phenomena.

The language means used for this purpose are called – in accordance with Hyland (1996) – content oriented hedges (1). Using such devices allows the writer to reflect the reality in academic writing as truthfully as possible.

(1) The effects of author prestige reported in the publication-based research, although small, are probably quite real. (L B)

The remaining elements – the writer and the reader – are concerned with the interpersonal macro-function of language. The following type is typically associated with the persona of the writer and is guided by the writer's need to protect himself/herself from possible criticism. By shifting the responsibility for the propositional content to impersonal structures rather than committing him/herself fully, the writer is shielded from the anticipated negative reactions on the side of the reader. Hedging expressions in this category can be referred to as writer oriented hedges (2).

(2) It seems, then, that Wordsworth is back in Prelude vein: just as when he comes upon Stonehenge or rows across Ullswater in that poem, he here finds himself oppressed by the monumental objects of the material world, when their location and origin are met unexpectedly. (LC A)

The third constituent – the reader – stands behind the motivation for using such structures that involve the recipient of the text in the process of meaning construal. By employing reader oriented hedges (3), the writer guides the reader throughout the text in an attempt to influence the receiver's perception of the information contained in the text.

(3) If we are to remove some of these difficulties, we must first have a clearer and more restrictive definition of the notion of idiom. (LC D)

The following table presents the semantic functions of hedging expressions as discussed above.

content oriented		participant oriented		
		writer oriented	reader oriented	
aim at greater accuracy (precision)	aim at the extent of applicability/ generalizability	protect the writer by depersonalizing the information presented in the proposition	appeal to the reader by employing various strategies of reader involvement	protect the writer by personalizing the information presented in the proposition

Table 1: Semantic classification of hedges

Hedging is very closely connected to another important analytical framework – Speech Act Theory (Austin 1962). The writer chooses a certain language means with a particular kind of pragmatic motivation, or motivations, in mind in order to create the desired perlocutionary effect on the reader. The more appropriate the writer's choice of language means, the higher is the possibility of achieving the ultimate goal of such a communicative situation, which is the acceptance of the proposition by the fellow members of the respective discourse community as defined by Swales (1990). The need for acceptance by the discourse community is what the writer strives for in order to increase his/her scientific credibility. The positive influence of hedging expressions on the reader's attitude has been experimentally proven by Crismore & Vande Kopple (1997), who carried out a survey of the reader's perception of a scientific text about a controversial issue. The results show that a hedged text is perceived by its readers in a more favourable way than the same text with the hedging expressions removed.

Previous research into hedging expressions in RAs (Hyland 1998) showed that one of the main features of almost any hedging expression is the difficulty in identifying the underlying pragmatic motivation for its use. A hedge can fulfil several functions at the same time and therefore trying to devise clear-cut categories of semantic functions of hedges may seem to be of little use. Instead, it is more useful to employ the fuzzy category concept proposed by Lakoff (1973), which reflects the true nature of hedging more precisely. This model takes into consideration both core and peripheral instances of hedges, thus allowing for a more refined investigation.

Based on what has been stated earlier, in the present study hedges are defined as grammatical and strategic means expressing epistemic stance, conveying different degrees of commitment to the truth judgement expressed in the proposition and thus enabling the writer to negotiate meaning in interaction with his/her intended readership.

3. Formal hedging means

It is necessary to point out at an early stage that the formal realizations of hedges, similarly to their semantic roles, are rarely seen as a system of discrete categories. Markkanen & Schröder (1998) claim that "almost any linguistic item or expression can be interpreted as a hedge" and that "no linguistic items are inherently hedgy but can acquire this quality depending on the communicative context" (ibid.: 6). Thus it follows that the interpretation of both semantic roles and forms of hedges is highly context bound. Moreover, one formal realization can fulfil multiple semantic roles and vice versa. Some authors take into consideration purely lexical means of hedging (e.g. Biber et al. 1999), others tend to include also strategic means such as questions, hypotheticals, admission to the lack of knowledge, etc. (Salager-Meyer 1994, Hyland 1996, 1998).

In accordance with Hyland (1998), the present study recognizes two principal formal categories of hedging expressions – lexical and strategic. It also attempts

to relate the principal lexical and strategic means to the above mentioned semantic categories (see Table 2).

content oriented	participant oriented	
	writer oriented	reader oriented
primarily expressed by:	primarily expressed by:	primarily expressed by:
precision adverbs epistemic lexical adjectives epistemic lexical nouns epistemic modal nouns limited knowledge	impersonal structures referring to model/method/theory/ experimental conditions abstract rhetors epistemic lexical verbs attribution to literature	personal engagement structures (personal attribution, reference to methods/models, etc.) reader engagement structures (assumption of shared goals, conditionals, questions, reference to testability, etc.)

