Researching Specialized Languages

Edited by Vijay Bhatia,
Purificación Sánchez Hernández
and Pascual Pérez-Paredes

Studies in Corpus Linguistics 47

JOHN BENJAMINS PUBLISHING COMPANY

Researching Specialized Languages

Studies in Corpus Linguistics (SCL)

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

For an overview of all books published in this series, please see *http://benjamins.com/catalog/scl*

Volume 47

Researching Specialized Languages
Edited by Vijay Bhatia, Purificación Sánchez Hernández and Pascual Pérez-Paredes

Researching Specialized Languages

Edited by

Vijay Bhatia
City University of Hong Kong

Purificación Sánchez Hernández
Pascual Pérez-Paredes
University of Murcia

John Benjamins Publishing Company
Amsterdam / Philadelphia

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

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

Library of Congress Cataloging-in-Publication Data

Researching specialized languages / edited by Vijay Bhatia, Purificación Sánchez Hernández, Pascual Pérez-Paredes.
p. cm. (Studies in Corpus Linguistics, ISSN 1388-0373 ; v. 47)
Includes bibliographical references and index.
1. Discourse analysis. 2. Professional employees--Language. 3. Corpora (Linguistics) 4. Oral Communciation. 5. Sublanguage. I. Bhatia, V. K. (Vijay Kumar), 1942- II. Hernández, Purificación Sánchez. III. Pérez-Paredes, Pascual.
P302.R484 2011
410.1'88--dc23 2011019491
ISBN 978 90 272 0352 6 (Hb ; alk. paper)
ISBN 978 90 272 8505 8 (Eb)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

List of contributors

Douglas Biber
Bethany Grey
Department of English
Northern Arizona University
Liberal Arts Building 18 – Room 140
PO Box 6032
Flagstaff, Arizona 86011, USA

Carmen Pérez-Llantada
Departamento de Filología Inglesa,
Facultad de Filosofía y Letras,
Universidad de Zaragoza
Pedro Cerbuna 12,
50009 Zaragoza, Spain

Sara Gesuato
Dip. di Lingue e Letterature
Anglo-Germaniche e Slave
University of Padua,
Via Beldomandi n. 1
35137 Padova, Italy

Mercedes Jaime-Sisó
Departamento de Inglés
Facultad de Veterinaria
University of Zaragoza
C/Miguel Servet, 177
50013 Zaragoza, Spain

Pascual Pérez-Paredes
Purificación Sánchez Hernández
Pilar Aguado
Departamento de Filología Inglesa
Campus de la Merced
Universidad de Murcia
30071 Murcia, Spain

Carmen Sancho
Dpto. de Lingüística Aplicada
ETSI Aeronáuticos
Universidad Politécnica de Madrid
Plaza del Cardenal Cisneros, 3
Ciudad Universitaria,
28040 Madrid., Spain

John Flowerdew
University of Leeds and
City University of Hong Kong
Department of English,
City University of Hong Kong
Room: Y7619
Tat Chee Avenue, Kowloon,
Hong Kong

Shaeda Isani
Université de Grenoble
17 rue des Jardins
38320 Eybens, France

Gillian Diane Lazar
English Language and Learning
Support (ELLS)
Middlesex University
Trent Park Campus
Bramley Road
Oakwood
London N14 4YZ, United Kingdom

Kris Buyse
Eva Saver
An Laffut
Herlinda Vekemans
Leuven Language Institute (LLI),
Fac. Letteren
K.U.Leuven
Blijde-Inkomststraat 21
Postbus 3308
B 3000 Leuven, Belgium

Elena Montiel-Ponsoda
Guadalupe Aguado
Ontology Engineering Group (OEG).
Universidad Politécnica de Madrid
Madrid, Spain

Specialized languages

Corpora, meta-analyses and applications

Vijay Bathia, Purificación Sánchez Hernández
and Pascual Pérez-Paredes
City University of Hong Kong and University of Murcia

Like any other mature discipline, LSP has also had its fair share of conflicts and controversies over the years. In its early days, LSP attracted a variety of comments and evaluative labels, some of them extremely favourable, others not so charitable. When referring to terms and labels such as English for Science and Technology, English for Business, etc., Brumfit (1980: 106) declared that one should not be 'swayed by impressive sounding titles'. Instead, he cautioned that the duty of 'serious teachers is that of investigating new vogue words and determining their value'. On the other side, Swales, as early as 1984, declared that 'ESP has come of age'. He maintained that it had acquired an independent status of its own, not only the status of a respectable teaching profession but also of a more established area of academic study.

Today LSP has moved into a new phase which is both independent, on the one hand, and multi-disciplinary, on the other; independent as a discipline in itself, with its own research agenda, methodology, and applications; and multi-disciplinary, in the sense that it draws upon insights from other related disciplines in discourse and genre analysis, corpus-based research, multi-modal discourse analysis, critical discourse analysis, communication studies, ethnography of communication, socio-cognitive research on teaching and learning of specialised languages, and, at the same time, a number of other academic and professional disciplines that it tends to serve.

This volume is an attempt to continue the tradition of publishing research as a basis for contributing to the development of principled and evidence-based applications to the pedagogic aspects of the teaching of specialised languages. These papers embody the range and methodological robustness which has characterised LSP over the last two decades, representing an insightful spectrum of research interests and featuring some of the major trends of research in the field.

The volume is divided into two sections in terms of the specific methodologies popularly used in current LSP research and practice. Corpus-based research, in recent years, has gained an increasing amount of prominence in LSP analytical

research.The first section therefore encompasses the research of scholars working with linguistic corpora with an emphasis on corpus-based data gathering as a means to gain insight into areas of interest in the field of specialised languages. This set of papers represents an excellent showcase of the insights that can be derived from corpus-based research, which looks at micro-genres such as conference abstracts, research articles across different disciplines, and other academic discourses. The volume also has an almost equal number of papers which focus on meta-analysis of lexico-grammar of specialised discourses. The second part of the volume thus includes research articles which put their focus on meta-analysis as research methodology and/or the application of research to the implementation of applications in the field of specialised languages, which is yet another important aspect of LSP research and practice.

Section One of the book includes the following six papers:

Douglas Biber and Bethany Gray adopt a diachronic, corpus-based methodology to challenge the stereotypical view of professional academic writing as being grammatically complex, packed with elaborated structures, and with meaning relations expressed explicitly. Their historical analysis shows that academic writing has changed dramatically over the past century to strongly prefer these less explicit styles of presentation. The finding, no doubt, is in line with corpus-based findings of Douglas Biber himself, who, in past publications, had already proved that, introspectively, *a priori* ideas about language do not always match language in use.

The authors use a corpus-based analysis of published research articles to describe the discourse styles of academic writing, investigating the extent to which professional academic writing employs grammatical devices for structural elaboration and explicitness. Their corpus of academic research articles, circa 3 million words, was sampled from four general disciplines: science/medicine, education, social science (psychology), and humanities (history). The authors collected texts from three 20-year intervals (1965, 1985, 2005) to enable the description of short-term historical change. They offer a picture of academic writing as not conforming to the stereotypes of 'literate' discourse which has developed a unique style, characterised by a reliance on nominal/phrasal rather than clausal structures.

Carmen Pérez-Llantada analyses a sample of lexico-grammatical patterns (*we*-subject patterns, anticipatory *it*-patterns, inanimate subject patterns and passive constructions) involving varying degrees of inter-subjective stance both in research articles written in English by scholars from an Anglophone-based context and a Spanish context, and in articles written in Spanish by Spanish scholars. The author examines the rhetorical functionality of these patterns, classifying this functionality as dialogically contractive, implying a 'heteroglossic disengagement', and dialogically expansive, implying 'heteroglossic engagement'.

The author built a corpus of biomedical research articles published in English-medium international journals by writers from an Anglophone-based context and another one of articles published by Spanish writers in the same international journals. She also built a comparable corpus of biomedical articles published by Spanish scholars in Spanish journals for inter-linguistic comparison. Pérez-Llantada combines contrastive rhetoric and genre analysis with corpus-based methodology to profile inter-subjective stance in research articles, and in particular in the introduction, method, results and discussion sections.

"Structures, Content and Functions of Calls for Conference Abstracts" is the title of **Sara Gesuato**'s paper on how 100 calls for Conference Abstracts are structured in the academic areas of Biology, Computing, History and Linguistics, her main goal being how authors conceptualise these communicative acts. To carry out her study the author created a corpus of 4 sub-corpora and examined the texts using a bottom-up approach and developing coding categories for the moves.

Gesuato concludes that Conference Abstracts are functionally complex texts, which act as announcements, offers, orders and invitations. The calls for abstract under study comprise 16 move types, whose distribution and frequency are similar across sub-corpora, showing stronger similarities between Biology and Computing and between History and Linguistics. This paper, which makes an important contribution to the study of academic discourse, highlights that Conference Abstracts can help establish and share community standards, practices and viewpoints.

Mercedes Jaime-Sisó's paper tackles the way biomedical scientists deal with their professional reading and writing. For the analysis of discourse as genre, the author follows Swales's and Bhatia's proposals and develops her research in two phases, by interviewing authors in order to verify to what extent scientists' reading behaviour is determined by the use of electronic journals and by building a small corpus of research articles to detect whether online journals offer specific features that may have contributed to the increase in scientists' reading efficiency.

Jaime-Sisó also conducted interviews with ten senior and experienced researchers in the areas of Biochemistry, Pathology, Genetics and Embryology following the procedures established by Berkenkotter and Huckin (1995). The informants were asked about the steps they follow when selecting journal articles online or on paper. The results show that they make the initial selection of research papers online, scanning the table contents, reading the abstract and looking at the tables, figures and legends. If they are interested in the paper they then read the results and the discussion section and skip the introductions and methodology section.

Regarding the analysis carried out on the corpus of the research papers, the author followed two different paths, reading either the title, abstract, results and discussion or the title, summary and adapted texts where available. The author

concludes that the structure of scientific research articles is progressively being adapted to the needs of online users.

The use of adverbials hedged in EAP students' oral performance is the contribution of **Pascual Pérez-Paredes, Purificación Sánchez-Hernández and Pilar Aguado Jiménez** to this volume. The authors focused on the comparison of the use of adverbial hedging between non-native students of English at university level and native British university students.

Following the Louvain International Database of Spoken English Interlanguage, the authors compiled two corpora of spoken English in academic contexts: one of Spanish students of English in their first year at university and a second of British students studying Modern Languages. The adverbial hedges *almost, maybe, sort of* and *kind of* were selected in both corpora and classified into the four types suggested by Biber et al. (1999): imprecision in word choice, approximators and quantifiers, uncertainty, and degree diminishing, to which a fifth category was added by the authors to assign some occurrences that could not be classified into the four existing ones.

The authors present abundant results of their research, which point to a lack of formal instruction on the part of the Spanish speakers of English in the use of hedging devices. In both corpora, hedges seem to be associated with the elicitation of personal information, opinions or views rather than with more objective, impersonal information. The differences between both groups were only significant when the five categories of adverbial hedging were taken as a unique set of data. However, no statistically significant differences were found when considering adverbial hedging individually.

Carmen Sancho-Guinda's paper highlights the need for explicit instruction in visual data transfer within ESP environments, discussing its status as a multi-skill and cross-disciplinary practice and shedding light on the challenges and limitations visual data commentary poses in the classroom setting. The paper illustrates some of the most innovative areas of linguistic research, namely that of multimodal communication and multi-modal corpus analysis. This micro-skill or specialised skill exemplifies the type of new specialised language areas which have attracted the attention of researchers worldwide. Sancho-Guinda discusses the taxonomy advocated by Systemic Functional Grammar where information transfer may be presented as multi-skill and cross-disciplinary practice.

The author reviews the state of the art in linguistic and didactic research, reporting on a preliminary case study conducted at the Madrid School of Aeronautical Engineering. In her research, 57 students of aeronautical engineering performed two tasks orientated towards investigating qualitatively their attempts at visual data transfer and commentary. All the participants were aged between 20 and 22, with an intermediate level of proficiency in the English language and taking a

sixty-hour and English-medium elective Technical English, probably the prototypical ESP language learner across the EU.

The author's findings reveal very little use of metadiscourse and presentational constructions, together with the adoption of compensatory tactics, a preference for common-core and pre-modified superordinate nouns and certain collocational interferences with the speakers' L1. The data collected by Sancho-Guinda corroborates the need for specialised language students to receive explicit training in interpretive data commentary, particularly in dealing with metadiscursive items and presentational structures. The author discusses an inventory of decisions concerning data transfer based on Hallidayan meta-functions in a continuum with an embedded cline for each sub-option.

Section Two of the book contains the following five papers:

John Flowerdew's paper deals with some dichotomies in Genre Analysis for languages for specific purposes. Considering the most important contributions to the field of genre analysis, the author addresses in great detail four of the most relevant theoretical issues related to dichotomies, especially in terms of move structure: Individual genres vs. genre networks; Written vs. spoken genres; Macro vs. micro levels of analysis; and Move structure vs. lexico-grammar.

Flowerdew offers an exhaustive discussion on each of the dichotomies presented, based on the most relevant studies published in the area and offers a wide variety of examples to make his points clear. He aims to contribute towards highlighting areas where further theoretical reflection and empirical enquiry might enhance the theory and at the same time offer even more for application.

Shaeda Isani's contribution adds to the growing literature on English for Legal Purposes. Departing from Konrad Lorenz's imprinting theories and Stephen Krashen's acquisition theories, the author analyses the 3-phase cognitive process behind the creation of the ersatz legal culture and identifies "the Continental Paradox" as the greater knowledge that Spanish, French, German and presumably other students have of American Legal culture than of their own. The singularity of the Continental Paradox lies in the fact that it is not a question of ignorance of the students' national legal system but involves a more complex situation of cultural substitution.

The author advances several reasons to explain the paradox, highlighting the fact that European law institutions and professionals remain anachronistically hermetic and remote as well as highlighting the arcane and obscure character of the legal world and the restricted access of journalistic media to European legal institutions.

The research carried out derives from a cross-disciplinary approach which overlaps law, ethnology and ESP didactic related to teaching ELP to European

students of law and proposes a shift in the conventional approach to promote greater learner awareness of the target legal culture.

Gillian Diane Lazar's paper addresses a crucial issue for those doing research and teaching academic writing in the tertiary sector. According to the author, the narrative genre has not been given much consideration in this educational context. This is even more the case if the narratives are of an oral nature.

Lazar begins her research by describing the difficulties of some British university students engaged in writing academic assignments which were grounded in reflective practice. During this activity, the use of oral narratives was found to be a useful bridge into academic writing for these students. The author considers some features of narrative discourse and the ways in which these can be exploited to develop academic writing skills which incorporate argument. Some of these considerations are that oral narratives are grounded in the tacit knowledge of situated everyday experience and that they can be a way of tapping into the implicit knowledge of genre that students bring with them. In her paper, some examples of practical classroom techniques and materials used with students from different disciplinary backgrounds are suggested.

Kris Buyse, Eva Saver, An Laffut and Herlinda Vekemans present an innovative solution to the problem of handling specialised lexicon in specialised languages: UrgentiAS (Urgentielexicon ArtsenStage), a lexicon for medical students in clinical placements, is an excellent example of specialised language applied to real life needs of training professionals. This lexicon was compiled to meet the needs of Dutch-speaking Belgian medical students attending classes of Medical Spanish/French/English to prepare for their clinical placements in medical areas such as obstetrics and gynaecology, paediatrics, and internal medicine in hospitals in different countries.

This solution is, from the end-user perspective, an online multilingual lexicon of representative terms which can help students acquire the vocabulary needed in their placements abroad, and which allows them to consult it abroad when needed. The authors carefully adopted a needs-analysis methodology in order to deliver a resource which can be instrumental in bridging the gap between lexicographical work and specialised language pedagogy. In this respect, the authors have taken great lengths to document their efforts, thus providing a valuable meta-resource for researchers and language professionals who wish to undertake similar paths.

The authors compiled an open, synchronic and multilingual corpus with material from different monolingual corpora in order to build a multilingual database. It is of interest that they use a very wide textual typology, unusual in specialised corpora which tend to focus on one particular domain, i.e. ER nursing, and on one register, i.e. research articles. The data collected included clinical cases, anamneses and diagnoses written by doctors and addressed to their colleagues

and, in addition, medical language used by and towards patients as well as cultural aspects of medical treatment. The article offers a comparison of test results for Medical Spanish which suggests a significant improvement in language acquisition and skills.

Elena Montiel-Ponsoda and Guadalupe Aguado de Cea's contribution deals with taxonomy and metonymy relations in the context of ontology development. The authors have followed a corpus-based onomasiological approach to discover patterns. Štekauer (2006) argues that this onomasiological approach has been neglected by linguists in the past, who, at least in the field of word formation studies, preferred other non-meaning-driven methods to investigate the genesis and development of lexicons. Naturally, the development and dominance of Chomskyan linguistics did not help at all, but that is another story. Lexical realisations, as argued by the authors, convey a certain relation between concepts and world objects. In Depecker and Roche (2007) we can find the foundations of some of the terms used by the authors.

The topic is of the utmost interest to those doing research in ESP-related areas such as terminology extraction, lexicography, corpus linguistics and natural language processing and specific domains. Not only is the use of ontologies a phenomenon which goes hand in hand with the birth and expansion of the semantic web, but it is also an opportunity for linguists to contribute to domains of knowledge by defining the concepts of that domain and the relations among them. As Montiel-Ponsoda and Aguado de Cea state, the knowledge represented in an ontology captures the consensual knowledge of a community of domain experts. In a recent article Pérez-Paredes (2010) has argued that ontologies can play an important role in providing text-external information (Bhatia 2004, 2008) in terms of both the conventions of a genre and the specific professional practice, the latter seen as text-external aspects of language use. This area will most likely expand in the near future.

In their research, the authors adopted the name Lexico-Syntactic Patterns (Hearst 1992), redefining the term as the linguistic schemas or constructs derived from recurrent expressions in natural language that consist of linguistic and paralinguistic elements that follow a certain syntactic order, and that permit us to extract some conclusions about the meaning they express. The main objective of their research was to create a repository of Lexico-Syntactic Patterns associated with the ontological structures that they express. In identifying Lexico-Syntactic Patterns that correspond to ontological structures or Ontology Design Patterns, those researching specialised languages wishing to go deeper into the application of ontologies can adopt a system that enables the automatic detection of the ontological relation expressed in the sentence introduced by the user.

We hope that this collection of very carefully selected studies will be useful to LSP researchers as well as practitioners.

References

Berkenkotter, C. & Huckin, T.N. 1995. *Genre Knowledge in Disciplinary Communication: Cognition, Culture, Power*. Hillsdale NJ: Lawrence Erlbaum Associates.

Bhatia, V. 2004. *Worlds of Written Discourse: A Genre-based View*. London: Continuum.

Bhatia, V. 2008. Genre analysis, ESP and professional practice. *English for Specific Purposes* 27:161–174.

Biber, D., Johansson, S., Leech, G., Conrad, S & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Brumfit, C.J. 1980. *Problems and Principles in English Teaching*. Oxford: Pergamon.

Depecker, L. & Roche, C. 2007. Between ideas and concepts: Towards ontology. *Langages* 168: 106–114.

Pérez-Paredes, P. 2010. Ontologies and the study of Legal English. In *Professional English in the European Context: The EHEA Challenge*, A. Linde López & R. Crespo Jiménez (eds), 221–243. Berlin: Peter Lang.

Štekauer, P. 2005. Onomasiological approach to word-formation. In *Handbook of Word-Formation*, P. Štekauer & R. Liber (eds), 207–232. Dordrecht: Springer.