Table 2: Formal classification of hedges

The most prominent lexical hedging devices include epistemic modal verbs, precision adverbs, epistemic lexical verbs, adjectives and nouns. Some lexical means are associated with certain semantic roles more often than others, e.g. precision adverbs typically fulfil the semantic function of content oriented hedges operating within the scope of the proposition; epistemic lexical verbs (especially evidential and judgemental) – very often in connection with abstract rhetors – can be regularly found in the position of writer oriented hedges, etc.

The strategic hedging devices comprise more complex structures such as personal and impersonal reference to methods/models/experimental conditions, conditionals, questions, admission to lack of knowledge and various reader involvement strategies. Here it is worth mentioning that, just like the lexical means, even the strategic means are usually associated with particular semantic functions. The personal/impersonal character of reference is, for instance, often the only reliable feature for distinguishing between writer and reader oriented hedges.

4. The genre of research articles

The term ‘genre’ is central to academic writing. However, the definition of genre varies to a considerable extent and is especially problematic in connection with the equally important concept of register. This paper works with the definition of Biber, who argues that genre refers to a “culturally recognized ‘message type’ with a conventional internal structure” (Biber 2006: 11).

The genre of the research article is enjoying a special status in academic settings. The exceptionality of the genre lies in its potential to serve as the means of publishing the results of the scientist’s work, thus promoting and sustaining his/her academic career. The career success of academia is crucially intertwined with the acceptance of one’s work by the discourse community. If the RA in

question is published and subsequently quoted in the work of fellow members of the discourse community, it significantly contributes to the scientist's credibility within the discourse community itself. In order to gain this credibility, the writer not only presents the reader with certain claims, but also provides him/her with more or less subtle guidance on how the claims should be interpreted. In fact, the author communicates with his/her readers via the text and there is actually a lot of interaction going on. The success of this acceptance negotiating strategy is – as previously mentioned – to a great extent dependent on the writer's choice of suitable language means. These language means, which are interactive in nature, are most commonly united under the umbrella term 'metadiscourse' (Hyland 2005). Ädel (2006) rather aptly notes that "metadiscourse is also used by the writer to interact with her imagined reader in ways that create and maintain a relationship with the reader and that allow the writer to influence him by addressing him directly in various ways" (ibid.: 20). Therefore, in accordance with the aforementioned authors, hedges (together with other language means) can be considered part of 'metadiscourse' and as such they help the writer influence the reader's attitude towards the information presented in the text.

It has already been mentioned that each new piece of information published in a RA is subject to careful examination and can be accepted or rejected by the reader. Hence it is considered provisional and has to be marked as such. Since the author has to tend both the propositional information and his/her readership, the two aforementioned semantic types of hedges – context and participant – are employed to present the claims in the RA as pending.

Scientists are obliged, based on community practice, to report the reality as accurately as possible. The first type of hedges – content oriented hedges – enables the writer to express the relationship between the information contained in the proposition and the real *status quo* of the world. This is achieved by adjusting either the precision or the scope of generalizability of the propositional information (4), (5). Such weakening of the strength of the claim may considerably lessen the risk of rejection of the reported findings.

(4) Sometimes, a small number of words form an oscillating pattern, where individual words move between the two states, but it is unusual for these oscillations to be very large. (L C)

(5) In general, the physical setting is mapped and controlled, by a fussy, neurotic, Crusoesque tendency in the narrative to counting, and to the measurement of distances and time. (LC E)

The second major type – participant oriented hedges – is more overtly connected to the relationship between the participants of the communicative situation. There are two major subtypes in this category of hedges: writer oriented and reader oriented hedges. The former subtype is directly aimed at shielding the author from possible criticism on the side of the reader by withdrawing full commitment to the propositional information. This is achieved by making the link between the persona of the author and the claim he/she presents in the text covert or indirect

to a greater or lesser degree by shifting the responsibility for the claim to various means such as impersonal subjects, impersonal reference to methods/models/ theories and experimental conditions, usually in connection with epistemic modal or lexical verbs (6). It is worth mentioning that the impersonal character of this kind of hedging expressions is their prominent feature.

(6) This result may seem inconsistent with the famous 'Matthew effect', named by Merton himself, which predicts that authors who are already well-cited will get a disproportionate share of additional citations (Merton 1968, 1998). (L B)

Reader oriented hedges, just like the writer oriented ones, also help the writer increase the possibility of the acceptance of his/her work. The main distinctive feature of reader oriented hedges is their personal character. These means either appeal to the reader by presenting the writer's claims as just one of the possibilities (7), or by more or less directly involving the reader (8). Treating the readers as members of the same discourse community increases the chance of the reader adopting the writer's point of view, thus increasing the possibility of favourable reception of the text.

(7) Both studies are concerned with mezzo institutional settings; and I suspect that 'culture' rather than 'discourse' is used in the latter only because there is a beginning of large culture difference, which ironically proves a red-herring. (L A)

(8) A 'to' is missing, too, at the beginning of the final line, where the ghostly iambic metre, as well as grammatical convention, would lead us to expect one.(LC B)

5. Material and methods

The study aims at comparing hedging expressions used in linguistics (L) and literary criticism (LC) research articles. The research corpus comprises ten RAs, five of them were obtained from *Applied Linguistics* and five from *Essays in Criticism*; all the articles were chosen randomly out of a larger corpus of linguistics and literary criticism RAs. Both journals are published by Oxford University Press, and since these journals are well-established within their discourse communities, they can be taken as prototypical examples of the genre. The selected articles were published between the years 1999 and 2009. The length of the corpus is approximately 75,000 words. As far as the L1 background of the writers is concerned, only native speakers of English were chosen for the present study to avoid interference since it has been proven that the use of hedges is to a certain extent culturally determined (e.g. Nikula 1998, Burrough-Boenish 2002).

As already mentioned, one of the characteristics of the genre is its conventional internal structure (Biber 2006). One of the most comprehensive theories on the structure of the genre of RAs is that of Swales (1990). As far as the linguistics RAs are concerned, the IMRD structure (ibid.) was followed in all instances. However, none of the literary criticism RAs has any clearly recognisable structure in terms of sections. The overall character of the literary criticism RAs is also rather

descriptive and when compared to linguistics RAs, it becomes apparent that there is a certain degree of intra-generic variation, with the literary criticism RAs being fairly non-prototypical with regard to the IMRD theory.

Both quantitative and qualitative analyses were carried out with the use of the previously mentioned formal and semantic classifications (see Tables 1 and 2). The qualitative analysis focused on discovering the semantic types of hedges occurring in both types of RAs as well as their principal formal realisations. As already pointed out, hedges are an extremely context bound phenomenon; therefore it was sometimes rather difficult to distinguish between the individual semantic categories. One of the feasible possibilities of coping with this difficulty seems to be the use of the fuzzy category model as proposed by Lakoff (1973). This approach enables us to take into consideration all instances of hedges ranging from the core (or prototypical) to the peripheral (less prototypical) ones and classify them with regard to their prevailing semantic function.

The study has worked with several hypotheses. Firstly, based on previous research (Malášková 2009) and the character of both types of RAs, it was assumed that both the formal and semantic aspects of hedging means will be manifested differently in the two types of examined texts. Secondly, it was expected that in literary criticism RAs the most common type of hedges will be reader oriented ones. Content oriented hedges were predicted to prevail in linguistics RAs since these are supposed to contain a larger amount of numeric data which are typically hedged by content oriented hedges.

6. Results and discussion

Table 3 summarizes the results of the quantitative analysis of both types of RAs. It shows that in both of them all three semantic types of hedges were present.

	content oriented	writer oriented	reader oriented	total
linguistics	289	377	253	**919**
literary criticism	142	250	143	**535**

Table 3: Results of the quantitative analysis

As can be seen from Table 3, writer oriented hedges were the most common semantic type in the linguistics RAs. At this point it is already obvious that the results refute the prediction that the content oriented hedges will be most common in linguistics RAs. It seems that the writer's need to shield him/herself from possible objections coming from readers exceeds other possible pragmatic motivations.

Moreover, this seems to be the case also with literary criticism RAs, where – just as in the linguistics RAs – writer oriented hedging expressions were the most frequently used semantic type. It is also worth mentioning that the linguistics

RAs are more heavily hedged than the literary criticism ones. A possible explanation seems to lie in the fact that literary research articles are inherently more subjective than linguistics ones; therefore literary critics do not need to employ so many hedging expressions since their opinions and claims are a priori perceived as just one of many possible alternatives.