Swales, J.M. 1984. 'ESP comes of age? 221 years after "some measurable characteristics of modern scientific prose"'. *UNESCO ALSED-LSP Newsletter* 7/2(19): 9–20.

SECTION ONE

Research based on corpora

The historical shift of scientific academic prose in English towards less explicit styles of expression

Writing without Verbs

Douglas Biber and Bethany Gray
Northern Arizona University

The stereotypical view of professional academic writing is that it is grammatically complex, with elaborated structures, and with meaning relations expressed explicitly. In contrast, spoken registers, especially conversation, are believed to have the opposite characteristics. Our goal in the present paper is to challenge these stereotypes, based on results from large-scale corpus investigations. First, we argue that both conversation and professional academic writing are structurally complex, but their complexities are dramatically different: in some ways, conversation is more structurally elaborated than academic writing (e.g., finite dependent clauses are more common in conversation than in academic writing). In contrast, written academic discourse is actually much more 'compressed' than elaborated, with phrasal (non-clausal) modifiers embedded in noun phrases being the major type of structural complexity found in academic writing. Our historical analysis shows that academic writing has changed dramatically over the past century to prefer these compressed discourse styles. Second, we argue that a consideration of the meaning relations among structural elements illustrates that academic written texts are anything but explicit at the grammatical level. Rather, the 'compressed' discourse style of academic writing is much less explicit in meaning than alternative styles that employ elaborated structures. Again, our historical analysis shows that academic writing has changed dramatically over the past century to strongly prefer these less explicit styles of presentation.

Keywords: Complexity, elaboration, explicitness, academic writing, conversation, noun phrase structures

1. Introduction

Researchers have claimed for decades that academic writing is more structurally elaborated than speech, shown by longer sentences, longer 't-units' (a main clause plus all associated dependent clauses), and a greater use of subordinate clauses (see, e.g., O'Donnell et al. 1967; O'Donnell 1974; Kroll 1977; Chafe 1982; Brown & Yule 1983). In addition, researchers have claimed that academic writing is more explicit than speech. That is, while speech is dependent on a shared situational context, academic writing lacks such shared context between participants and is thus claimed to be 'decontextualized', 'autonomous', or 'explicit', with all assumptions and logical relations being overtly encoded in the text (see, e.g., DeVito 1966; Olson 1977; Kay 1977; Johns 1997: 58–64). Hughes (1996: 33–34) notes both of these characteristics, writing that spoken grammar employs 'simple and short clauses, with little elaborate embedding', in contrast to written discourse, which employs 'longer and more complex clauses with embedded phrases and clauses', 'explicit and varied marking of clause relations', 'explicit presentation of ideas', and 'explicit indication of text organization'.

The belief that academic writing is 'elaborated' and 'explicit' persists up to the present time. For example, Hyland (2002: 50), in a major ESP teacher education textbook, documents the widespread perceptions that academic writing is 'structurally elaborate, complex, abstract and formal', with 'more subordination' and 'more explicit coding of logical relations'. Hyland goes on to describe how there are no clear-cut distinctions between speech and writing, but he does not specifically challenge the characterizations of academic writing as 'structurally elaborate' and 'explicit'.

Our goal in the present paper is to challenge these perceptions, based on the results of large-scale corpus investigations. First, we argue that *both* conversation and professional academic writing are structurally complex – but their complexities are dramatically different. This difference relates primarily to structural elaboration, and here the findings are surprising: in some ways, conversation is more structurally elaborated than academic writing (e.g., subordinate clauses – especially finite dependent clauses – are much more common in conversation than in academic writing). In contrast, written academic discourse is actually much more 'compressed' than elaborated, with phrasal (non-clausal) modifiers embedded in noun phrases being the major type of structural complexity found in academic writing. Our historical analysis shows that academic writing has changed dramatically over the past century to prefer these compressed discourse styles.

Second, we argue that a consideration of the meaning relations among structural elements illustrates that academic written texts are anything but explicit at the grammatical level. Rather, the 'compressed' discourse style of academic writing is much less explicit in meaning than alternative styles that employ elaborated

structures. This generalization holds for a wide range of different grammatical devices that are especially common in academic writing (but rare in speech). Again, our historical analysis shows that academic writing has changed dramatically over the past century to strongly prefer these less explicit styles of presentation.

The following sections present the results of large-scale corpus analyses that document these patterns of use. Section 2 introduces the corpora and linguistic features used for the analyses. Then, the analyses themselves are discussed in two main sections: Section 3 surveys the historical findings for the use of features associated with structural elaboration versus compression, and Section 4 discusses the lack of explicitness resulting from the use of these compression devices in academic writing. In conclusion, we briefly discuss functional motivations for these historical changes and implications for teaching academic reading (and writing) at the university level.

2. Corpus and grammatical features used for the analysis

We employ corpus-based analysis to describe the typical discourse styles of academic writing, investigating the extent to which professional academic writing (as represented by published research articles) employs grammatical devices for structural elaboration and explicitness. Previous corpus-based studies have documented the different complexities of spoken and written registers. For example, multi-dimensional studies of register variation (e.g., Biber 1988, 1992, 2006) have shown repeatedly that certain dependent clause types (e.g., *because*-clauses and WH-clauses) are more strongly associated with speech than writing. The *Longman Grammar of Spoken and Written English* (Biber et al. 1999) provides more detailed descriptions of the grammatical features that are common in conversation versus those that are common in academic writing.

Building on this previous research, the present study focuses on the grammatical devices in English that are associated with structural elaboration and explicitness. Many of the descriptions below contrast the patterns of use in professional academic writing to those in conversation, based on analysis of a large corpus of texts for each of these two registers.

We constructed a corpus of academic research articles (see Table 1), sampled from four general disciplines: science/medicine, education, social science (psychology), and humanities (history). We collected texts from three 20-year intervals (1965, 1985, 2005) to enable the description of short-term historical change. However, for the purposes of the present study, we consider these as a single group (429 texts, c. 3 million words), contrasted with conversation.

The conversation subcorpus is taken from the *Longman Spoken and Written Corpus* (see Biber et al. 1999: 24–35). The subcorpus includes 723 text files and

Table 1. Corpus of Research Articles*

Year	Academic Discipline	Number of Research Articles	Totals for Subcorpus
1965	Science/Medicine	70	158 articles
	Education	30	ca. 923,000 words
	Psychology	28	
	History	30	
1985	Science/Medicine	44	132 articles
	Education	30	ca. 810,000 words
	Psychology	28	
	History	30	
2005	Science/Medicine	52	139 articles
	Education	27	ca. 1,206,000 words
	Psychology	30	
	History	30	

*Research journals sampled for each discipline: Science/Medicine (*Journal of Cell Biology, Biometrics, American Journal of Medicine, Journal of Animal Ecology, Journal of Physiology)*; Education (*American Educational Research Journal, Journal of Educational Measurement*); Psychology (*American Journal of Psychology, Developmental Psychology)*; History (*Journal of Contemporary History, Journal of the History of Ideas*).

around 4.2 million words of American English conversation. These are conversations collected by participants who agreed to carry tape recorders for a 2-week period. The corpus thus represents one of the largest collections of natural face-to-face conversations available.

The corpora were grammatically annotated ('tagged') using software developed for the *Longman Grammar of Spoken and Written English* and earlier corpus studies of register variation (e.g., Biber 1995). Then, more specialized computer programs were developed for detailed linguistic analyses of specific types of structural elaboration.

Table 2 lists the types of dependent clauses that we considered for our analysis of structural elaboration. These dependent clauses can serve three major syntactic functions: complement clauses, which usually function as the direct object of a verb; adverbial clauses, which modify the main verb; and post-nominal relative clauses, which modify a head noun.

We also considered grammatical devices that result in a 'compressed' rather than 'elaborated' discourse style, illustrated in Table 3. These are all phrases rather than dependent clauses, used to modify a head noun. Attributive adjectives and pre-modifying nouns occur before the head noun ('pre-modifiers'), while prepositional phrases occur after the head noun ('post-modifiers').

Most of these features could be identified accurately using automatic computer programs. However, prepositional phrases required hand coding to determine

Table 2. Selected grammatical features associated with structural elaboration

Grammatical feature	Examples
Finite complement clauses	*I wonder how he is today.* *I thought that was just too funny.*
Non-finite complement clauses	*I'd like to get one of these notebooks.* *Do you want to elaborate on that more?*
Finite adverbial clauses	*She won't narc on me, because she prides herself on being a gangster.* *You can have it if you want.*
Finite relative clauses	*A method that would satisfy the above conditions...* *a repressor substance which prevents the initiation...*
Non-finite relative clauses	*The results shown in Tables IV and V add to the picture...* *The factors contributing to the natural destruction of microbes...*

Table 3. Selected grammatical features associated with structural compression

Grammatical feature	Examples
Attributive adjective (adjective as noun pre-modifier)	*a large number, unusual circumstances*
Noun as noun pre-modifier	*human actions, membrane structure*
Prepositional phrase as noun post-modifier	*the scores for male and female target students in the class* *the mechanism for penetration of protein through the ovariole wall*

when the phrase was functioning as a noun modifier versus adverbial. This analysis was based on a sub-sample of tokens (every 4th occurrence) from a sub-sample of the corpus (41 academic research articles, and 48 conversations). Appositive noun phrases were also coded by hand (to distinguish them from other noun phrases with different syntactic functions). The counts for all linguistic features were converted to a 'normed' rate of occurrence (per 1,000 words) for each text (see Biber, Conrad & Reppen 1998: 263–264).

3. Structural elaboration and compression in academic writing

The first belief to be challenged is the claim that academic writing is structurally elaborated. Researchers have usually focused on dependent clauses (or subordinate clauses) as the primary measure of structural elaboration. Surprisingly, there is extensive clausal embedding in conversation. In particular, complement clauses (also called 'nominal clauses') are very common, especially *that*-clauses and

WH-clauses. Complement clauses normally fill a direct object slot, making it possible for a relatively short utterance to have multiple levels of embedding. For example, the following short utterance has two embedded complement clauses:

You know [you could get [what you wanted]]

Unlike adverbial clauses and relative clauses, complement clauses are not optional structures; rather, they take the place of a required noun phrase. In conversation, the complement clause usually occurs with a transitive verb (e.g., *think, know,* or *want*): the complement clause substitutes for the noun phrase as the direct object of the verb. As a result, these structures can contain multiple levels of structural embedding. For example, the following relatively short sentence from conversation has four embedded complement clauses, each occurring as the object of the preceding main verb:

But I don't think [we would want [to have it [sound like [it's coming from us]]]].

In contrast, the typical discourse style of modern-day academic prose employs surprisingly little clausal embedding. If we employ traditional measures, such as the number of embedded dependent clauses, academic prose would not be characterized as elaborated. This is because the structural elaboration of academic writing is realized as phrases without verbs. For example, consider the following sentence from a Biology research article:

The knowledge of tissue distribution of each novel molecular species is the first step toward the understanding of its possible function.

This sentence consists of only a single main clause, with the main verb *is*. There are no dependent clauses in this sentence. The sentence is relatively long because there are multiple prepositional phrases:

of tissue distribution
of each novel molecular species
toward the understanding
of its possible function

In addition, many of the noun phrases include extra nouns or adjectives as premodifiers before the head noun:

tissue distribution
novel molecular species
possible function

In their main clause syntax, sentences from academic writing tend to be very simple. Thus, consider the following sentence from a Psychology research article:

This may indeed be ***part*** *[**of** the reason [**for** the statistical link [**between** schizophrenia and membership [**in** the lower socioeconomic classes]]]].*

The clausal syntactic structure of this sentence is extremely simple, with only one main verb phrase:

X may be Y (*This may be part*)

All of the elaboration here results from prepositional phrases added on to noun phrases. Thus, unlike conversation, academic writing does *not* often employ dependent clauses for structural elaboration. Rather, we find a more 'compressed' style, employing embedded phrases rather than fuller dependent clauses.

Our corpus investigations show that the patterns illustrated above represent strong general differences between academic writing and conversation. Figure 1 shows that complement clauses and adverbial clauses are much more frequent in conversation than in academic writing. These differences are strongest for finite clauses (e.g., *that*-clauses and *WH*-clauses functioning as complement clauses; *because*-clauses and *if*-clauses functioning as adverbial clauses). However, the same general pattern holds for non-finite complement clauses (*to*-clauses and *ing*-clauses). In contrast, relative clauses are more frequent in academic writing than in conversation (especially non-finite relative clauses, such as *the concept of society proposed here*). Overall, Figure 1 shows that there are around twice as many dependent clauses in conversation as in academic writing.

Instead of dependent clauses, academic writing relies heavily on non-clausal phrases to add information. Figure 2 shows that most of these phrases occur embedded in noun phrases. Many of these structures are adjectives modifying a head noun (e.g., *theoretical orientation*) or nouns pre-modifying a head noun (e.g., *system perspective*). But the most striking difference from conversation is for the use of prepositional phrases as noun post-modifiers. Many of these are *of*-phrases (e.g., *an interpretation of the general form of mitochondria*), but other prepositions are also commonly used for this function (e.g., *the complex relations between three components; understanding rational approach to politics*). Prepositional phrases used as adverbials (e.g., *From the systems perspective, these stages are marked by...*) are also more common in academic writing than in conversation, but the difference is much less strong.

In summary, the stereotype that writing is more elaborated than speech is not supported by corpus evidence. In fact, using traditional measures of elaboration – considering the use of dependent clauses – we would conclude that the opposite was the case: that conversation is more elaborated than academic writing.

However, that conclusion would be misleading. That is, the apparent lack of elaboration in academic writing is in part an artifact of inadequate measures, rather than an accurate characterization of academic writing. That is, elaboration has

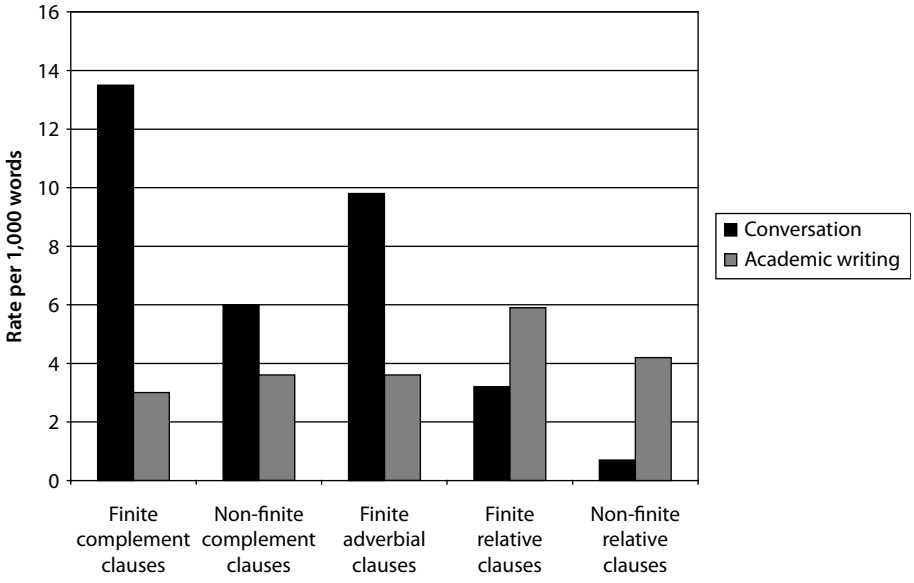


Figure 1. Common dependent clause types

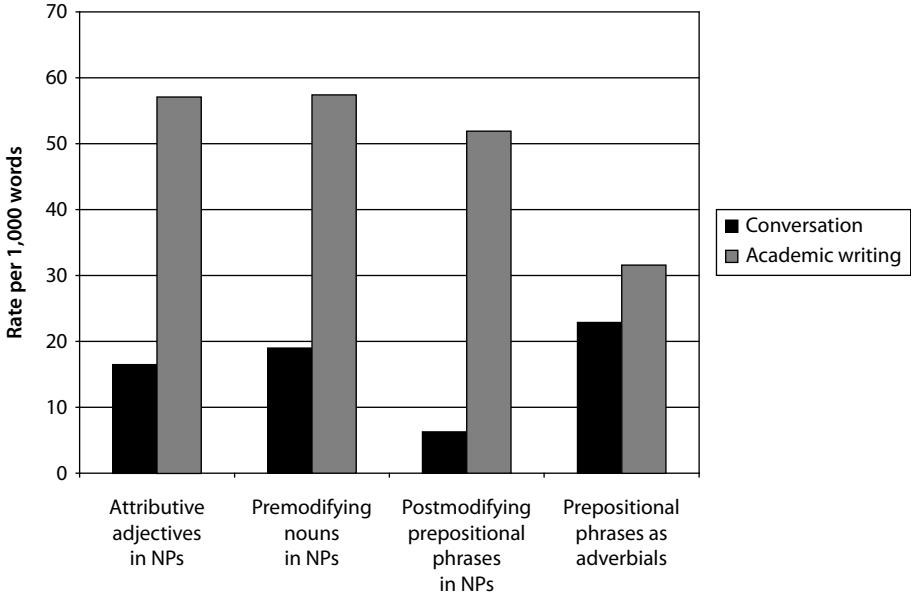


Figure 2. Common dependent phrase types

normally been analyzed by considering the extent to which dependent clauses are used in a text. By that measure, we would conclude that academic writing is actually less elaborated than conversation. However, that measure misses the most important structural characteristic of academic written discourse: the reliance on phrasal rather than clausal elaboration. Most sentences in academic prose are elaborated in the sense that they have optional *phrasal* modifiers, especially nominal pre-modifiers (adjectives or nouns) and post-modifiers (prepositional phrases).

Interestingly, academic research writing has not always been unelaborated and implicit in the expression of meaning relations. In fact, this is a recent innovation. Corpus-based studies, which allow us to track these historical changes, have shown that the shift to the compressed, inexplicit style of discourse described above is largely a 20th century phenomenon (see Atkinson 1992, 1996, 1999; Biber & Finegan 1997, 2001; Biber & Clark 2002; Biber & Conrad 2009, Chapter 6).

Figures 3 and 4 document some of these historical changes, based on analysis of science/medical texts in the ARCHER Corpus and astronomy texts in the *A Coruña* Corpus. Figure 3 shows that nouns as pre-modifiers were generally rare in academic writing through the 18th and 19th centuries. They began to increase in the early 20th century, and have increased strongly in use over the course of that century. Similarly, Figure 4 shows that prepositional phrases as noun modifiers have increased strongly in use over the past century. *Of*-phrases were already prevalent in the 18th century, and their frequency has remained relatively constant. However, the use of other prepositions as noun modifiers has increased strongly across these centuries.