It is of interest that in both linguistics and literary criticism RAs the number of content oriented hedges is relatively equal to the number of reader oriented ones. The reason for this is somewhat difficult to grasp and, with regard to the pragmatic motivation for the use of the two semantic types of hedges, it can only be speculated that the writers feel the need to pay equal attention to both their readership and the propositional content itself. The following figures represent the division of individual semantic types of hedges per individual article.

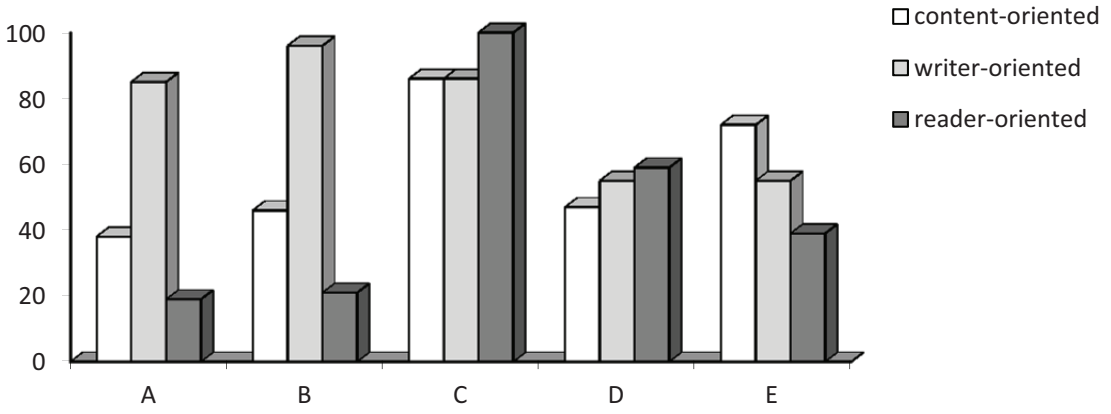


Figure 1: Linguistics RAs (individual articles breakdown)

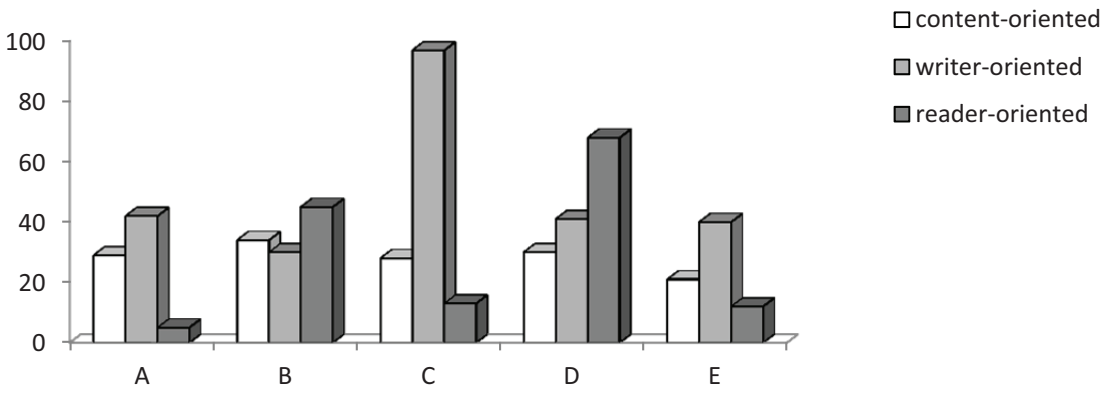


Figure 2: Literary criticism RAs (individual articles breakdown)

The breakdown of all three semantic categories of hedges per individual article indicates that some articles within each of the fields tend to be more heavily hedged than others. In some cases the discrepancy is quite striking, especially in linguistics (see Figure 2). It is obvious from the results that the individual

writer's preferences are reflected in the use of hedges, but based on the size of the corpus it is not possible to make any generalizations. In the field of literary criticism there are also notable differences in the frequency of use of individual semantic types, which indicates that the pragmatic motivation for the use of hedges varies among the members of the same discourse community. Each semantic type of hedge will now be treated in more detail with regard to the differences occurring in the two types of texts.

type of device	occurrences L	occurrences LC
precision adverbs (style disjuncts, content disjuncts, downtoners)	146	68
epistemic modal verbs	100	46
epistemic modal adjectives	21	10
limited knowledge	9	4
other (epistemic lex. verbs, nouns, etc.)	13	14
total	**289**	**142**

Table 4: Content oriented hedges

Content oriented hedges are present in both types of texts with the precision adverbs being the most common, closely followed by epistemic modal verbs. This category of hedges operates within the scope of proposition and expresses probability or possibility (9), limited extent (10) and degree of precision (11).