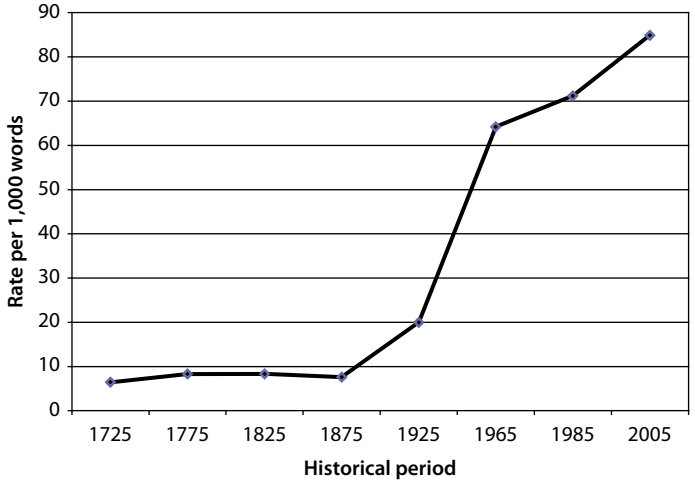


Figure 3. Historical change in the use of nouns as nominal pre-modifers in academic prose

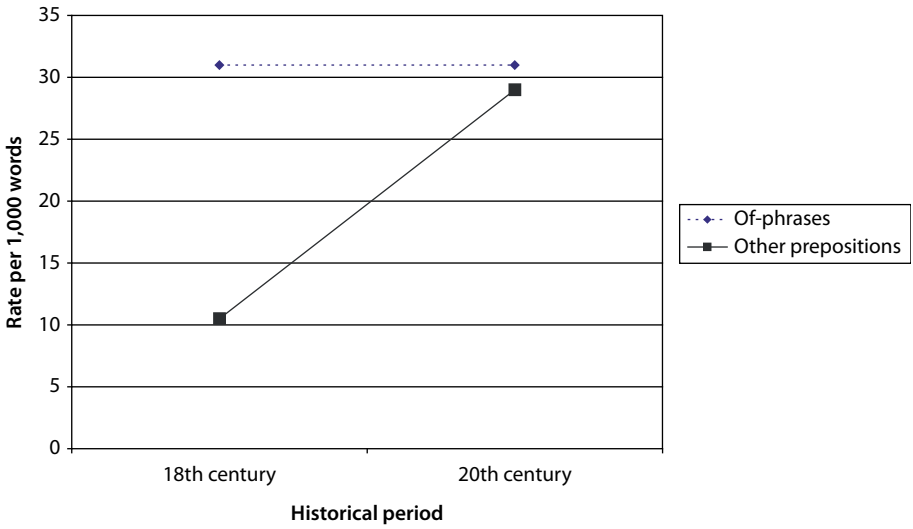


Figure 4. Historical change in the use of prepositional phrases as post-nominal modifers in academic prose (medical)

There are good reasons why compressed, phrasal expressions are preferred over elaborated clausal expressions in academic writing: they are more economical; they allow for faster, more efficient reading; and they are equally comprehensible to the expert reader. However, it is often difficult to specify the exact meaning relationship between the head noun and the modifiers. With compressed, phrasal modifiers, those meaning relationships are implicit, while clausal modifiers require a fuller explicit specification of the meanings. As a result, this discourse style is at odds with another major stereotype: that academic writing is maximally explicit in meaning.

4. Explicit and implicit meaning in academic writing

The second stereotype to be challenged is the belief that academic writing is maximally explicit, in contrast to conversation, which relies heavily on implicit meaning. Similar to the stereotypes about elaboration, this stereotype about explicitness fails to fully capture the discourse patterns of academic writing.

On the one hand, persons, objects and places are referred to explicitly in academic prose (e.g., with full noun phrases), in contrast to the practice in speech to use less explicit, situation-dependent referring expression (such as pronouns or deictic adverbials). Thus, in conversation there are frequent pronouns referring situationally to the conversational participants (I, you) or other

people and objects (e.g., he, she, it, that, those). In contrast, inexplicit referring expressions are avoided in academic prose. The reader does not share time and place with the author of academic prose; if the author failed to use explicit referring expressions, readers would have no basis for understanding the intended reference. Thus, in this sense academic prose may be considered more explicit than conversation.

However, if we consider the meaning relations among grammatical constituents, academic writing is anything but explicit. In fact, it has changed historically to strongly prefer grammatical styles that are dramatically less explicit than in earlier periods. In particular, all of the forms of phrasal (as opposed to clausal) modification described in the preceding section result in a loss of explicitness.

For example, consider the meaning relationship between a head noun and a pre-modifying noun, as in the noun phrase *heart disease*. There is no grammatical clue to help the reader know what the relation is between the two nouns. However, a wide range of meanings is possible, and the use of clausal elaboration can specify the exact meaning relationship; for example:

lung cancer	*cancer located in the lung*
tin can	*a can made from tin*
family photos	*photos taken of a family*
prison officials	*officials who work in a prison*
computer keyboard	*a keyboard used with a computer*
retail outlet	*an outlet which sells retail merchandise*
union assets	*assets that belong to a union*

As a result of this inexplicitness, the exact same pre-modifier can have a completely different relation to the head noun, as in:

pressure hose	*a hose able to withstand pressure*
pressure ratio	*a ratio that measures pressure*

As shown in Figure 2 above, pre-modifying nouns are extremely frequent in academic writing but generally rare in conversation; Figure 3 further shows that academic writing has changed radically in the last 100 years to strongly prefer this structural device that is much less explicit in meaning than alternative expressions.

As noted above, all phrasal forms of modification preferred in academic writing are inexplicit in similar ways. The most important of these is prepositional phrases used as nominal modifiers, which are considerably less explicit than alternative clausal modifiers. For example, consider the three prepositional phrases beginning with *for* in the following sentence:

Another reason to use Ohio as a surrogate [for the country as a whole] is that the data base [for hazardous waste generation and flow] [for the State] is fairly good.

In their grammatical form, these three phrases are identical, but they express three different meaning relationships. Clausal modifiers would make those meaning relationships explicit:

a surrogate for the country as a whole →
a surrogate that represents the country as a whole

the data base for hazardous waste generation →
the data base that documents hazardous waste generation

the data base ... for the State →
the data base that the State uses

Other meaning interpretations are possible here, at least for the non-expert reader. For example, *the data base for hazardous waste generation* could mean 'the data base used to generate hazardous waste'. Similarly, *hazardous waste generation and flow for the State* could mean 'the hazardous waste that is generated and flows in the State'. The grammatical forms used in this academic sentence provide no clues to the intended meaning. Rather, meaning is implicit and must be inferred by the reader. Here again, as shown in Figure 4, academic writing has changed dramatically over the past 100 years to prefer discourse styles with these inexplicit structures.

5. Summary and pedagogical implications

Present-day academic writing is one of the most distinctive registers in English. In its grammatical characteristics, it is dramatically different from all spoken registers and most other written registers. It does occasionally use 'spoken' features (like first person pronouns), but the basic grammatical structure of discourse is nominal/phrasal rather than clausal, and thus it is unlike spoken discourse (even speech in informational/academic settings; see, e.g., Biber 2006).

Academic writing is certainly complex, elaborated, and explicit, but it does not conform to our stereotypes about these characteristics:

- complex: yes, but not in its use of the traditional measures – see 'elaborated'
- elaborated:
  - in the use of embedded phrases, especially in noun phrases: yes
  - in the use of clausal subordination – the traditional measure of elaboration: no

- explicit:
  - in specifying the identity of referents: yes
  - in the expression of logical relations among elements in the text: no

Thus, on the one hand, academic writing is dramatically different from speech, but on the other hand, it does not conform to the stereotypes of 'literate' discourse. Rather, it has developed a unique style, characterized especially by the reliance on nominal/phrasal rather than clausal structures.

One important motivation for the development of this discourse style is that it is highly efficient for academic professionals: because prose is so compact, an expert reader can quickly scan through a research article and extract the essential information. The lack of explicitness causes few (if any) problems, because the expert reader anticipates the expected readings that will occur in this context. A compact (and inexplicit) style actually helps expert readers keep up with the increasing volume of information produced by scientific researchers.

However, these compact, inexplicit discourse styles are difficult for novice students. This is because students lack the specialist knowledge that would allow them to readily infer the expected meaning of compact, inexplicit constructions. Thus, there are clear implications of these findings for the teaching of academic reading and writing: whether we like these discourse styles or not, it is a certainty that our students will encounter professional writing with these characteristics. Their academic success will depend on their ability to extract the intended meaning from these texts (reading) and to eventually produce such texts (in writing).

The main goal of the present study has been descriptive, to challenge the widely-held perceptions that academic writing is elaborated and explicit, and to instead document the ways in which it is compressed and inexplicit. In order to succeed in an academic profession, students must learn to efficiently read and to eventually write this style of discourse. Future research is required to determine how that process can best be facilitated through EAP instruction.

References

Atkinson, D. 1992. The evolution of medical research writing from 1735 to 1985: The case of the Edinburgh Medical *Journal. Applied Linguistics* 13: 337–374.

Atkinson, D. 1996. The Philosophical Transactions of the Royal Society of London, 1675–1975: A sociohistorical discourse analysis. *Language in Society* 25: 333–371.

Atkinson, D. 1999. *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London, 1675–1975.* Mahwah NJ: Lawrence Erlbaum Associates.

Biber, D. 1988. *Variation across Speech and Writing.* Cambridge: CUP.

Biber, D. 1992. On the complexity of discourse complexity: A multidimensional analysis. *Discourse Processes* 15: 133–163.

Biber, D. 1995. *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge: CUP.
Biber, D. 2006. *University Language: A Corpus-based Study of Spoken and Written Registers* [Studies in Corpus Linguistics 23]. Amsterdam: John Benjamins.
Biber, D. & Clark, V. 2002. Historical shifts in modification patterns with complex noun phrase structures. In *English Historical Syntax and Morphology* [Current Issues in Linguistic Theory 223], T. Fanego, J. Pérez-Guerra & M.J. López-Couso (eds), 43–66. Amsterdam: John Benjamins.
Biber, D. & Conrad, S. 2009. *Register, Genre and Style.* Cambridge: CUP.
Biber, D., Conrad, S. & Reppen, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: CUP.
Biber, D. & Finegan, E. 1997. Diachronic relations among speech-based and written registers in English. In *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*, T. Nevalainen & L. Kahlas-Tarkka (eds), 253–275. Helsinki: Societe Neophilologique.
Biber, D. & Finegan, E. 2001. Diachronic relations among speech-based and written registers in English. In *Variation in English: Multi-dimensional Studies*, S. Conrad & D. Biber (eds), 66–83. London: Longman.
Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English.* London: Longman.
Brown, G. & Yule, G. 1983. *Discourse Analysis.* Cambridge: CUP.
Chafe, W. 1982. Integration and involvement in speaking, writing, and oral literature. In *Spoken and Written Language: Exploring Orality and Literacy*, D. Tannen (ed.), 35–54. Norwood NJ: Ablex.
DeVito, J.A. 1966. Psychogrammatical factors in oral and written discourse by skilled communicators. *Speech Monographs* 33: 73–76.
Hughes, R. 1996. *English in Speech and Writing: Investigating Language and Literature.* London: Routledge.
Hyland, K. 2002. *Teaching and Researching Writing.* Harlow: Longman.
Johns, A. 1997. *Text, Role, and Context: Developing Academic Literacies.* Cambridge: CUP.
Kay, P. 1977. Language evolution and speech style. In *Sociocultural Dimensions of Language Change*, B. Blount & M. Sanches (eds), 21–33. New York NY: Academic Press.
Kroll, B. 1977. Ways communicators encode propositions in spoken and written English: A look at subordination and coordination. In *Discourse across Time and Space* [SCOPIL no. 5], E.O. Keenan & T. Bennett (eds), 69–108. Los Angeles CA: University of Southern California.
O'Donnell, R.C. 1974. Syntactic differences between speech and writing. *American Speech* 49: 102–110.
O'Donnell, R.C., Griffin, W.J. & Norris, R.C. 1967. A transformational analysis of oral and written grammatical structures in the language of children in grades three, five, and seven. *Journal of Educational Research* 61: 36–39.
Olson, D.R. 1977. From utterance to text: The bias of language in speech and writing. *Harvard Educational Review* 47: 257–281.

Heteroglossic (dis)engagement and the construal of the ideal readership

Dialogic spaces in academic texts

Carmen Pérez-Llantada Auría
University of Zaragoza

Appropriating White's (2003) distinction of 'heteroglossic engagement/ disengagement' modes of expression, this paper analyses a sample of grammatical patterns and their rhetorical functionality for the expression of intersubjective stance in research articles written in English by scholars from two cultural contexts (an Anglophone-based context and a Spanish context) and articles written in Spanish by Spanish scholars. Results show that the scholars from the Anglophone context prefer patterns that construct a heteroglossic disengagement mode, suggesting the construal of their readership as sharing similar values, beliefs and attitudes. The texts in Spanish display patterns that create a heteroglossic engagement mode, indicating that writers conceive their readership as having potentially dissenting views. Remarkably, the texts written in English by the Spanish scholars strategically combine a heteroglossic disengagement mode with a heteroglossic engagement mode, hence creating a hybrid dialogic space for writer/reader interaction.

Keywords: Contrastive rhetoric, stance, dialogism, heteroglossia, academic writing

1. Introduction

The transmission of scientific knowledge was rightly described by Lyotard (*The Postmodern Condition*, 1984) as a language game in which "the sender is supposed to be able to provide proof of what s/he says, and on the other hand s/he is supposed to be able to refute any opposing or contradictory statements concerning the same referent", hence creating "a horizon of consensus to be brought to the debate between partners" (in Sarup 1993: 135–136). Echoing Lyotard's post-structuralist reflections, this dialectic language game has largely been postulated by sociorhetoricians, applied linguists and genre theorists in EAP contexts

(cf. Berkenkotter & Huckin 1995; Hyland 2001; Swales 2004), who have conceived academic writing as rhetorically conventionalised – "a universe of discourse which sets up imperatives for the language forms and structures" (Wilkinson 1991: xvii). These linguistic forms and rhetorical configurations do not simply help writers transmit new knowledge but also construct subtle social relationships with their audiences for reasons of acceptance and recognition within a given disciplinary community.

One area of research that has paid considerable attention to the dialogic and interpersonal nature of academic writing practices 'is the contrastive rhetoric (CR) field'. CR has identified specific linguistic and rhetorical preferences across cultural contexts and languages and has brought to the fore not only the view that academic prose is interpersonal and socially-oriented but also that the text-linguistic realisations of such social orientation entail culture-specific variability (Connor 2004; Breivega et al. 2002; Dahl 2004; Fløttum 2005; Vold 2006; Connor et al. 2008). As Duszak (2005: 17) puts it, "varying clines of commitment and detachability" surface when we compare the way scholars from Anglophone and non-Anglophone contexts project their stances onto their texts to construct knowledge, establish credibility and gain the acceptance of their audiences.

To our knowledge, however, very limited research has focused on the rhetorical functionality of grammatical expressions of intersubjective stance and on the way these expressions involve varying clines of commitment/detachability according to the writers' construal of their ideal readership – supporting White's (2003: 259) contention that "language construes social roles and relationships". This paper analyses a sample of grammatical patterns involving varying degrees of intersubjective stance in research articles written in English by scholars from two cultural contexts (an Anglophone-based context and a Spanish context) and in articles written in Spanish by Spanish scholars. Relying on White (2003: 258–259), the paper examines the rhetorical functionality of these patterns – i.e. the degree of authorial attachment to or detachment from propositional contents that they convey – and classifies this functionality into two types: dialogically contractive and dialogically expansive. Grammatical patterns with a dialogically contractive functions help writers build up a solidarity relationship with the audience – implying a 'heteroglossic disengagement' mode of expression and the construal of the readership as one sharing similar views, beliefs and attitudes. Grammatical patterns with dialogically expansive function establish deferential rather than solidarity relations with readers and instantiate a 'heteroglossic engagement' mode that opens up a space for alternative views on the part of the audience. In agreement with previous CR research (e.g. Vold 2006; Dahl 2004; Fløttum 2005; Vázquez Orta & Giner 2008, 2009), we hypothesised that the texts published by the scholars from a non-Anglophone context were going to follow the standard (Anglophone-based)

academic grammatical patterns. Yet, we further hypothesised that, through the use of phraseological elements accompanying those patterns at a textual level, they were going to display a different intersubjective stance from that used by the scholars from the Anglophone-based context. These differing stances would instantiate two different culture-specific textual responses to different audience construals. As for the cross-linguistic comparison of the texts written by Spanish scholars in English for international journals and in Spanish for Spain-based journals, we initially hypothesised that, regardless of the different target audiences of these articles, the actual rendering of the texts – i.e. their textual development – was going to be similar in the two subsets of texts due to the scholars' ingrained, culture-related, intellectual style of formal academic writing. Therefore, the discursive hybridity of the texts written in English by the Spanish academics would be consistent with previously reported hybridisation features in the academic writing practices of other groups of non-Anglophone academics – cf. Mauranen's study (1993) on Finnish academics, Clyne (1996) on German scholars, Gianonni (2008) on Italian academics, and Carciu (2009) and Pérez-Llantada (2010b) on Spanish scholars.

2. Corpus and methodology

For the analysis, we selected 24 biomedical research articles (RA henceforth) published in English-medium international journals by writers from an Anglophone-based context (ENG corpus) and 24 articles published by Spanish writers in the same international journals (SPENG corpus). We also selected 24 biomedical articles published by Spanish scholars in Spanish journals (SP corpus) for interlinguistic comparison with SPENG. The 72 texts belonged to the biomedical component of SERAC (the *Spanish English Research Article Corpus*).[1] To better identify the rhetorical functionality of the grammatical patterns of intersubjective stance, we segmented the texts into rhetorical sections following the methodological approach of Biber et al. (2007) (cf. Table 1)

Table 1. Number of words per rhetorical section across the three subsets of texts

	Introductions	Methods	Results	Discussions	Totals
ENG	11,407	22,719	35,811	24,259	94,196
SPENG	9,065	25,909	26,239	23,352	84,565
SP	9,379	13,038	15,091	22,197	59,705
Totals	29,831	61,666	77,141	69,808	238,466

1. For a full description on the selection criteria of SERAC see Pérez-Llantada (2008).

To explore the two modes of heteroglossic (dis)engagement we selected a sample of patterns (*we*-subject patterns, anticipatory *it*-patterns, inanimate subject patterns and passive constructions) considered by the aforementioned CR literature as linguistic resources that writers use to express varying degrees of authorial commitment/detachment. We classified these patterns using White's (2003) distinction of dialogically contractive and dialogically expansive resources. *We*-subjects and anticipatory *it*-patterns were conceived of as dialogically contractive resources by means of which writers convey commitment to propositions. *We*-subjects (e.g. *[W]e report a reduction*) would indicate committed authorial positioning. Anticipatory *it*-patterns (e.g. *It is important to remember*; *It is likely that*) would represent dialogically contractive resources with an implicit emphatic function at a discourse level. On the other hand, inanimate subjects and passive constructions would be instances of dialogically expansive resources through which writers express detachment from propositions. Inanimate subjects (e.g. *Molecular analysis of chimerism showed*; *The lack of consistency supports*) would implicitly articulate authors' lack of involvement towards the new knowledge claims. Similarly, passive constructions (e.g. *Nelarabine has consistently been reported to be*) were considered to convey complete authorial detachment as they embed no explicit agent indicators. The commitment/detachment scale would then be represented as follows.

We first tracked the occurrence of these patterns across RA sections using Wordsmith Tools (Scott 1999). Since both ENG and SPENG and SPENG and SP respectively were homogeneous and comparable in all their contextual values, we also applied a likelihood-ratio test to identify the overuse/underuse of the patterns in the ENG corpus relative to the SPENG corpus and in the SPENG corpus relative to the SP corpus. In the comparison between ENG and SPENG, ENG was considered the normative corpus. In the comparison between SPENG and SP, SPENG was taken as the normative corpus (cf. Rayson & Garside 2000). Drawing inspiration from Mauranen's (1993) study of text-linguistic differences in English and Finnish academic texts, we also deemed it necessary to enquire into the textual developments and the phraseology accompanying the selected grammatical patterns. As described below, this text-linguistic analysis helped delimit more accurately the specific modes of expression in the two cultural contexts and in the two languages.

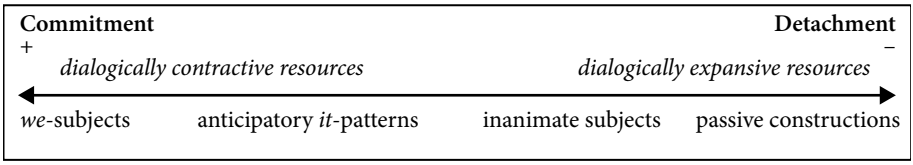


Figure 2. The commitment/detachment scale

3. Results

The comparison of the three corpora indicated an overall uniformity in the use of patterns across RA sections. Passive constructions were shown to prevail throughout rhetorical sections – being particularly prominent in Methods and Results sections but also scoring the highest comparative frequencies and the highest expected frequencies in Introductions and Discussion sections (cf. Tables 3 and 4). Overall results further imply that the section-specific use of patterns in the three corpora is highly standardised, possibly due to medical writers' adherence to the *Uniform requirements for manuscripts submitted to biomedical journals* (http://www.icmje.org/). The Introductions of the three groups of texts only showed variation in the use of *we*-pronoun patterns and passive constructions. In Methods, variation was only found in the use of passive constructions. In Results and Discussions, only inanimate subject patterns showed statistical variation. The remaining patterns indicated no significant differences across the three sets of texts.

The results of the likelihood-ratio test (Table 4 on next page) showed that in Introductions the overuse of *we*-pronoun patterns in ENG was significant at the level of $p < 0.001$ compared to SPENG and that the overuse of passive constructions in SPENG was highly significant ($p < 0.0001$) compared to SP. In Methods sections, the overuse of passives in ENG was significant at $p < 0.0001$ when compared to SPENG and significant in SPENG when compared to SP ($p < 0.05$). In Results sections, the overuse of inanimate subjects in SPENG was significant at $p < 0.01$ when compared to ENG, and highly significant ($p < 0.0001$) when compared to SP. Variation across the two sets of texts written by Spanish scholars was

Table 3. Comparative percentages of use of grammatical patterns

		***we*-subject patterns**	**anticipatory *it*-patterns**	**inanimate subject patterns**	**passive patterns**
Introductions	ENG	19.02%	3.93%	7.87%	69.18%
	SPENG	13.53%	5.26%	7.52%	73.68%
	SP	9.56%	12.12%	19.12%	59.56%
Methods	ENG	0.82%	0.00%	0.10%	99.08%
	SPENG	2.49%	0.00%	0.32%	97.19%
	SP	2.94%	0.74%	0.00%	96.32%
Results	ENG	13.10%	1.75%	4.80%	80.35%
	SPENG	18.77%	3.90%	7.96%	69.37%
	SP	16.01%	3.92%	0.33%	79.74%
Discussions	ENG	15.83%	9.91%	22.26%	52.00%
	SPENG	14.68%	9.67%	23.83%	51.81%
	SP	18.65%	9.92%	7.74%	63.69%

Table 4. Expected frequencies and log-likelihood (LL) values

		expected frequencies		LL	expected frequencies		LL
		ENG	SPENG	value	SPENG	SP	value
Introductions	*we*-pron.	52.38	31.88	20.58*	24.08	26.44	10.48
	antic. *it*-patt.	14.49	11.51	0.96	14.74	15.26	0.07
	inanim. subj.	24.52	19.48	0.02	22.61	23.39	0.59
	passives	226.78	180.22	2.47	136.14	140.86	53.22**
Methods	*we*-pron.	14.48	16.52	5.73	23.28	11.72	0.01
	antic. *it*-patt.	0.93	1.07	0.01	2.66	1.34	2.88
	inanim. subj.	1.87	2.13	0.80	7.98	4.02	8.65
	passives	872.73	995.27	19.88**	859.49	432.51	5.51
Results	*we*-pron.	136.78	100.22	10.46	110.47	63.53	5.45
	antic. *it*-patt.	23.66	17.34	7.40	24.12	13.88	0.41**
	inanim. subj.	54.25	39.75	7.53*	34.28	19.72	40.22**
	passives	1078.08	789.92	25.84	725.65	417.35	123.22*
Discussions	*we*-pron.	89.68	86.32	0.01	91.77	87.23	0.10
	antic. *it*-patt.	57.58	55.42	0.01	54.34	51.66	0.10
	inanim. subj.	135.53	130.47	0.85	90.74	86.26	53.79**
	passives	305.21	293.79	0.26	318.37	302.63	2.17

*significant at the level of $p < 0.001$
**significant at the level of $p < 0.0001$

also found in this section, being the overuse of passives in SPENG significant ($p < 0.001$). In Discussions, only inanimate subject patterns showed statistically significant differences, with an overuse of this particular pattern in SPENG ($p < 0.0001$) compared to SP.

Below, we describe the rhetorical functionality of the patterns across RA sections and relate it to the specific textual developments and the co-occurring phraseology of these patterns in each set of texts.

In Introductions, writers in the two cultural contexts choose dialogically expansive passive constructions to indicate a research gap, raise a question or set the grounds for occupying the research space. They then opt for dialogically contractive *we*-pronouns to express commitment when they refer to the rationale of the study at the end of the section. Correlating with contrast markers (*however*, *until now*), clausal negations and restrictive adverbs (*not*, *only*) and activity verbs implicitly indicating writers' involvement in the research (*compare*, *analyse*), the functionality of *we*-patterns is targeted at foregrounding research agents when the purpose of the paper is announced:

(1) To date, however, the role of cyclic nucleotides in HbF induction, which has been well established in K562 cells, **has** not **been** rigorously **explored** in primary erythroid cultures. In this report, **we compare** the actions of HU, SB, and AZA on HbF induction [...] (ENG23)

(2) Until now, these compounds **have been tested** mostly in cell lines, with only two reports describing HA14–1 as inducing apoptosis in a small number of primary acute myeloid leukemia samples [26, 36]. Here**, we analyze** the effects of the Bcl–2 inhibitors HA14–1 [...] (SPENG8)

The overuse of *we*-subject patterns in the ENG Introductions creates a collegial writer/reader relationship that implicitly closes down the space for alternative views on the part of the readership. ENG writers promote their research through *we*-subject references and justify the need for the current research by explicitly referring to their own previous findings. These references are accompanied by possibility modals (*could*, *can*) and intensifying adverbs (*greatly*) and by this means the textual development suggests that these writers regard their audience as potentially consenting:

(3) **We** recently **developed** a procedure to produce hematopoietic cells by coculturing hESCs with fetal human liver clone B (FH-B-hTERT) cells17 [...]. **We demonstrated** that **we** could reproducibly **obtain** about 150,000 CD34 cells [...]. **We** also **demonstrated** that increasing the time of coculture of the hESCs from 1 to 3 weeks was associated with a partial globin switch [...]. **We report** here that the hematopoietic cells produced by coculture on FHB- hTERT can be greatly expanded in vitro into fully mature primitive erythroid cells. (ENG22)

Resembling the ENG authors, the SPENG writers express commitment to claims by means of *we*-pronoun patterns that co-occur with other dialogically contractive stance markers (stance adjectives and adverbs), hence instantiating a heteroglossic disengagement mode of expression similar to that of the ENG Introductions. It could be argued at this point that this shared mode of expression could result from the fact that in international journals "originality (especially in theory) tends to be highly prized, competition tends to be fierce, and academic promotionalism and boosterism are strong" (Swales 2004: 226). However, the textual developments of the SPENG Introductions are recurrently constructed upon tightly-knit cause-effect relationships. As shown in the extract below, writers' stance is mitigated through probability modals (*may*) and epistemic lexical verbs (*speculate*). Even if this particular mode of expression is primarily heteroglossically disengaged, the use of *we*-pronouns allows writers to open up some space for dialogism. The

extract also illustrates the overuse of passive constructions in the SPENG Introductions compared to SP, with impersonal expressions (e.g. *It can thus be speculated that*) that convey authorial detachment from claims:

(4) **We have** recently **shown** that the methylation of cytosine nucleotides in ALL cells may be the most important way of inactivating cancer-related genes in this disease. In fact, this epigenetic event can help to inactivate tumor-suppressive apoptotic or growth-arresting responses [...] Interestingly, the Dkk–3 gene which negatively modulates Wnt7A signaling is frequently silenced by methylation in ALL15, 23. **It can** thus **be speculated that** the functional loss of Wnt antagonists can contribute to activation of the Wnt pathway in ALL, and may play a role in the pathogenesis and prognosis of the disease; however, this hypothesis has never been previously explored. In this study, **we demonstrate** that silencing [...] (SPENG1)

The SP Introductions show slightly higher comparative frequencies of inanimate subjects to the detriment of *we*-pronoun patterns. Combined with passive constructions, inanimate subject patterns function as dialogically expansive resources. Co-occurring with cause/effect markers (*debido a*/[*due to*]), restrictive dependent clauses (*aunque*/[*although*]), possibility modals and semimodals (*puede deberse*/[*can be due to*], *parece que*/[*it seems that*])), inanimate patterns help writers restrict the space for dialogism considerably when they state the purpose of the study at the end of the Introduction (e.g. *El objetivo de este estudio es*/[*The aim of this study is*]).

(5) Aunque el hemocultivo **se considera** la base para el diagnóstico de la bacteriemia, el valor de los hemocultivos en pacientes en que se sospecha bacteriemia es cuestionable, debido a que la verdadera sensibilidad del test no puede **ser determinada**, el rendimiento global es relativo y el valor predictivo para patógenos verdaderos es muy bajo. Además, los resultados de los hemocultivos pueden no tener ningún impacto en el tratamiento o, incluso, llevar a un tratamiento inapropiado. A pesar de estas limitaciones, parece que el uso de hemocultivos puede llegar a ser excesivo en los pacientes adultos hospitalizados. **El objetivo de este estudio es** determinar la utilización y el rendimiento de los hemocultivos [...]. (SP20)

In Methods sections, scholars in the two cultural contexts and in the two languages contextualise the study and provide details for replication and validation. Quantitative results showed an all-encompassing use of passive constructions (with comparative frequencies higher than 95% in the three corpora) (cf. Figure 3 above). This homogeneous authorial detachment, more marked in the two sets of texts published in international journals, substantiates that the central communicative purpose of this RA section is to provide an objective reporting of research

procedures with the aim of facilitating replication of the experiment and, ultimately, verification of the study. The overuse of passives in ENG with respect to SPENG and in SPENG with respect to SP might be attributed to reasons of precision and accuracy when writers report the protocol for conducting the study. The two sets of English texts minutely follow the conventions pertained to this section and writers in the two cultural contexts are very accurate in details on data collection, methodological procedures, statistical support, reference standards, statistical analysis, criteria used for the evaluation and interpretation of results, as well as the calculations performed and the statistical tests done to conduct the analysis:

(6) Northern blotting **was carried out** using 25 lg of total RNA (McKie et al, 2000). RNA from the tissues **was denatured** in formaldehyde-containing buffer and electrophoresed on 1% agarose/formaldehyde gels. RNA from the gels **was transferred** on to a nylon membrane. (ENG9)

(7) Briefly, whole BM samples (106 cells in 100 ll/tube) **were incubated** for 10 min in the dark (room temperature; RT) with saturating amounts of the following combinations of fluorescein isothiocyanate (FITC)-/ phycoerythrin (PE)-conjugated monoclonal antibodies. (SPENG22)

(8) La determinación de lactato **se hizo** [*was carried out]* de manera inmediata cuando existió una sospecha diagnóstica clínica y posteriormente se recogieron otras muestras al tercer y séptimo día, así como a la tercera semana y en controles posteriores de seguimiento. (SP9)

In the Results sections of the three corpora, except for inanimate subject patterns, the remaining patterns showed similar uses and rhetorical functions. Mainly through passive constructions and, to a lesser extent, through *we*-subjects and anticipatory *it*-patterns writers in the three corpora summarise the most significant results, generalise and interpret, exemplify, state limitations, evaluate findings and make overt claims. ENG and SPENG writers opt for *we*-pronouns co-occurring with perception verbs and attach similar rhetorical functionality to this pattern – to summarise the main research procedures and outcomes. As exemplified below, dialogically contractive stance adverbials (*particularly*, *of interest*) initially make these writers' research attractive and important. However, as the text evolves, *we*-pronoun patterns do no longer perform self-promotional goals (as happened in Introductions) but strictly convey reporting of past research procedures. Instead of *we*-subject patterns, inanimate subjects (e.g. *data*, *results*, *analysis*, *findings*, etc.) act as dialogically expansive resources that convey detachability and restrict the scope of the writers' claims in international publications:

(9) The combination of forskolin and HU proved to be particularly toxic to the cells so as an alternative **we utilized** rolipram, a PDE4 inhibitor, to stabilize cAMP in the cultures with HU as an induction agent. **Results were** consistent with the forskolin data in that rolipram augmented the actions of HU (Fig. 7B) [...]. Together **these data support** a role for involvement of cAMP in the action of HbF induction agents in the CD34 system. (ENG23)

(10) Of interest, 4 1 integrin and CXCR4 costimulation did not increase MMP–9 secretion above the levels induced by either stimulus, but resembled the effect of CXCL12 alone (Figure 5A). To confirm this, **we studied** the signaling pathways involved in MMP–9 upregulation by costimulation via 4 1 and CXCL12. [...] All together, **these results established** that CXCL12 and 4 1 enhanced MMP–9 production via different signaling pathways. (SPENG2)

While a subtle shift from dialogically contractive to dialogically expansive resources becomes noticeable at the very end of the ENG Results sections, the textual developments of the SPENG Results appear to create a manifest open space for writers/readers dialogism. Assessment of research outcomes is abundant but, whenever it occurs, the SPENG writers tend to keep a balance between detachment from propositions (through inanimate subject patterns) and commitment to them (through anticipatory *it*-patterns). The former patterns combine with dialogically expansive probability modals (*might*) and the latter co-occur with dialogically contractive stance lexis (*noteworthy*, *highlight*), ultimately creating a mixed disengagement/engagement mode of expression:

(11) **We observed** that edelfosine was more potent than perifosine in its ability to recruit apoptotic molecules into membrane rafts (Figure 4A). **This differential protein** translocating activity of the ALPs might **explain** the higher antitumor activity of edelfosine when compared to perifosine in MM cells (Figure 1, A and C). **It is noteworthy to** highlight the almost complete translocation of Fas/CD95, DR5, FADD and Bid (SPENG4)

While Bazerman's (1990: 78) observation that "the writing-up of results was more of an after-the-fact reconstruction to make one's results seem attractive, important, and true to the consumer of knowledge" applies to ENG Results sections and, to a lesser extent, to SPENG Results, it does not seem to be valid for the SP corpus. In SP Results, inanimate subjects hardly occur. This section is characterised by a detached positioning in which authorial evaluation is kept to a minimum (f.i., in the example below, evaluation is embedded in a sentential relative clause). In addition, the discourse is mainly sustained upon cause-effect and probability markers (*posible efecto*/

[*possible effect*], *sería escasa*/[*would be scarce*], *probablemente*/[*probably*]) whenever writers refer to specific observations. This limited authorial commitment builds up a consistent heteroglossic engagement mode of expression throughout this section:

(12) El área bajo la curva SROC (fig. 3B) fue <de 0,80 (IC 95% 0,77–0,82), y no cambió sensiblemente cuando **se construyeron** las curvas SROC asimétricas (ABC 0,90 [IC 95% 0,76–0,82]), lo que indica que la influencia de un posible efecto umbral sería escasa. Por el contrario, los resultados expresados en términos de sensibilidad y especificidad mostraron una clara heterogeneidad (p < 0,001 en ambos casos), probablemente en relación con el mencionado efecto umbral. (SP16)

In Discussion sections, the three groups of writers re-state specific research procedures and findings through agentless passive patterns, and at times opt for anticipatory-*it* patterns to provide readers with interpretations of results. Even so, specific uses and functions of dialogically contractive *we*-pronouns and, above all, dialogically expansive inanimate subjects in ENG and SPENG – barely used in the SP texts – suggest intercultural and interlinguistic differences when writers interpret research outcomes and highlight their significance.

The ENG writers initially turn to *we*-subject patterns and inanimate subject patterns that co-occur with stance markers (*new*, *useful*, *possibilities*, *considerable advantage*, *important*, etc.) and possibility modals (*can*, *could*, etc.) to convey, with self-promotional overtones, a positive evaluation of research outcomes. This co-occurring phraseology confers assertiveness to the authorial voices and invites readers to accept the writers' points of view. Writers' commitment to propositions suggests that they construe their audience as potentially consenting towards the authors' interpretations of the new findings:

(13) **We have developed** a new method to produce relatively large numbers of human erythroid cells in liquid culture from undifferentiated hESCs. [...]. **This new method** should **prove** useful as an experimental system to study many questions related to human erythropoiesis such as the mechanism of globin switching, hemoglobinopathies, iron transport, and enucleation to name just a few possibilities. One considerable advantage of producing RBCs by directed differentiation of hESCs is that the cells produced are genetically homogenous and can be genetically manipulated either directly in the hESCs or using lentiviruses during the culture. Another important aspect of this work is that it is a first step toward the production of RBCs for transfusion. (ENG 22)

At the end of the section, the ENG scholars use interpersonal resources to establish solidarity positions towards their readership. Using a heteroglossic disengagement

mode of expression, writers convey collegiality towards their readership through *we*-pronouns that recurrently combine with positively evaluative lexis (*efficiently*, *potent*, *promising*, *important*), stance expressions (*in fact*, *of note*) and possibility modals. Overall, authorial commitment prevails when writers re-state their main findings. This mode of expression is also recurrent when writers summarise the main research outcomes at the end of the Discussion. Writers opt for inanimate subject patterns (e.g. [*I*]*n summary, this study shows* [...]) and the discourse is devoid of any pragmatic qualification. Though stated in an assertive way (e.g. *will require further study*), limitations help authors indirectly to confer relevance to the study in question. Through emphatic anticipatory *it*-patterns, the ENG writers show commitment to propositions when suggesting future research lines that may overcome limitations. As illustrated below, these patterns close down the space for dialogism, suggesting that the writers' ideal readership is construed as sharing similar value positionings:

(14) In the present study, **we show that** large numbers of CACs and EPCs can be efficiently collected from mobilized peripheral blood by leukapheresis. The leukapheresed CACs retain their in vivo angiogenic potential and, in fact, on a per-cell basis are similar to freshly harvested CACs. Finally, **we show that** CACs but not EPCs can be cyropreserved and retain functional activity after thawing. Of note, **we** used a DMSO-based cryopreservation protocol; [...] In summary, **this study shows** that AMD3100 is a potent and rapid mobilizer of angiogenic cells, providing a promising strategy to stimulate therapeutic angiogenesis following acute vascular injury. **This study** also **shows** the feasibility of obtaining and storing large numbers of angiogenic cells from the blood of cytokine-treated individuals. Whether mobilized peripheral blood mononuclear cells are superior to bone marrow cells in stimulating angiogenesis will require further study. Ultimately, it will be important to definitively identify the cell type(s) that mediate angiogenesis in vivo and develop techniques to isolate them. (ENG10)

While quantitative results indicated that the SPENG Discussions shared the linguistic preferences of the ENG writers, differing rhetorical functionalities of the preferred grammatical patterns and their accompanying phraseology can be easily tracked in the SPENG corpus. At the beginning of this section, *we*-pronoun patterns exclusively perform a summarising function (i.e. re-statement of the overall issue being researched). This was not the case with the ENG writers, who also used *we*-pronoun patterns for showing commitment to propositions when research results are interpreted (see, f.i., the extract above). Very timid evaluative references stand in stark contrast with ENG's recurrent use of evaluative markers – e.g. *In this study, **we***

defined *a novel surface antigen* (SPENG14); *In this report,* ***we have studied*** *the role and regulation of MMP–9 in B-CLL cells.* ***We show*** *for the first time* (SPENG2).