(9) The size of the recognition window may range from a few words on either side of the reference to a multi-sentence passage. (L B)
(10) In some cases, perhaps the majority of cases given the prevalence of metaphor, this reinterpretation will be done in terms of an image (or image schemata, Gibbs 1992). (L D)
(11) Indeed some may feel that, for many of our students, Titus has become almost too popular, too central, a text. (LC D)

The most frequent type of hedges in both examined text types – writer-oriented hedges – serve mainly as writer protection. This effect is achieved primarily by the employment of impersonal structures in connection with epistemic lexical or modal verbs. As follows from Table 5, the most common impersonal devices in the analyzed texts were abstract rhetors (12), attribution to literature (13) and empty subjects (14), closely followed by impersonal reference to method/model/ experimental conditions (15).

type of device	occurrences L	occurrences LC
epistemic lexical verb	121	83
abstract rhetors	62	55
epistemic modal verbs	44	25
attribution to literature	44	27
empty subjects	34	15
modalized passive voice	28	19
hypotheticals	24	19
impersonal reference to method/ model/experimental conditions	17	2
other	3	5
total	**377**	**250**

Table 5: Writer oriented hedges

The frequent occurrence of these self-protective items indicates that writers in both fields take the risk of possible criticism seriously and attempt to lower the risk by reducing the commitment to the propositional content and greater invisibility in the text.

(12) These later, popular, place poems forcefully suggest that the common understanding of his later career as a decline into Toryism needs revision. (LC A)
(13) Hamp-Lyons (1996) has suggested that miscommunication between editors and authors is common for both native and non-native writers [...] (L E)
(14) It seems natural to Elizabeth, and to the other English in the novel, that Europeans should hold sovereignty over the East, just as it seems natural that human beings should stand above and dominate the natural world. (LC E)
(15) In entangled models, the basic features of a Boolean Network model are maintained, but some constraints are imposed on the way words are linked to other words, and a new type of Boolean unit is introduce. (L C)

As previously mentioned, the prime motivation for the use of reader-oriented hedges is to appeal to the implied readership, thus enhancing the possibility of favourable reception of the research article. The writer tries to draw the reader into the process of negotiating the meaning by the use of various reader involvement devices. These devices most often include first person personal pronouns in plural (16) (17) and personal attribution, very often in connection with epistemic lexical verbs (cf. Table 6).

type of device	occurrences L	occurrences LC
reader involvement devices	71	92
personal attribution	59	14
epistemic lexical verbs	49	12
personal reference to method/ model/experimental conditions	27	0
epistemic modal verbs	21	12
hypotheticals	19	13
offering alternative	5	0
assumption of shared goals	2	0
total	**253**	**143**

Table 6: Reader oriented hedges

It is worth noticing that some of the formal means expressing reader-oriented hedges in linguistics RAs are not present in literary criticism RAs. For instance, there are no means expressing personal reference to method, model or experimental conditions in literary criticism RAs, while in linguistics texts these are quite common. This may as well be explained by the highly subjective character of literary criticism texts where the only 'method' is the subjective perception of the literary work under examination.

(16) And once more the observer in betraying his own romantic insight is left alienated from himself – for he has acted, we might say, against his better nature – and give over to self-pity and guilty isolation [LC E]
(17) If we are to remove some of these difficulties, we must first have a clearer and more restrictive definition of the notion of idiom. [L D]

7. Conclusion and further research questions

The present study focused on detecting possible differences in the use of hedges between linguistics and literary criticism research articles with special regard to the semantic classification and formal realization of hedging expressions. The survey worked with several hypotheses concerning namely the frequency of occurrence of individual semantic categories. The results confirmed that all three semantic categories of hedges identified in the investigation are present in both types of research articles. Quite surprisingly, the expectation that reader oriented hedges would be the most common type in literary criticism RAs was refuted by the results of the analysis. This may be due to the more subjective character of literary criticism research articles; literary critics present their work to their implied readership as just one of a range of possible interpretations, which in a way frees them from the need to use reader oriented hedges on a large scale.

The most common semantic category of hedging expressions in both types of texts analyzed here was the category of writer oriented hedges. This finding suggests that the writers' need to protect themselves transcends disciplines and possibly overrides other motivations that underlie the use of hedging expressions. On the other hand, the remaining two semantic types of hedging expressions are represented in comparatively similar numbers in both fields, which might lead to the conclusion that writers of both linguistics and literary criticism pay equal attention to both the accuracy of the information and the intended recipients of the texts in which the information is presented.