At the end of the Discussion, this committed stance merges with detached authorial positionings. At this point, the SPENG writers recurrently opt for dialogically expansive inanimate subject patterns to introduce explicit assessment of research results. Both solidarity and deferential positions are maintained between writers and readers. Inanimate subject patterns co-occur with both dialogically contractive (evaluative lexis) and expansive resources (probability modals, concessive dependent clauses and clausal negations indicating limitations of the research presented in the paper). The modelled reader is thus construed as both consenting to and dissenting from the authors' mode of thinking:

(15) **The present report** therefore **identifies** novel physiological mechanisms contributing to B-CLL progression and **provides** evidence for a key role of MMP–9 in this malignancy. MMP–9 degrades extracellular matrix and mediates B-CLL migration and invasion, as **we show** here, but may also release matrix-bound angiogenic factors, such as VEGF and bFGF, and contribute to angiogenesis.11 Although still not proved in B-CLL, these angiogenic factors in turn up-regulate MMP–9 in other cell systems, 11 thus establishing a feedback mechanism for tumor expansion. (SPENG2)

Even if the SPENG Discussions are initially built upon a dialogically contractive argumentation – in this respect, resembling the heteroglossic disengagement mode of expression of the ENG Discussions – their mode of expression fluctuates between disengaged and engaged positionings at the very end of the section. As shown in the example below, the SPENG writers interpret results through dialogically expansive resources such as inanimate subject and epistemic markers (*suggest*, *could*) that counteract the rhetorical effects of dialogically contractive overt authorial evaluation (*significant*, *indeed*, *advantage*). The concluding statement exemplifies how the writers finally open up the space for the audience's alternative views (*remains to be proved*). This mixed heteroglossic disengagement/engagement mode suggests that the SPENG writers explicitly promote their research but, at the same time, acknowledge greater vulnerability to criticism than the ENG writers do at the end of this section:

(16) In conclusion, **our findings suggest** that intensity of conditioning regimen and chronic GVHD have a significant effect on T-cell reconstitution in the early and late periods after transplantation. [...] Whether such a rapid immune recovery could indeed contribute to the apparent therapeutic advantage of RIC-SCT over MA-SCT in patients with a high risk of transplant-related mortality remains to be proved. (SPENG10)

The SPENG authors' construal of the readership as having potentially dissenting views may easily explain the writers' subtle combination of inanimate subject patterns, at times collocating with oblique *we*-forms (e.g. *our findings*) and detached passive constructions. Recurrent modal markers (*could*, *would*), reason-result markers (*as a consequence*, *therefore*) and other hedging expressions (*virtually, further studies are required*) create, as exemplified below, a heteroglossic engagement mode of expression:

(17) Overall, **these results** obtained through the analysis of purified MCs could **contribute** to an explanation of previously observed discrepancies in the literature. As a consequence**, our findings** would also **indicate** that treatment with imatinib, which **has been shown** to be ineffective in vitro and in vivo against MCs carrying the D816V c-kit mutation, would therefore not be indicated in virtually all patients with SM, except most WDSM. However, further **studies are required** before definitive conclusions can **be drawn** in this regard. (SPENG3)

Although statistical data indicated differences in the use of inanimate subjects between SPENG and SP, with an overuse of these patterns in the former texts, the use of dialogically contractive grammatical patterns in the SPENG Discussions resembles the SP Discussions as regards the textual development and accompanying phraseology. At the beginning of the SP Discussions, *we*-pronoun patterns collocate with epistemic lexical verbs (*creemos*/[*we believe*]) and often co-occur with concessive dependent clauses and clausal negations (*a pesar de que*/[*although*], *no se realizó*/[*was not conducted*]). Also, possibility modals combine with dialogically expansive passive constructions (*los resultados pueden ser representativos*/[*results can be representative*]). The discourse lacks authorial evaluation and interpretation:

(18) A pesar del reducido tamaño muestral de este estudio, que no se realizó de forma prospectiva y que se produjeron pérdidas para el seguimiento, **creemos que** los **resultados** obtenidos pueden **ser** representativos de las modificaciones tanto clínicas como epidemiológicas ocurridas durante una década en las mujeres infectadas por el VIH y sus hijos. (SP13)

Such mode of expression recurs till the very end of the SP Discussions. Argumentation at the end of this section is highly impersonal, typically characterised by the use of passive constructions in the statement of limitations, thereby showing writers' manifest responses to potentially dissenting views on the part of the audience:

(19) En conclusión, a pesar de que nuestra tasa de extracción de hemocultivos y su rendimiento **se ha situado** en niveles ligeramente mejores en comparación con los estudios realizados en pacientes críticos, todavía parece

> que puede haber margen para mejorar nuestros resultados mediante la implementación de un protocolo de extracción capaz de identificar aquellas situaciones clínicas que más se asocian con hemocultivos positivos (SP17)

4. Discussion and implications

The comparative analysis across the two cultural contexts and the two languages has shown to illustrate the distinction between White's modes of heteroglossic (dis)engagement – one relying on dialogically contractive resources that close down the space for dialogism, and the other sustained upon dialogically expansive mechanisms that open up the dialogic space and invite readers to take different views from those of the writers. The analysis has further shown cross-cultural and cross-linguistic variability according to the rhetorical functionality of the selected sample of grammatical patterns and its relationship with the co-occurring phraseology and the specific textual developments of the texts.

The ENG and SPENG writers have been shown to prefer similar grammatical patterns to express clines of commitment/detachability in Introductions. In seeking objectivity, these writers establish the research territory mainly through passive constructions. However, the ENG scholars shift to self-promotional *we*-patterns at the end of this section and construct a dialogic space in which promotionalism stands as the writers' primary rhetorical goal. Through this pattern, they become more visible writers and favour a heteroglossic disengagement mode of expression that conveys commitment towards contents and establishes a collegial relationship with the readership. As for the textual development of the ENG Introductions, the text-linguistic analysis has further shown that *we*-pronoun patterns correlate with other dialogically contractive resources (contrast markers and clausal negations creating the research gap, activity verbs implying writers' agency, and intensifying and stance adverbs expressing overt commitment to claims) that strengthen the persuasive authorial voices. By presupposing a potentially consenting audience, the ENG writers tinge their Introductions sections with an assertive tone in order to gain the readers' credibility and acceptance of the validity of the new knowledge claims.

In the SPENG Introductions, *we*-patterns have shown to co-occur with some of the dialogically contractive resources that the ENG scholars use (restrictive adverbs creating the research gap, activity verbs implying writers' agency and stance adverbs expressing overt commitment) but also with dialogically expansive resources (probability modals and epistemic lexical verbs) that mitigate the self-promotional intentions attached to this RA section.

By comparison, the SP Introductions display non-committed voices through anticipatory *it*-patterns and, above all, through inanimate subjects in the statement of purpose that combine with other dialogically expansive resources such as cause/effect markers, restrictive dependent clauses, modals and semimodals. In sum, the dialogic space in the ENG Introductions is constructed upon a heteroglossic disengagement mode of expression; the SPENG Introductions represent a mixed heteroglossic disengagement/engagement mode and the SP Introductions instantiate a heteroglossic engagement mode of expression that opens up a space for alternative positionings on the part of the readership. Opposing the SP writers' intersubjective positionings, relatively similar reader modellings in ENG and SPENG may confirm that writing Introductions in international journal publications involves a highly complex rhetorical exercise with two intersecting communicative goals (cf. Askehave & Swales 2001): the transmission of new knowledge, and the establishment of a suitable writer-reader interaction for convincing the audience of the validity of this new knowledge. The three dialogic spaces are represented in Figure 5.

Concerning Methods sections, the standardised use of the grammatical patterns of intersubjective stance has revealed similar writing procedures for this

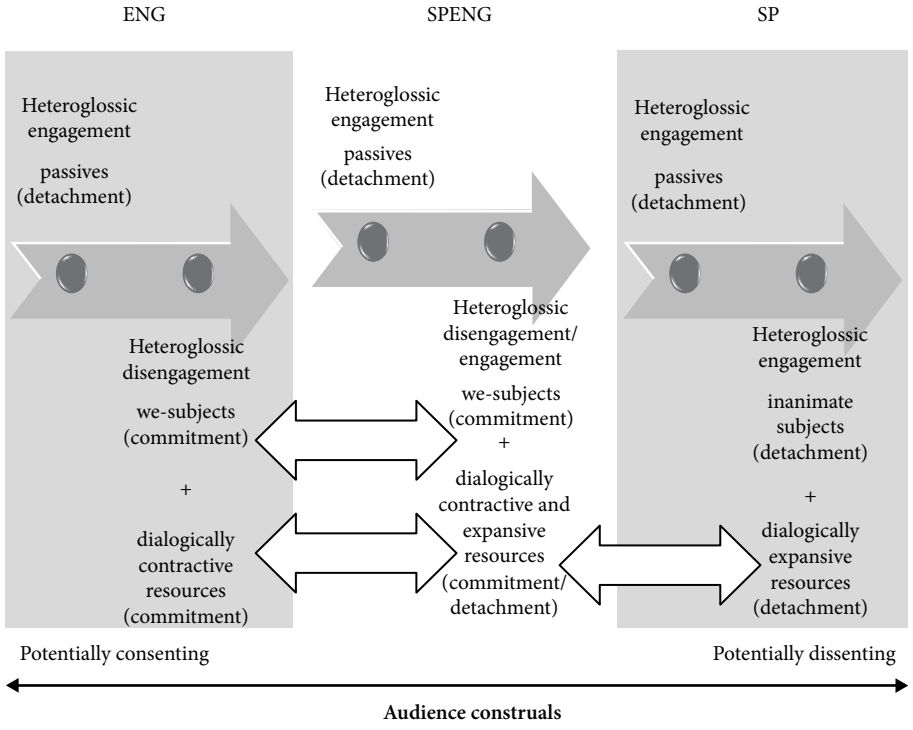


Figure 5. Evolving dialogic spaces in Introductions

particular RA section, confirming that this section involves a routinised writing process in the biomedical field across the cultural contexts and languages analysed. The overuse of passives in ENG and SPENG compared to SP may be taken as an indication of the existence of very strict requirements for an impersonal reporting of methodological protocols in international publications.

The rhetorical functionality of the selected sample of patterns in Results and Discussion sections also suggests the construction of different dialogic spaces. Except for some initial passive reporting of research outcomes, the overall mode of expression in the ENG Results and Discussions tends to be heteroglossically disengaged. Accompanied by other dialogically contractive resources such as positively evaluative markers and possibility modals, *we*-subject patterns help ENG writers show committed stances when they state research findings, comment on them in Results sections and interpret them and suggest future research lines in the Discussion sections. Despite some sparse expressions of detachment through inanimate subject patterns for interpreting results at the end of the ENG Discussions, this mode of expression indicates the construction of the audience as potentially consenting towards the writers' views. The SPENG Discussions, though, have shown a gradual shift from disengaged to less disengaged (or more engaged) authorial stances. In Results sections these writers use inanimate subjects – statistically significant compared to that of the ENG and SP authors – to convey authorial detachment and construct a heteroglossic engagement mode of expression in their English-medium publications. A similar dialogic space has been constructed in the SPENG Discussions. Initially, *we*-subject patterns convey authorial commitment when writers re-state research findings but, noticeably, the presence of evaluative lexis accompanying this self-promotional pattern is relatively scarce. As regards their rhetorical functionality, *we*-pronoun patterns in SPENG Discussions have been shown to be used for reporting past research procedures but not for conveying any interpretation or overt assessment of research outcomes – as happened in the ENG Discussions. At the end of the section, the SPENG writers approach the heteroglossic engagement mode of expression of the SP writers. Inanimate subject patterns co-occur with both dialogically contractive stance adverbs but, above all, with dialogically expansive resources (probability modals, concessive dependent clauses, and clausal negations indicating limitations of the research presented in the paper). This dialogic space suggests that the modelled reader is construed as both possibly dissenting and possibly consenting to the authors' mode of thinking – hence the writers' mixed solidarity-deferential positionings.

The cross-linguistic comparison of SPENG and SP reveals a mixed mode of expression in the SPENG texts. The text-linguistic analysis has shown that the SPENG writers' use of *we*-pronouns merely involves reporting and summarising of past research procedures. When the textual voices engage in persuasion to consolidate

research findings at the end of the Discussion, inanimate subjects conveying authorial detachment are mitigated by dialogically expansive devices (modal verbs). The Spanish writers writing in English and in Spanish hardly interpret research outcomes through *we*-pronoun patterns and, if they do so, they acknowledge limitations of their study using a heteroglossic engagement mode of expression which is built upon abundant cause-effect markers and restrictive clauses. This indicates that the SP authors and, to some extent, the SPENG ones, construe the communicative context as heteroglossically diverse, particularly towards the end of the Discussion. As happened in Introductions, but indeed more markedly in Results and Discussions, the initial heteroglossic mode of expression of the SPENG writers ultimately becomes less disengaged and a much more engaged mode subtly opens up the space for dialogism – ultimately mirroring the overall mode of expression in the SP Discussions. These diverse dialogic spaces in Discussion sections are represented in Figure 6.

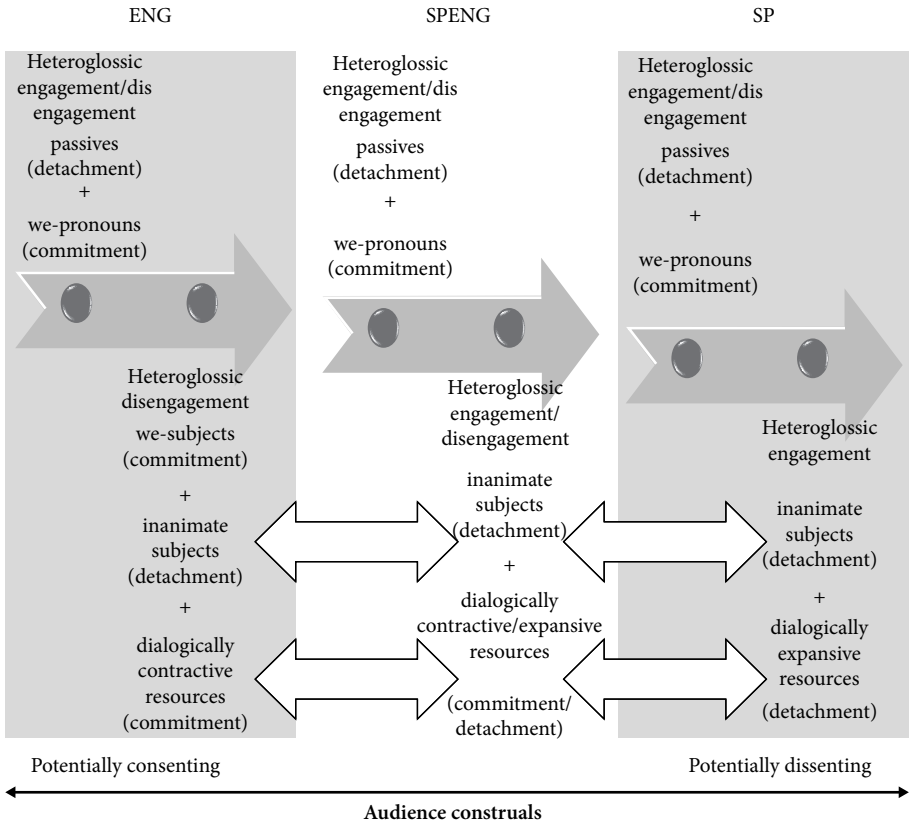


Figure 6. Evolving dialogic spaces in Results and Discussions

The hybrid mode of expression in the SPENG corpus – consistent across RA sections – may instantiate how, despite the benefits of English as a global lingua franca for academic and research communication, the non-native scholars transfer some of their local L1 rhetorical traditions to their texts in English (cf. also Mauranen 1993; Clyne 1996; Bennett 2007; Gianonni 2008; Carciu 2009; Pérez-Llantada 2010a, 2010b). As suggested by Kerans (2002), Ferguson (2007) and Mauranen et al. (2010), whether or not the hybrid features of the SPENG texts involves a gradual loss of the non-native scholars' specific cultural traits in favour of the use of English for international scholarly publishing should be further examined in future research.

Pedagogically, the results of this study should be taken into consideration when helping both native and non-native English scholars understand how the use of language for transmitting scientific knowledge is always socially-situated. In this respect, it would be advisable to sensitise them as regards these different modes of heteroglossic (dis)engagement with audiences and with the range of dialogically contractive/expansive patterns and resources that writers can choose to express varying clines of involved/detached intersubjective stances across RA sections. Recognition of addressee features and the kinds of social relationships that can be established according to specific audience construals may facilitate the transmission of disciplinary knowledge as well as the construction of preferred dialogic spaces for writer-reader interaction across cultural contexts and languages.

Acknowledgements

I am most grateful to Prof. John M. Swales for his guidance on this paper during my research spell at the English Language Institute (University of Michigan) in 2008. The stay was fully funded by a Research Mobility Grant that I received from the University of Zaragoza (Spain). The paper is also a contribution to project FFI2009–09792 (Spanish Ministry of Education) and project UZ2008-HUM–06 (University of Zaragoza).

References

Askehave, I. & Swales, J. 2001. Genre identification and communicative purpose: A problem and a possible solution. *Applied Linguistics* 22(2): 195–212.

Bazerman, C. 1990. Discourse analysis and social construction. *Annual Review of Applied Linguistics* 11: 77–83.

Bennett, K. 2007. Epistemicide! The tale of a predatory discourse. *The Translator* 13(2): 151–169.

Berkenkotter, C. & Huckin, T.N. 1995. *Genre Knowledge in Disciplinary Communities: Cognition, Culture, Power*. Hillsdale NJ: Lawrence Erlbaum Associates.

Biber, D., Connor, U. & Upton, T.A. (eds). 2007. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure* [Studies in Corpus Linguistics 28]. Amsterdam: John Benjamins.

Breivega, K.R., Dahl, T. & Fløttum, K. 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. *Applied Linguistics* 12(2): 218–239.

Carciu, O.M. 2009. An intercultural study of first-person plural references in biomedical writing. *Ibérica* 18: 71–92.

Clyne, M. 1996. *Inter-cultural Communication at Work: Cultural Values in Discourse*. Cambridge: CUP.

Connor, U. 2004. Intercultural rhetoric research: Beyond texts. *Journal of English for Academic Purposes* 3: 291–304.

Connor, U., Nagelhout, E. & Rozycki, W.V. 2008. *Contrastive Rhetoric: Reaching to Intercultural Rhetoric* [Pragmatics & Beyond New Series 169]. Amsterdam: John Benjamins.

Dahl, T. 2004. Textual metadiscourse in research articles: A marker of national culture or of academic discipline? *Journal of Pragmatics* 36: 1807–1825.

Duszak, A. 2005. Between styles and values: An academic community in transition. In *Identity, Community, Discourse*: *English in Intercultural Settings*, G. Cortese & A. Duszak (eds), 69–94. Bern: Peter Lang.

Ferguson, G. 2007. The global spread of English, scientific communication and ESP: Questions of equity, access and domain loss. *Ibérica* 13: 7–38.

Fløttum, K. 2005. The self and the others: Polyphonic visibility in research articles. *International Journal of Applied Linguistics* 15(1): 29–44.

Gianonni, D. 2008. Medical writing at the periphery: The case of Italian journal editorials. *Journal of English for Academic Purposes* 7(2): 97–107.

Hyland, K. 2001. Bringing in the reader: Addressee features in academic writing. *Written Communication* 18(4): 549–74.

Kerans, M.E. 2002. Close to home: Notes on the post-publication withdrawal of a Spanish research paper. *Ibérica* 4: 39–54.

Mauranen, A. 1993. *Cultural Differences in Academic Rhetoric: A Text-linguistic Study*. Bern: Peter Lang.

Mauranen, A., Swales, J.M. & Pérez-Llantada, C. 2010. Academic Englishes: A standardised knowledge? In *The Routledge Handbook of World Englishes*, A. Kirkpatrick (ed.), 634–654. London: Routledge.

Pérez-Llantada, C. 2008. Humans vs. machines? A multi-perspective model for ESP discourse analysis in intercultural rhetoric research. *ESP Across Cultures* 5: 91–104.

Pérez-Llantada, C. 2010a. The 'dialectics of change' as a facet of globalisation: Epistemic modality in academic writing. In *English for Professional and Academic Purposes*, M. Ruiz, J.C. Palmer & I. Fortanet, 25–42. Amsterdam: Rodopi.

Pérez-Llantada, C. 2010b. The discourse functions of metadiscourse in published writing: Issues of culture and language. *Nordic Journal of English Studies* 9(2): 41–68).

Rayson, P. & Garside, R. 2000. Comparing corpora using frequency profiling. In *Proceedings of* the *Workshop on Comparing Corpora, Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000)*. 1–8 October 2000, Hong Kong, 1–6.

Sarup, M. 1993. *An Introductory Guide to Post-Structuralism and Postmodernism*. New York NY: Harvester Wheatsheaf.

Scott, M. 1999. *Wordsmith Tools*. Oxford: OUP.

Swales, J.M. 2004. *Research Genres: Explorations and Applications*. Cambridge: CUP.

Vázquez Orta, I. & Giner, D. 2008. Beyond mood and modality: Epistemic modality markers as hedges in research articles. A cross-disciplinary study. *Revista Alicantina de Estudios Ingleses* 21: 171–190.

Vázquez Orta, I. & Giner, D. 2009. Writing with conviction: The use of boosters in modelling persuasion in academic discourses. *Revista Alicantina de Estudios Ingleses* 22: 219–237.

Vold, E.T. 2006. Epistemic modality markers in research articles: A cross-linguistic and cross-disciplinary study. *International Journal of Applied Linguistics* 16(1): 61–87.

White, P.R.R. 2003. Beyond modality and hedging: A dialogic view of the language of intersubjective stance. *Text* 23(2): 259–284.

Wilkinson, A. 1991. *The Scientist's Handbook for Writing Papers and Dissertations*. Englewood Cliffs NJ: Prentice Hall.

Structure, content and functions of calls for conference abstracts

Sara Gesuato
University of Padua

This paper examines the pragmatic functions and structural organisation of 100 Calls for Conference Abstracts (CfCAs) in Biology, Computing, History and Linguistics. These goal-oriented communicative acts focus on the representation-dissemination of disciplinary values and goals. They comprise the communicative core, the invitation – textually marginalised – and various auxiliary, textually prominent moves: informative, regulatory and persuasive-argumentative (announcements, offers, orders and requests). These qualify the CfCAs as representative, directive and commissive texts. The CfCAs are structurally similar – they share moves appearing in typical sequences – but do not instantiate an identical text type or structure potential, as no functional component is common to them all. The individual CfCAs thus exemplify the genre to different degrees of prototypicality, which paves the way for generic innovation.

Keywords: Conference abstracts, structure, genre

1. Introduction

CfCAs are standard forms of communication among scientists. Written by experts for their peers, they constitute invitations to participate in academic gatherings in order to present, share and discuss recent developments in given fields. However, CfCAs do not merely serve an organisational function for academic communities of practice. They also publicly display these communities' self-perception. In fact, CfCAs offer snapshots of what scholars consider valid work and relevant issues in their disciplines: they point out and delimit research domains (i.e. they identify trends and breakthroughs, and signal topics worth exploring), and also determine how to tackle future research questions (i.e. they set out conditions to be observed for proposals to be accepted); thus they formulate, circulate and maintain academic standards. As value-laden texts, CfCAs are worth examining on a par with other forms of scientific writing.

Within academic discourse, various written genres have been examined, including research article introductions (Swales 1990), abstracts (Bondi 2004), titles (Haggan 2004), review articles (Noguchi 2006), book reviews (Römer 2008), manuscript submission letters (Swales 1996), editorial letters to journal contributors (Flowerdew & Dudley-Evans 2002), recommendation letters (Trix & Psenka 2003), acknowledgements (Giannoni 2006) and back-cover blurbs (Gesuato 2007). However, CfCAs represent a relatively unexplored genre. To my knowledge, only two studies have taken them into consideration.

Räisänen (2002: 84–90) has defined CfCAs as a regulatory genre, which aims to direct the target readership's behaviour. Indeed, CfCAs manifest their directive illocutionary force in three ways: they *offer* scholars opportunities to participate in academic events, *impose* conditions for participation in those events, and also *request* scholars' participation and contributions. From a complementary perspective, Schneider (2007) has shown the importance of the genre's physical layout features in automatic information extraction. The typographic format of CfCAs, indeed, is a strong determinant of their recognisability and guides the reader in the identification and processing of their information units. However, the genre has not been *described* in terms of the topics it mentions or the moves it realises. This paper analyses the content, wording, functions and structure of CfCAs on the basis of corpus data so as to outline their prototypical instantiation, illustrate their degree of internal variation and exemplify their lexico-grammatical encoding.

2. Data

The texts analysed comprise 100 CfCAs in English, representative of four disciplines. I collected most of them on the internet by querying the Google search engine with the phrase *call for papers* combined with the terms *biology*, *computing*, *history* and *linguistics*.[1] In addition, I gathered some linguistics CfCAs through email, received partly from mailing lists and partly from colleagues. The corpus thus consists of four sub-corpora – Biology (BIO), Computing (COMP), History (HIST) and Linguistics (LING) – each including 25 CfCAs (total words: 56,366).

The amount of data examined – in terms of number of exemplars and number of words – is that appropriate for a specialised corpus (Bowker & Pearson 2002: 48) and is justified by the formal and functional homogeneity of the texts

1. The original output included calls for the contribution of chapters to books or articles to journal issues, which I disregarded.

considered. Previous studies have shown that valuable insights can be gained into the linguistic and textual nature of communicative acts by examining a small amount of well-selected data (e.g. Ghadessy et al. 2001), that is, data that comprises full texts (rather than extracts) and that represents the same type of specialised discourse material (i.e. relevant to the same topic, oriented towards the same purpose, and produced under similar contextual circumstances).[2]

The texts examined in this study meet the above requirements. They are comparable in content – the CfCAs are all relevant to the topics, goals, policies and logistics of academic conferences – they exemplify the same specialised language – i.e. academic English produced by conference organisers, presumably university lecturers – and they are motivated by the same communicative rationale – i.e. conveying information to a clearly defined target readership so as to convince them to participate in the events being organised.

As Table 1 shows, the average length of a CfCA is 564 words. But there are marked differences in length values within and among sub-corpora. In each sub-corpus, the longest CfCA is several times longer than the shortest one (i.e. 10, 14, 8 and 20 times longer in BIO, COMP, HIST and LING, respectively). Also, COMP and HIST have the shortest and longest CfCAs, respectively, while BIO and LING display intermediate values.

The distribution of the CfCAs across word length values (see Table 2) shows intra-corpus similarities between BIO and COMP, on the one hand, and HIST and LING, on the other. The former tend to cluster around 301–900 and 401–1,000 word length values, respectively, while the latter concentrate around 101–500 and 201–600 word length values, respectively. Overall, the corpus preference is for CfCAs ranging between 101 and 800 words (79%).

Table 1. Length of CfCAs in number of words

No. of words	BIO	COMP	HIST	LING	Overall
Total	15,368	18,768	9,570	12,660	56,366
Average per CfCA	615	751	383	506	564
In longest CfCA	1,475	2,739	990	1,628	1,708
In shortest CfCA	146	187	119	79	133

2. Similarly, Biber (1990: 269) has shown that in the study of language variation, too, "relatively short text lengths and small corpus size are often adequate". Also, Carter and McCarthy (1995: 143) have argued that even a corpus of tens of thousands of words can easily reveal grammatical patterns.

Table 2. Distribution of CfCAs across word length values

No. of words	BIO	COMP	HIST	LING	Overall
1–100	0	0	0	1	1
101–200	2	1	8	1	12
201–300	0	1	4	4	9
301–400	6	1	5	7	19
401–500	3	3	2	3	11
501–600	2	4	0	2	8
601–700	5	2	1	0	8
701–800	2	4	3	3	12
801–900	2	2	1	2	7
901–1,000	1	4	1	0	6
1,001–1,100	0	2	0	1	3
1,101–1,400	0	0	0	0	0
1,401–1,500	2	0	0	0	2
1,501–1,600	0	0	0	0	0
1,601–1,700	0	0	0	1	1
1,701–2,700	0	0	0	0	0
2,701–2,800	0	1	0	0	1
Total	25	25	25	25	100

3. Approach: Identification of moves

I adopted a bottom-up approach to analyse the texts. I familiarised myself with the content of the CfCAs through repeated readings. The recursive passes through the CfCAs enabled me to identify recurrent patterns – in terms of content and wording – and thus to codify text segments expressing specific functions, that is, moves (see sections 3.1–3.4).[3] I produced a first coding scheme of such inductively developed move categories on a subset of 40 texts, and later refined my categorisation on the basis of the larger corpus. Next, I tagged the whole corpus for its functional components. To increase intra-rater reliability, I re-tagged the corpus again from scratch after a year. When the comparison of my classifications revealed mismatches, I re-examined the relevant text segments, and corrected classification oversights, where applicable, or re-coded the text segments as appropriate (see Section 3.4). Finally, I compared the findings across the sub-corpora.

3. Hyland (2004) and Hyland and Tse (2004) adopted a similar approach in their studies of acknowledgements.

The identification and classification of moves in the CfCAs was meant to determine how their authors conceptualise these communicative acts. Given their average length (see Section 2), the CfCAs were likely to express more than invitations to participate in conferences, and thus to convey complementary notions sustaining the main communicative intent.

3.1 The rationale for the genre

The rationale motivating a CfCA is the invitation to participate in a given conference. The expression of this illocution constitutes the main move of the text, coded in the corpus as \Invitation\.[4] This counts as a request – because contributing to and attending the conference is in the interests of the conference organisers – and as an offer – because obtaining information and professional opportunities is advantageous to the target academic community. In the corpus, \Invitation\ is encoded as a sentence which represents the mental action of 'requesting, suggesting, wanting and/or accepting' – as signalled by such verbs as *call for, consider, encourage, invite, seek, solicit, welcome* – and which expresses the notion of 'object of invitation' – namely scholars and/or their work – as a verb argument. \Invitation\ can occur in either the active or the passive voice, depending on whether the semantic object of invitation is syntactically encoded as the object or subject, respectively, of the predicate; e.g.:

(1) With this call for papers the Organising Committee 1) cordially invites you to submit talks and posters (BIO–1)
(2) [...] welcomes original papers [...] (COMP–5)
(3) Early career researchers are encouraged to participate [...] (HIST–17)
(4) Original and unpublished research papers are solicited [...] (LING–3).

\Invitation\ is supported by informative moves, which qualify CfCAs as announcements, and regulatory moves, which aim to influence the target readership's behaviour.

3.2 Informative moves

The corpus exemplifies 10 types of informative moves: \Context\, \Definition\, \Detail\, \Goal\, \History\, \Introduction\, \Notice\, \Technical\, \Theme\ and \Topics\.

4. Here and elsewhere, move tags appear surrounded by backward slashes.

\Context\ refers to or identifies the spatio-temporal setting of the conference.[5] It is marked by the occurrence of predicates encoding the notion of 'taking place', accompanied by adverbials specifying the time and place of the conference; e.g.:

(5) This year ISC'06 will be held in Samos Island, Greece (city of Pythagoras, Doryssa - Bay Hotel www.doryssa-bay.gr) (COMP–5)
(6) PQCrypto 2006 will be held at the Katholieke Universiteit Leuven in Belgium from Wednesday 24 May 2006 through Friday 26 May 2006. (Eurocrypt 2006 will begin on Sunday 28 May 2006 in St. Petersburg) (COMP–7; original emphasis)
(7) The 2006 meeting of the Consortium on the Revolutionary Era, 1750–1850 will be taking place in Atlanta, GA, March 2–4, 2006 (HIST–4).

\Definition\ is a mini-classification or description of the conference, which indicates what this is about (i.e. its theme and/or goal). It can consist in a specification of what the conference is – in which case the relevant text segment includes a linking verb – or it may signal what the conference does – in which case it pivots on an action verb. The content of \Definition\ can be relevant to the conference as a regularly occurring event – as a sort of timeless statement – or to the next event being promoted – and thus projected into the future; e.g.:

(8) [...] will focus on [...] it will provide [...] (BIO–9)
(9) [...] is the world's largest conference and exhibition [...] (BIO–21)
(10) ISC is an annual international conference covering research in and applications of Information Security (COMP–5)
(11) [...] serves as a venue for learning about state-of-the-art in security [...] (COMP–6)
(12) [...] provides an opportunity for postgraduates and established academics to discuss aspects of their research (HIST–8)
(13) [...] promotes research on critical discourse and language matters [...] (LING–13).

\Detail\ specifies details of the conference programme, either its organisational aspects or its academic highlights. The move, therefore, indicates what the conference comprises (i.e. the "package" it offers) and/or reveals part of the background that characterises it; e.g.:

(14) It is supported by [...] (COMP–13)
(15) It consists of 14 workshops [...] (COMP–24)
(16) Keynote speakers will include [...] (HIST–4)

5. This move may also qualify as promotional, if it includes an attractive description of the conference venue; see Section 3.3.2.

(17) Some highlights already planned for the program include [...] (HIST–9)
(18) Additional event [sic] at the conference include, for example, a special session of film journal editors to talk with participants about publication opportunities and goals [...] (HIST–12).

\Goal\ points out what the conference aims to achieve and/or who it is for. It is marked by the employment of nouns or adjectives (e.g. *aim, goal, purpose, objective, devoted*), or verbs (e.g. *seek, intend, hope, would like, wish, want*) and verb forms (i.e. *will*-future) encoding the notion of 'orientation to a final, purposeful result'; these then introduce the methodological, social and/or scientific goals that officially justify the organisation of the conference; e.g.:

(19) The workshop intends to bridge gaps and difficulties (BIO–9)
(20) We would like to bring forth [...] (COMP–20)
(21) The major objective is to start exploring [...] (HIST–24)
(22) The objective of the symposium is to bring the linguists, computational linguists and the computer scientists together who are working on Indian languages in different parts of India and its neighboring countries (LING–3)
(23) The workshop will furthermore serve to connect [...] (LING–15).

\History\ is a descriptive account of the conference, which reports on its (regular) past activities and/or those of its promoting association. It combines terms denoting the unfolding, or chronological context of events (e.g. *last, previous, former, first, after, follow, take place*) with the use of present perfects and past tenses; e.g.:

(24) This successful series first started [...] (BIO–10)
(25) The first [...] workshop took place [...] (COMP–4)
(26) The scope [...] has been broadened since the inaugural 2005 event (COMP–6)
(27) The call for papers is posted internationally, and speakers have been known to come from across Western Canada, the United States, and Europe. Non- presenting attendees usually include local students and faculty. The venue rotates on a four year cycle, between [...] (LING–1).

\Introduction\ is an announcement about the decision to hold the conference. The text segment encoding it, indeed, conveys such notions as 'announcing, opening, hosting' or 'organising', followed by reference to 'the conference' as a newsworthy topic. \Introduction\ is always realised as a clause and occurs after \Notice\ (see below); e.g.:

(28) [...] we are extremely happy to host [...] (BIO–15)

(29) The 2007 annual meeting of the Michigan Academy of Science, Arts, and Letters will be held March 9–10 at Ferris State University in Big Rapids, Michigan (BIO–17)
(30) [...] is now open [...] (BIO–18)
(31) [...] is issuing a call for papers [...] (BIO–21)
(32) [...] is organizing its second annual conference [...] (COMP–3)
(33) The Society for the History of Technology will hold its annual meeting [...] (HIST–3)
(34) [...] are pleased to announce the hosting [...] (LING–3).

\Notice\ is a succinct text unit informing readers of the upcoming conference. It includes the minimum information that helps prospective participants orientate themselves (i.e. the name, venue and date of the conference). Unlike \Introduction\, it is encoded by means of phrases – rather than clauses – and also occurs at the very beginning of the CfCA, *before* the main body of the text; e.g.:

(35) 2nd IEEE International Workshop on Heterogeneous Multi-Hop Wireless and Mobile Networks 2006 (IEEE MHWMN'06) [...]) (COMP–22)
(36) REVISED Call for Papers for ICOHTEC 2006 // The International Committee for the History of Technology's 33rd Symposium in Leicester, U.K., 15–20 August 2006 // Deadline for proposals is 1 MARCH 2006 (HIST–1).[6]

\Technical\ informs readers about the logistics of the conference (e.g. contact details, accommodation options, facilities, the scientific committee, special arrangements). The practical issues addressed in the move usually form the topic portion of the message – encoded as subjects – but occasionally it is the conference participants that are mentioned first, addressed as direct interlocutors; e.g.:

(37) All the meeting information is posted [...] (BIO–15)
(38) Hotel accommodations are not included in the workshop registration fee [...] (COMP–7)
(39) Visit the website at [...] (HIST–12)
(40) The conference offers an honorarium to graduate students [...] (HIST–21)
(41) WARNING: Unfortunately, getting a Russian visa for a business trip is not easy [...] (LING–12)
(42) You will only have to take care of your travelling costs [...] (LING–15).

\Theme\ is the move that synthesises the focus of the conference. The message encoding it usually starts off with a topic portion identifying the conference or the general notion of 'theme', followed by a comment portion that indicates the specific

6. Here and elsewhere, double slashes signal the beginning of a new paragraph.

theme of the conference; the use of linking verbs is thus frequent in instances of this move; e.g.:

(43) A central theme will be [...] (BIO–4)
(44) The IMECS 2006 multiconference has the focus on the frontier topics in the theoretical and applied engineering and computer science subjects (COMP–24)
(45) ***Transforming Economies and Civilizations: The Role of Technology*** is the broad general theme of ICOHTEC's 33rd symposium (HIST–1; original emphasis)
(46) The symposium intends to focus on challenges in modelling [...] (LING–3)
(47) The leitmotif of this edition is going to be **"Reaching Far: Distant Countries, Distant Disciplines"** (LING–10; original emphasis).

\Topics\ presents the topics and/or approaches relevant to the conference. This move is worded as an introductory statement expressing what the conference organisers recommend as suitable for – or pertaining to – the conference, followed by a list of the suggested topics; e.g.:

(48) Its scope includes computational problems in the interpretation and analysis of molecular biological data including: // Data management methods and systems [...] (BIO–6)
(49) This year's program format is designed to address multiple topic areas including [...] (BIO–18)
(50) We particularly encourage submissions on any of the following topics [...] (COMP–12)
(51) Submissions might address [...] (HIST–5)
(52) From the important topics for research, we can suggest the following (this list is not exhaustive) [...] (LING–7).

3.3 Regulatory moves

Regulatory moves are meant to determine the addressee's responsive behaviour. They include three types: directive, promotional and persuasive-argumentative.

3.3.1 *Directive moves*

Directive moves are prescriptive, as they inform prospective participants of what is expected of them and, more generally, what "ideal model" the conference is supposed to conform to. They include \Policies\ and \Request\, but also \Theme\ and \Goal\.