It must be pointed out that possible future research questions emerged during the investigation. For instance, individual writer style appears to be in need of further exploration. Based on the results of this comparative study, it seems that writer style has a considerable impact on the choice of hedging expressions used in both types of RAs with slightly wider variation in the field of literary criticism. Since the present study is not large enough to provide representative data, further extensive investigation is needed to cast light on the role of the individual writer's preferences in connection with hedging.

In conclusion, hedges – even though thoroughly researched in some respects – still remain a language phenomenon that is difficult to address in depth. The comparison of linguistics and literary criticism research articles attempted to provide more insight into the way writers in these disciplines construe meaning in communication with their readers. Nevertheless, there are still issues worth further exploration in order to avoid a simplistic view of this unquestionably complex matter.

Martina Malášková
Faculty of Education
Masaryk University
Brno
malaskom@gmail.com

References

Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam/Philadelphia: John Benjamins.

Austin, J. L. (1962). *How to Do Things with Words*. Harvard: Harvard University Press.

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finnegan (1999). *Longman Grammar of Spoken and Written Language*. London: Longman.

Biber, C. (2006). *University Language*. Amsterdam/Philadelphia: John Benjamins.

Brown, P. & S. C. Levinson (1978). *Politeness. Some Universals in Language Usage.* Cambridge: Cambridge University Press.

Burrough-Boenish, J. (2002). *Hedging*. Dissertation Thesis. NL: Lot Publications.

Crismore, A. & W. J. Vande Kopple (1997). Hedges and Readers: Effects on Attitudes and Learning. In: R. Markkanen & H. Schröder (eds.), 83-114.

Crompton, P. (1997). Hedging in Academic Writing: Some theoretical problems. *English for Specific Purposes* 16 (4), 271-287.

Halliday, M. A. K. (1978). *Language as Social Semiotic. The Social Interpretation of Language and Meaning*. London: Edward Arnold.
Hübler, A. (1983). *Understatements and Hedges in English.* Amsterdam: John Benjamins.
Hyland, K. (1996). Writing Without Conviction? Hedging in Science Research Articles. *Applied Linguistics* 17, 433-454.
Hyland, K. (1998). *Hedging in Scientific Research Articles.* Amsterdam/Philadelphia: John Benjamins.
Hyland, K. (2005). *Metadiscourse.* London: Continuum.
Hyland, K. (2005). Stance and Engagement: A Model of Interaction in Academic Discourse. *Discourse Studies* 7, 173-192.
Hyland, K. (2010). Community and Individuality: Performing Identity in Applied Linguistics. *Written Communication* 27*,* 159-188.
Ifantidou, E. (2005). The semantics and Pragmatics of Metadiscourse. *Journal of Pragmatics* 37 (9), 1325-1353.
Lakoff, G. (1973). Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic* 2, 458-508.
Malášková, M. (2009). Expressing Imprecision in Research Article Introductions: Formal and Semantic Analysis. *Topics in Linguistics* 4, 63-67.
Markkanen, R. & H. Schröder (eds. 1997), *Hedging and Discourse. Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts.* New York: Walter de Gruyter.
Markkanen, R. & H. Schröder (1997). Hedging. A Challenge for Pragmatics and Discourse Analysis. In: R. Markkanen & H. Schröder (eds.), 3-18.
Myers, G. (1989). The Pragmatics of Politeness in Scientific Articles. *Applied Linguistics* 10 (1), 1-35.
Nikula, T. (1997). Interlanguage View on Hedging. In: R. Markkanen & H. Schröder (eds.), 188-207.
Salager-Meyer, F. (1994). Hedges and Textual Communicative Function in Medical English Written Discourse. *English for Specific Purposes* 13, 149-171.
Skelton, J. (1988). The Care and Maintenance of Hedges. *ELT Journal* 41(1)*,* 37-43.
Skelton, J. (1997). How to Tell the Truth in The British Medical Journal: Patterns of Judgement in the 19th and 20th Centuries. In: R. Markkanen & H. Schröder (eds.), 42-63.
Swales, J. (1990). *Genre Analysis. English in Academic and Research Settings.* Cambridge: Cambridge University Press.
Ventola, E. (1997). Modalization: Probability – an Exploration into its Role in Academic Writing. In: A. Duszak (ed.), *Culture and Styles of Academic Discourse*. Berlin/New York: Mouton de Gruyter, 157-179.