\Policies\ spells out conference rules to be complied with (e.g. about the language of the conference, submission of abstracts, selection procedure, membership

requirements, publication guidelines), and thus specifies what prospective participants can, must or must not do. The impositional role of the move may be conveyed directly – through the use of deontic modality – or indirectly – through unhedged predictive statements that inform the readership about ready-made plans and non-negotiable future courses of action, single-handedly decided on by the conference organisers; e.g.:

(53) You can submit [...] (BIO–1)
(54) Submissions must not be concurrently under review [...] (COMP–6)
(55) [...] Submissions will be judged on correctness, originality, technical strength, significance and quality of presentation (COMP–22)
(56) The abstracts, in Italian or English, will be sent in Word format by e.mail to the Scientific Seminary Secreteriat [sic] [...] (HIST–18)
(57) The language of the forum is English (LING–16).

\Request\ is an appeal for help – uttered from a one-down interactional position – in which the addressee's inclination to comply is represented as hypothetical (e.g. through the use of the conditional) or spontaneous (e.g. through the use of the verb or noun *volunteer*). When \Request\ is encoded as a directive, this is qualified by the politeness marker *please*, and when realised as a want statement, this comes across as non-impositional because it is associated with verbs that tone down the strength of its illocutionary point; e.g.:

(58) If you want to [...] chair a session, evaluate papers [...] or any other offer to help please send an email [...] (BIO–12)
(59) [...] we ask presenters to volunteer [...] (BIO–16)
(60) We also welcome any volunteers to chair and/or comment on sessions (HIST–9)
(61) Please forward this message to any person you know who would be interested in the event (HIST–11).

\Theme\ and \Topics\ influence the readership's behaviour in that, by delimiting the domain within which research contributions are allowed, they serve a gatekeeping function. However, since their primary role is to *inform* readers about the conference (see Section 3.3), they can be classified as indirect directive moves.

3.3.2 *Promotional moves*

Promotional moves advertise assets of the conference and arouse the reader's interest. They include \Context\ and \Detail\.

\Context\ plays a promotional role when it describes the conference location as an attractive sight-seeing destination and/or the conference programme as an intellectually appealing event; e.g.:

(62) Historical buildings and castles in Innsbruck include [...] (BIO–1)
(63) Stellenbosch is situated in the fertile Jonkershoek river valley, surrounded by vineyards, orchards and mountains. [...] Besides the spectacular beauty of Stellenbosch and the nearby city of Cape Town we also promise you an exciting scientific programme presented by international experts (LING–11).

\Detail\ serves a promotional function when its content is presented as advantageous to the reader, that is, when it highlights what makes the conference a worthwhile experience; e.g.:

(64) Our two keynote speakers are [...] Also, the dates of our conference are chosen so that interested students will have the opportunity to attend the annual AAAS Forum on S&T Policy [...] (BIO–20)
(65) The conference will feature original, refereed papers, software demonstrations [...] (BIO–25).

3.3.3 *Persuasive-argumentative moves*

Persuasive-argumentative moves aim to ensure the success of the conference. They comprise \Background\, \Exhortation\ and \Gap\, and indirectly \Goal\ and \Theme\.

\Background\ sets the stage for the conference by outlining the current state of a discipline. This involves making reference to past events (possibly with current relevance), present situations, timeless circumstances and/or future perspectives represented as pertaining to the theme of the conference; e.g.:

(66) Artificial life is a field of study devoted to understanding [...] (BIO–9)
(67) Technology diversity is growing fast [...] (COMP–3)
(68) Will large quantum computers be built? [...] (COMP–7)
(69) The research area of mobile computing has become more important following the recent widespread drive towards wireless sensor networks and location tracking technologies and their applications [...] (COMP–20)
(70) Although British cinemas were initially closed in at the beginning of war [...] (HIST–12)
(71) [...] two opposing views on linguistic facts have been competing with each other [...] (LING–21).

\Exhortation\ is an explicit appeal to potential participants to attend the event. It is encoded as a wish statement – in the indicative or conditional mood – which expresses the organisers' hopes about the conference, or as a directive – in the imperative or conditional mood – which points out what is in the conference for prospective attendees; e.g.:

(72) We would greatly appreciate your visit! (BIO–5)
(73) We look forward to seeing familiar faces as well as meeting new students and hearing about all the exciting projects that are currently underway. // Sincerely, (BIO–15)
(74) [...] ATTEND [...] AND ENJOY [...] A TOTAL SOLAR ECLIPSE (BIO–25; original emphasis)
(75) Look forward to seeing you (LING–8).

\Gap\ identifies an issue to be addressed or a goal worth pursuing. Its role in the CfCA is to point out options, problems and doubts, which provide a niche for academics to occupy with their research. It can be worded in the form of questions – which academics may want to answer, and which should attract the wider readership's attention – or be exploited to create a bond with prospective participants – through the use of an inclusive *we* – or presented more neutrally as a statement of fact, which introduces an issue for the scholarly community at large to tackle; e.g.:

(76) We now have the ability to recreate [...] (BIO–2)
(77) But is the governance of genomics special? If so, what makes it special? What modes of regulating the biosciences are currently emerging? Does 'the public' have a special role to play in the governance of genomics? [...] (BIO–22)
(78) [...] Exactly which of these systems are secure? (COMP 7)
(79) [...] Can the actual research carried out on Occitan lexicology and lexicography lead to solutions? (LING–7)
(80) However, so far such systems do not meet the expectations of ordinary users [...] (LING–15).

Finally, \Goal\ and \Theme\ are classifiable as indirect persuasive-argumentative moves because, in signalling what is hot in a given field – what is worth pursuing both in general and at a specific conference – they reveal the relevance and value of the conference.

3.4 Double moves

Occasionally, a text segment may require a twofold move classification because it fulfils two functions simultaneously. Typically, this means that the text segment is one clause or sentence long, and that its functions pertain to different syntactic constituents of it. For instance, in Example 81 the main clause encodes a \Definition\, while the subordinate one a \Goal\; in Example 82 the subject + predicate + object part of the clause encodes an \Invitation\, while the adverbial encodes a \Policies\ move; finally, in Example 83 the subject + predicate + subject_complement part encodes an \Invitation\, and the rest of the clause a \Theme\:

(81) The international conference on Field-Programmable Technology is a forum to present and discuss new research on the technology and the use of field-programmable devices for high performance and/or low energy computation (COMP–9: \Definition + Goal\)

(82) We invite submissions of abstracts for 20 minute presentations (plus 10 minute discussion) on any topic pertaining to formal semantics (LING–9: \Invitation + Policies\)

(83) We are especially interested in parallel algorithms, memory-efficient algorithms, large scale data mining techniques, and design of high-performance software (BIO–24: \Theme + Invitation\).

4. Findings

The tagging of the CfCAs reveals that the corpus contains 825 move tokens, and thus, on average, 207 per sub-corpus. However, while the number of move tokens in BIO, COMP and LING approximates the corpus average value, that of HIST stands out for being much lower. The move token frequency hierarchy is BIO > COMP > LING > HIST. The number of moves per text ranges between 2 and 17, the average being 8. Here too, BIO, COMP and LING display similar range values, while HIST has a much lower one. Over half of the CfCAs contain repeated moves, that is, multiple instantiations of one or more move types. This finding applies both to the corpus as a whole and to the BIO, COMP and LING sub-corpora. HIST, too, contains repeated moves, but in less than 50% of its texts (see Table 3).

The average number of words per move token can vary from 20 to 304. The value ranges in the sub-corpora are quite comparable, except for COMP – the longest sub-corpus – whose highest average value per move token is much higher. The global average values per move token and move type are 67 and 79 words, respectively. BIO, HIST and LING approximate these values, while COMP displays

Table 3. Frequency and distribution of moves, and length of CfCAs in number of moves

No. of	BIO	COMP	HIST	LING	Overall	Average
Move tokens	233	221	159	213	825	207
Repeated move tokens in same CfCA	72	61	28	49	210	53
Repeated move types in same CfCA	33	28	13	23	97	24
CfCAs with repeated moves	17	17	10	13	57	14
Moves in move-longest CfCA	17	15	9	15	–	14
Moves in move-shortest CfCA	5	5	2	4	–	4

Table 4. Length of moves in number of words

No. of	BIO	COMP	HIST	LING	Average
Words per move token: lowest average value in single CfCA	26	20	20	20	22
Words per move token: highest average value in single CfCA	155	304	165	130	146
Words per move token: overall average value	66	85	60	59	67
Words per move type: overall average value	81	100	66	68	79

higher ones (see Table 4). The distribution of the moves over length value ranges shows that most moves in the corpus consist of 100 or fewer words. In particular, BIO, COMP and LING and almost half of the corpus show a preference for moves up to 100 words long, and HIST for moves up to 50 words long. Only BIO and COMP occasionally have very long moves (see Table 5).

The relative frequency of occurrence of the various moves is homogenous across the sub-corpora. Only \Detail\ and \Technical\ have percentage frequency values ranging over more than 5 points, and only \Introduction\, \Notice\ and Double Moves approximate that range of values (see Table 6).

Table 7 shows the distribution of move types in the corpus, that is, the number of CfCAs in which they are instantiated. The most frequent move types, i.e. \Invitation\, \Notice\, \Policies\, \Technical\ and \Topics\, are instantiated in over 50% of the texts of each sub-corpus. Their distribution across the sub-corpora suggests that they are probably constitutive of the genre. On the other hand, move types whose distribution values range between 31% and 50% differ across sub-corpora, even if each sub-corpus shares at least one such move type with the others (i.e. \Introduction\, \Definition\, \Background\, \Context\, \Detail\ in BIO; \Background\,

Table 5. Distribution of moves across average word length values

Average No. of words per move	BIO	COMP	HIST	LING	Overall
1–50	9	5	12	10	36
51–100	15	12	8	13	48
101–150	0	5	4	2	11
151–200	0	1	1	0	2
201–250	1	1	0	0	2
301–350	0	1	0	0	1
Total	25	25	25	25	100

Table 6. Frequency of move tokens across sub-corpora

Move	BIO	COMP	HIST	LING	% Range
\Background\	9	15	7	15	3.9–7.0
\Context\	9	6	3	9	1.9–3.9
\Definition\	10	11	3	4	1.9–4.3
\Detail\	10	14	9	22	4.3–10.3
\Exhortation\	3	1	1	4	0.5–1.9
\Gap\	2	4	8	8	0.0–3.8
\Goal\	17	16	11	12	5.6–7.2
\History\	10	8	2	23	0.9–4.3
\Introduction\	11	1	6	6	0.5–4.7
\Invitation\	21	19	19	18	8.5–11.9
\Notice\	24	23	24	23	10.3–15.1
\Other\	7	5	1	2	0.6–3.0
\Policies\	34	38	28	35	14.6–17.6
\Request\	2	1	3	0	0.0–1.9
\Technical\	38	22	20	28	9.9–16.6
\Theme\	2	3	3	04	0.9–1.9
\Topics\	21	22	15	17	8.0–9.9
Double moves	3	12	4	7	1.2–5.5

\Definition\, \Detail\ and \History\ in COMP; \Detail\ and \Goal\ in HIST; and \Background\, \Goal\ and \Gap\ in LING). The distribution values of these moves suggest that these enhance the genre without being crucial to it. Finally, move types with a low distribution value (i.e. between 11% and 30%) differ markedly across the sub-corpora (i.e. \Other\ and \Exhortation\ in BIO; \Context\, \Gap\, \Theme\ and \Other\ in COMP; \Background\, \Introduction\, \Context\, \Definition\, \Request\ and \Theme\ in HIST; and \Gap\, \Context\, \Introduction\, \Definition\, \Exhortation\ and \Theme\ in LING). These moves can be said to expand on information already available elsewhere in the CfCAs.

The move types identified in the CfCAs are numerous and varied; therefore, none stands out prominently in the corpus. Table 8 shows their comparative frequency of occurrence, i.e. their share of textual space, both within and across sub-corpora. Only 5 of the 18 move types identified (i.e. \Invitation\, \Topics\, \Goal\, \Detail\ and \Background\) have a frequency value between 5% and 10%, while only 3 (i.e. \Policies\, \Technical\, and \Notice\) account for 10% or more of the move tokens instantiated. Similar frequency values can be found in the four sub-corpora, with only minor variations (e.g. \Technical\ is more frequent than \Policies\ in BIO; \Technical\ is less frequent than \Notice\ in COMP and HIST; and \Detail\ is more frequent than \Invitation\ in LING).

Table 7. Distribution of move types in the corpus in decreasing order of frequency

	No. of CfCAs in				
Move type	**BIO**	**COMP**	**HIST**	**LING**	**Corpus**
\Notice\	24	23	24	23	94
\Policies\	23	24	22	25	94
\Technical\	23	19	18	20	80
\Invitation\	17	18	16	18	69
\Topics\	19	19	14	17	69
\Goal\	14	14	8	11	47
\Detail\	8	10	9	17	44
\Background\	9	12	6	12	39
\Definition\	10	11	3	4	28
\Introduction\	11	1	6	6	24
Double moves	3	10	3	7	23
\Context\	8	5	3	6	22
\History\	6	8	2	2	18
\Gap\	2	4	0	8	14
\Theme\	2	3	3	4	12
\Other\	7	3	1	2	13
\Exhortation\	3	1	1	4	9
\Request\	2	1	3	0	6

A clearer picture of move type comparative frequency emerges if the moves are grouped into the main categories presented in sections 3.1–3.3.3, as in Table 9. Informative moves make up over 60% of the texts, while direct persuasive-argumentative and direct directive moves occupy over 22% and 17% of the corpus, respectively. \Invitation\, instead, is relevant to less than 10% of the corpus. The genre's crucial move is thus technically marginalised in the texts – it probably can be taken for granted by expert readers – while the bulk of the texts consists of information units meant to orientate readers in their approach to and participation in the conferences.

Despite the similarities in the frequency of the moves in and across the sub-corpora, it is not possible to outline the textual organisation of the CfCAs in the form of a generic structure potential. This is due to the distribution and realisation of the moves. No move type is shared by all the CfCAs – in particular, of the four sub-corpora, only LING has one move shared by all its CfCAs, namely \Policies\ – and a few moves occur infrequently (e.g. \Exhortation\ and \Request\). Also, certain moves recur in the same CfCA, or may occupy discontinuous text segments. Finally, some text units serve two communicative purposes simultaneously, showing a lack of correspondence between text excerpts and communicative functions.

Table 8. Comparative frequency of moves

	Frequency in %				
Move tokens	**BIO**	**COMP**	**HIST**	**LING**	**Overall**
\Policies\	14.6	17.2	17.6	16.4	16.3
\Technical\	16.3	9.9	12.6	13.2	13.1
\Notice\	10.3	10.4	15.1	10.8	11.4
\Invitation\	9.0	8.6	11.9	8.5	9.3
\Topics\	9.0	9.9	9.4	8.0	9.1
\Goal\	7.2	7.2	6.9	5.6	6.8
\Detail\	4.3	6.3	5.7	10.3	6.7
\Background\	3.9	6.8	4.4	7.0	5.6
\Definition\	4.3	5.0	1.9	1.9	3.4
Double moves	1.2	1.4	0.8	1.1	3.1
\Context\	3.9	2.7	1.9	2.8	2.9
\Introduction\	4.7	0.5	3.8	2.8	2.9
\History\	4.3	3.6	1.3	0.9	2.7
\Other\	3.0	2.3	0.6	0.9	1.8
\Gap\	0.9	1.8	0.0	3.8	1.7
\Theme\	0.9	1.3	1.9	1.9	1.4
\Exhortation\	1.3	0.5	0.6	1.9	1.1
\Request\	0.9	0.5	1.9	0.0	0.7

Table 9. Comparative frequency of move categories

Move category	**Frequency in %**
Informative	60.4
Directive	17.0
(Indirect directive)	(14.5)
Promotional	4.9
\Invitation\	9.3
Persuasive-argumentative	8.4
(Indirect persuasive-argumentative)	(8.2)
Total	100

However, the moves cluster in preferred sequences, which reveal typical textual patterns. Tables 10 and 11 show, respectively, the frequency of occurrence of text-initial and text-central/final move sequencing patterns in percentage values. That is, they specify how often certain (pairs or clusters of) moves occur in the texts before or after other (pairs or clusters of) moves. The CfCAs appear to be

Table 10. Text-initial move sequencing patterns

Move(s)		Move(s)	Frequency*
\Notice\	before	Rest of the CfCA	96%
\Notice\	before	\Introduction\ ~ \Context**	100%
\Notice\ ~ \Introduction\ ~ \Context\	before	\Background\ ~\Definition\ ~ \Goal\ ~ \History\ ~ \Theme\	99%
\Background\	before	\Gap\	93%
\Background\ (\Gap\)***	before	\Definition\ ~ \Goal\ ~ \History\ ~ \Theme\	100%
\History\ ~ \Definition\	before	\Goal\ ~ \Theme\	84%

*Percentages refer to the subset of CfCAs containing given moves.

**Tildes signal combinations of or alternation between moves.

***Parentheses indicate optionality.

Table 11. Text-central/final move sequencing patterns

Move(s)		Move(s)	Frequency*
\Background\ ~ \Definition\ ~ \Goal\ ~ \History\ ~ \Theme**	before	\Invitation\	82%
\Invitation\	before	\Topics\	79%
\Topics\	before	\Policies\	96%
\Policies\	before	\Exhortation\ ~ \Request\ ~ \Detail\	55%
\Policies\ ~ \Exhortation\ ~ \Request\ ~ \Detail\	before	\Technical\	96%
\Technical\	after	Rest of text	84%

*Percentages refer to the subset of CfCAs containing given moves.

**Tildes signal combinations of or alternation between moves.

characterised by a favourite sequential structure, suggesting that the prototypical sequence of moves in a CfCA is the following:

\Notice\ > \Introduction\ > \Context\ > \Background\ > \Gap\ > \History\ ~ \Definition\ > \Theme\ ~ \Goal\ > \Invitation\ > \Topics\ > \Policies\ > \Technical\.[7]

7. Here and elsewhere, tildes signal possible combinations of or alternatives between moves.

The patterns relevant to pairs or clusters of moves apply to the vast majority of the texts, with one exception: the occurrence of \Exhortation\ ~ \Request\ ~ \Detail\ after \Policies\ applies to just over half of the texts instantiating those moves. Indeed, the moves \Exhortation\ ~ \Request\ ~ \Detail\ do not appear to be tied to the others in any semi-fixed order.

5. Discussion

CfCAs are functionally complex texts. They act as announcements (referring to developments and perspectives in given fields), offers (providing scholars with opportunities to exchange ideas), orders (establishing criteria for conference participation) and invitations (manifesting organisers' desire for scholars to attend conferences). This justifies the variety of their moves.

In the data examined, the rationale for a CfCA is provided by the main move, \Invitation\, which identifies the genre's communicative purpose (i.e. inciting the audience to respond by taking action beneficial to the CfCA's authors and addressees). This core move is supported by auxiliary ones relevant to the text's multiple functions, which are: to inform of the conference (e.g. \Notice\, \Definition\, \History\), promote it (e.g. \Context\), impose its rules (e.g. \Policies\), and persuade prospective participants of its value (e.g. \Background\). The auxiliary moves highlight the multi-faceted communicative nature of the CfCA and sustain, motivate and promote the acceptability and effectiveness of the main move.

In the present corpus, the supporting moves may be pragmatically elaborate; for instance, they may serve two purposes simultaneously (e.g. \Topics\ is directly informative, but also indirectly regulatory) or they may serve different general purposes depending on their degree of elaboration (e.g. \Context\ may be succinct and informative, or detailed and evaluative-promotional). Furthermore, they may be textually complex; for example, they may recur in the CfCA or be realised over discontinuous texts segments, or alternatively, a text segment may realise two moves simultaneously, that is, be relevant to different functions (e.g. \Invitation + \Policies\; see Example (82)). Most of the time, however, they can be easily identified and classified on the basis of their content and wording.

The CfCAs examined comprise 16 move types (plus \Other\ and double moves), whose frequency and distribution values are similar across sub-corpora, but with stronger similarities between BIO and COMP, and between HIST and LING. No move is common to all the CfCAs, which paves the way for generic innovation (e.g. CfCAs written in the form of letters). Yet, each CfCA shares moves with other CfCAs. Also, the number, ordering and lexico-grammatical encoding of moves in the CfCAs is predictable across the sub-corpora (e.g. 97% of the texts

are under 1,000 words; 81% have between 6 and 12 moves; all the CfCAs exhibit certain moves in specific sequences: \Notice\ > \Theme\ ~ \Goal\ > \Topics\ > \Policies\ > \Technical\). All of this favours generic stability. Therefore, conformity to a textual construct is favoured across the texts and sub-corpora. The CfCAs thus appear to instantiate the same genre, but to different degrees of prototypicality, some texts being central and others peripheral exemplars of the same textual category (Paltridge 1995).

The findings reveal inter-disciplinary similarities and differences across the sub-corpora. HIST stands out as the sub-corpus with the lowest number of words and moves, and COMP for having the highest number of words and, on average, the longest moves. Also, BIO and COMP, on the one hand, and HIST and BIO, on the other, display similar word length values. But with regard to the frequency of occurrence of move tokens, the distribution of move types and the combinatorial possibilities of pairs or clusters of moves, the four sub-corpora reveal similar preferences. This suggests that, while discipline-specific norms and habits may impact the content and style of a genre – adapting it to local needs – its basic structure remains constant across fields, as determined by its primary purpose and as evidenced by the means through which its success is ensured, namely, its auxiliary moves.

The supporting moves reveal the communicative complexity of the CfCAs, serving orientational-organisational and promotional-advertising goals.

Moves with an orientational goal guide prospective participants into the conceptual space-time of the conference (e.g. \Notice\, \Definition\). Moves serving an organisational purpose are relevant to the imposition of rules (e.g. \Policies\), the provision of helpful instructions (\Technical\), and the involvement of the readership in the upcoming events (\Invitation\, \Exhort\).

Other moves are meant to favour the success of the conference by providing a solid justification for it or by serving as bait. Thus, some present the conference-to-be as a worthwhile event, motivated on rational grounds. They assert the scientific soundness of the conference and the well-foundedness of the decision to hold it; for instance, \History\ attests to the conference's credentials via reference to its past successes; \Background\ and \Gap\ – possibly inspired by the reporting disciplinary genre of academic paper introductions – create a niche, respectively, for the topic and purpose of the CfCA by identifying intellectual, scientific and professional needs to be met at the conference; and \Definition\ and \Goal\ motivate the text by pointing out how those needs will be satisfied.[8]

8. \Background\ and \Gap\ are reminiscent, respectively, of the moves 'Establishing the field or territory' and 'Preparing for current research' or 'Establishing a niche' in research article introductions (Swales 1990).

Other moves, instead, increase the effectiveness of \Invitation\ by acting as offers of incentives for prospective participants – while also establishing the conference organisers' credentials (e.g. \Detail\ and \Context\) – or even by directly soliciting a reaction from the readership (i.e. \Exhortation\).

Thus, on the one hand, the CfCAs frequently instantiate the rhetorical acts of informing and directing (see Table 9). This is in line with their expected purposes of orientating about and organising upcoming events, and makes them primary members of the genre colonies of Notices/Announcements and Instructions. On the other hand, CfCAs may instantiate additional rhetorical acts (i.e. describing the essence and importance of the conferences, narrating their history and reporting on the evolution of their fields.), and this fosters the success of the future events. As a result, the CfCAs also display features of promotional-advertising texts, of which they can be considered secondary members.

The appropriation of promotional discursive resources by CfCAs is an instance of genre colonization (a widespread phenomenon across discursive domains; see Bhatia 2004). Its occurrence may be accounted for by the ever-increasing number of such events being organised and the consequent need for conference organisers to compete with one another for visibility, appeal, credibility, resources, and turnout, as if in a marketplace: promotion guarantees survival and fosters success. Given the nature of the "product" to be "marketed", most of the textual strategies adopted aim to rationally appeal to the target audience's professional needs, although occasional appeals to their emotions may also occur (e.g. \Exhortation\ and \Context\).

This does not mean that the genre loses its integrity (Bhatia 1999: 23; Bhatia 2004). Indeed, it is highly recognisable due to its conventional, standardised (co(n))textual characteristics. Yet, the intertwining of promotional-persuasive purposes with informational-directive ones bends the CfCA genre towards advertising discourse, thus giving rise to genre-mixing.

6. Conclusion

This work represents a corpus-based exploratory study of the CfCA genre. Its findings provide a succinct account and illustration of the structure, lexico-grammatical features and textualisation patterns of CfCA exemplars. Therefore, the description and interpretation of the genre offered here can be further enhanced by adopting a multi-dimensional, multi-perspective framework (Bhatia 2004: 160–168).

First of all, from a textual perspective, my analysis has been carried out manually, and without the benefit of a second data-coder's contribution. By adopting the alternative approach of analysing the texts with corpus software, complementary

findings can be obtained providing evidence in support of or against those presented here. For example, a frequency word list, a cluster analysis and a keyword analysis of the texts, respectively, could identify the content vocabulary most frequently instantiated in the corpus, its typical multi-word units, and its especially (i.e. statistically) prominent terms. This could reveal the phraseologies giving the texts their distinctive lexical flavour and also the recurrent topics contributing to a definition of the aboutness of the texts. Also, concordances of especially frequent or key terms could highlight textual patterns, both grammatical patterns (e.g. the occurrence of terms having a cohesive or text-organising function in given message positions – theme or rheme) and semantic patterns (e.g. recurrent associations between the representation of certain types of events with given participants and the roles these are cast in).

Also, a deeper and more thorough understanding of the genre could be achieved by examining socio-cultural aspects of CfCAs not considered here. Thus, a replication study could be set up with a comparable set of CfCAs, for the purpose of comparing genre exemplars across disciplines (e.g. humanities, social sciences, hard sciences) and cultures (e.g. English, Australian, US, Indian). This type of analysis would identify features typical of the genre and distinguish them from those attributable to discipline- or culture-specific sub-genres, and thus better determine the possible degree of intra-generic variation and its correlation with elements of the context of situation.

Furthermore, the peculiarity and integrity of the CfCA genre could better be defined – from a tactical, socio-cognitive perspective – by comparing it with domain-specific sub-genres (e.g. calls for abstracts for volumes or journals, calls for papers), but also with genres having similar communicative purposes (e.g. degree programme and course descriptions) or focused on the same topic (e.g. post-conference reports) as well as neighbouring genre colonies (e.g. invitations, announcements, instructions, ads) so as to identify their recurrent topics and phraseologies. This would make it possible to determine the specificity of the genre "from the outside", that is, by identifying the linguistic-textual boundaries within which it "takes place".

More generally, from a professional-ethnographic perspective, the genre could be more accurately and comprehensively examined by interviewing CfCAs addressers and addressees on the values and expectations attached to such texts, on the functions envisaged and recognised in them, on their intended and perceived effectiveness, and on the production and reception strategies employed in using them.

Finally, the analysis of a greater number of CfCAs would lead to more reliable generalisations about the genre and a better characterisation of its prototypical exemplars; in addition, it would make it possible to carry out systematic comparisons of writing practices across disciplines.

Further analyses can thus provide deeper valuable insights into the CfCA genre. Within its limitations, this study shows how academic texts apparently produced and used for "housekeeping" purposes are not necessarily value-neutral, and can, in fact, help establish and share community standards, practices and viewpoints. In the case of CfCAs, this can be related to their goal-oriented nature and their co-textualisation. Among other things, CfCAs are meant to give out instructions, and thus impose rules of professional conduct. This is apparent in their auxiliary moves, which serve a logistic and a statutory role in the corpus. On the one hand, they favour a smooth running of the conference and foster its success. On the other, they shape and spread disciplinary values: by identifying recent relevant trends and issues, signalling topics worth exploring, and setting out conditions to be complied with for future work to be acceptable, they publicly display what academics perceive their work to be about and they help determine in what direction future work should proceed.

CfCAs thus appear to find their rightful place in the system of scholarly, scientific genres (Bazerman 1994), which collectively shape and regulate the lives of academics: in managing disciplinary practices, they influence experts' publicly accepted involvement in them.

References

Bazerman, C. 1994. Systems of genres and the enactment of social intentions. In *Genre and the New Rhetoric*, A. Freedman & P. Medway (eds), 79–101. London: Taylor & Francis.

Bhatia, V.K. 1999. Integrating products, processes, purposes and participants in professional writing. In *Writing: Texts, Processes and Practices*, C.N. Candlin & K. Hyland (eds), 21–39. London: Longman.

Bhatia, V.K. 2004. *Worlds of Written Discourse: A Genre-based View*. London: Continuum.

Biber, D. 1990. Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing* 5(4): 257–269.

Bondi, M. 2004. The discourse function of contrastive connectors in academic abstracts. In *Discourse Patterns in Spoken and Written Corpora* [Pragmatics & Beyond New Series 120], K. Aijmer & A.-B. Stenström (eds), 139–156. Amsterdam: John Benjamins.

Bowker, L. & Pearson, J. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

Carter, R. & McCarthy, M. 1995. Grammar and the spoken language. *Applied Linguistics* 16(2): 141–158.

Flowerdew, J. & Dudley-Evans, T. 2002. Genre analysis of editorial letters to international journal contributors. *Applied Linguistics* 23(4): 363–489.

Gesuato, S. 2007. Structural and generic complexity in back-cover blurbs of academic books. In *Cityscapes: Islands of the Self*, L. Jottini, G. Del Lungo & J. Douthwaite (eds), 419–432. Cagliari: CUEC.

Ghadessy, M., Henry, A. & Roseberry, R. 2001. *Small Corpus Studies and ELT: Theory and Practice* [Studies in Corpus Linguistics 5]. Amsterdam: John Benjamins.

Giannoni, D.S. 2006. Books acknowledgements across disciplines and texts. In *Academic Discourse Across Disciplines*, K. Hyland & M. Bondi (eds), 151–175. Bern: Peter Lang.

Haggan, M. 2004. Research paper titles in literature, linguistics and science: Dimensions of attraction. *Journal of Pragmatics* 36: 293–317.

Hyland, K. 2004. Graduates' gratitude: The generic structure of dissertation acknowledgements. *English for Specific Purposes* 23: 303–324.

Hyland, K. & Tse P. 2004. "I would like to thank my supervisor". Acknowledgements in graduate dissertations. *International Journal of Applied Linguistics* 14(2): 259–275.

Noguchi, J. 2006. *The Science Review Article: An Opportune Genre in the Construction of Science*. Bern: Peter Lang.

Paltridge, B. 1995. Working with genre: A pragmatic perspective. *Journal of Pragmatics* 24: 393–406.

Räisänen, C. 2002. The conference forum: A system of interrelated genres and discursive practices. In *The Language of Conferencing*, E. Ventola, C. Shalom & S. Thompson (eds), 69–93. Bern: Peter Lang.

Römer, U. 2008. Identification impossible? A corpus approach to realisations of evaluative meaning in academic writing. *Functions of Language* 15(1): 115–130.

Schneider, K.-M. 2007. Information extraction from calls for papers with conditional random fields and layout features. *Artificial Intelligence Review* 25(1–2): 67–77.

Swales, J. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: CUP.

Swales, J. 1996. Occluded genres in the academy: The case of the submission letter. In *Academic Writing: Intercultural and Textual Issues* [Pragmatics & Beyond New Series 41], E. Ventola & A. Mauranen (eds), 45–58. Amsterdam: John Benjamins.

Trix, F. & Psenka, C. 2003. Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse & Society* 14: 191–221.

Summarizing findings

An all-pervasive move in open access biomedical research articles involves rephrasing strategies

Mercedes Jaime-Sisó
University of Zaragoza (Spain)

This study shows how the interdisciplinary nature of scientific research, the tremendous growth in the number of journals and the easy access to information via online publications have influenced the way biomedical scientists tackle their professional reading and writing. Research articles online incorporate new presentational components - blurbs, authors' summaries, synopses - sharing with traditional sections the need to attract the busy reader's attention to the main results. Findings are anticipated in the title and reformulated in all article sections, using different linguistic structures to avoid semantic and syntactic repetitions. From interviews with scientists and the analysis of 20 articles from two prestigious journals, I conclude that this rephrasing phenomenon, which mainly affects open access publications, complicates RA writing for NNSs and may be announcing a new form of scientific information presented in well-defined cognitive textual modules.

Keywords: Scientific research articles, summarizing findings, rephrasing, online publications, reading behaviour, discourse analysis

1. Introduction

In the last three decades, a huge number of studies have been published on the structure and functions of scientific research articles (RAs). These generic analyses have provided valuable insight into how and why members of the scientific discourse communities communicate in the way they do, and have been the basis for descriptive frameworks that can be used in educational settings to assist learners and researchers in the interpretation and construction of this specific genre (e.g. Swales 1990, 2004; Bhatia 1993; Swales & Feak 1994, 2000). Thus, the so-called IMRD or IMRAD (Introduction-Methods-Results and Discussion) structure

is considered -and taught- today as the standard form for the writing and publishing of original scientific research articles where each of the sections, in turn, comprises a series of rhetorical movements (moves), meaningful functional and semantic units which help both writer and reader to organize the information following the process of scientific discovery.

Furthermore, the observation that to fully understand what constitutes an appropriate text in any discipline we should consider the intellectual activity which the text is part of (Bazerman 1988: 3) has given rise to a widely followed research line on the disciplinary differences observed within the macrostructure of scientific RAs: chemical engineering (Peng 1987), medicine (Skelton 1994; Nwogu 1997; Li & Ge 2009), computer science (Posterguillo 1999), and biochemistry (Kanoksilapatham 2005), to name only a few.

Most commonly, however, studies on RA disciplinary variation have narrowed the scope of the research to the analysis of specific features within individual sections of the RA (and/or abstract), focusing on a specific field of study. There is now a substantial body of work on aspects of the different RA article sections, including the *abstract* (e.g. Salager-Meyer 1990; Santos 1996; Anderson & Maclean 1997; Hyland 2000; Huckin 2006; Cross & Oppenheim 2006; Ayers 2008), *introduction* (e.g. Swales & Najjar 1987; Dudley-Evans & Henderson 1990; Bhatia 1997; Anthony 1999; Ozturk 2007), *methods* (e.g. Lim 2006; Bruce 2008), *results* (e.g. Thompson 1993; Brett 1994; Williams 1999; Bruce 2009), and *discussion* (e.g. Hopkins & Dudley-Evans 1988; Dubois 1997; Yang & Allison 2003). Finally, numerous contrastive studies have shown differences between disciplines in the construction of a particular section (e.g. Crookes 1986; Fortanet et al. 1998; Holmes 1997; Samraj 2002a, 2002b, 2005).

Undoubtedly, all of these studies have facilitated the work of EAP instructors and contributed to an improvement in the writing and reading practices of scientific scholars, since readers make most of their interpretive decisions about the substance of prose based on clues they receive from its structure.

Some researchers, however, (e.g. Miller 1984; Bazerman 1988, 1994) have warned against explicitly defining and prescriptively teaching generic structures, stressing the importance of genre awareness over genre instruction and emphasizing the dynamic rather than static nature of genres, whose rhetorical structures "can be manipulated according to the conditions of use" (Berkenkotter & Huckin 1993: 477).

The *conditions of use* have been recently modified by the introduction of new technologies that have revolutionized scholarly publishing. Open access peer-reviewed scientific publications have proliferated due to the establishment of organizations committed to making the world's scientific and medical literature freely available (e.g. Public Library of Science, BioMed Central, PubMed Central). As for today, 7,746 journals are included in the Directory of Open Access Journals (DOAJ).

It seems reasonable to suggest that RAs offered exclusively in hypertext may have developed special structural and presentational characteristics as a result of their modular configuration, as some researchers predicted: "the future scientific article will look noticeably different from the current printed version" (Gross et al. 2002: 232).

This paper is an attempt to account for the presence of an overwhelming number of summarizing statements of the research findings which appear scattered through all the sections of RAs on Biology and related Life and Health Sciences, a phenomenon observed in the academic context of EAP teaching in an experimental sciences faculty, which does not correspond to the rhetorical descriptions of this genre, nor to identified distinctive disciplinary variations.

The study has been conducted in the Faculty of Veterinary Sciences of the University of Zaragoza (Spain), where I have been teaching ESP and EAP courses and collaborating with different departments in preparing their research manuscripts for the last 20 years. The difficulties that this reiteration of information entails for non-native writers of RAs in these particular disciplines are also considered.

2. The need for genre flexibility: The case of biology and related life and health sciences

Since the publication of the first real scientific journal in the 17th century, the typical linear, essay-type form of the research article has basically maintained its form and function, progressively introducing changes to accommodate this genre to the changing patterns of communication, the new forms of data handling and the readers' sociocognitive needs and expectations (Kircz 1998; Harter 1998). The understanding of genre as a social phenomenon (Miller 1984; Bazerman 1988, 1994) whereby authors, audiences and critics serve to the communicative purpose explains the necessary balance between the existence of standard models which offer solutions to the recurring rhetorical problems of writing science, and the need to accommodate the discourse to the rhetorical exigencies of the situation. This tension between stability and change lies at the heart of genre use and genre knowledge (Berkenkotter & Huckin 1993: 501) and is more easily perceived by those engaged in disciplinary activity. When genres are examined from the perspective of their users, changes in textual patterns derived from changes in the intellectual content of the field become evident, which suggests that, to fully understand specific genre variations, a close collaboration between the members of the corresponding discourse community and genre analysts is more than advisable.

These considerations find a clear example in the evolution observed in RAs from Biological and related Life and Health Sciences. In a recent study

(Jaime-Sisó 2009), I described the process of convergence of a large number of scientific disciplines under the general umbrella of Developmental Biology that started in the late 1970s, with the result that their corresponding investigations and findings would affect all biological areas simultaneously. The immediate consequence was that the number of publications that could be of interest for scientists working in these fields marked a dramatic rise, making it impossible for them to read even the abstracts of all of these works. This has led to the development of time-saving strategies which parallel those of newspaper readers, searching for the most newsworthy information in a top-down manner. Scientists mostly restrict themselves to searching and reading particular parts of an article, rather than reading front to back (e.g. Bazerman 1985; Huckin 1987; Berkenkotter & Huckin 1995; Hackos & Redish 1998; Hartley 1999).

Accordingly, text features have progressively been adapted to facilitate article selection. Not only titles have become more informative, but whenever possible (up to 80% in highly ranked journals of these areas) they are micro texts summarizing the main results and conclusions (Jaime-Sisó 2009). Main findings also appear in the introduction (Kanoksilapatham 2007: 76), specific results are announced in informative subheadings, often cast as full sentences, and methodology sections are often relegated to the end of the article, usually printed in smaller type. In short, RAs in these areas have incorporated elements of promotional culture, in a process defined as "commodification" (Fairclough 1992, 1993) of institutional discourse.

3. How do scientists select, read and write articles today?

The time elapsed since the publication of the above-mentioned studies on the reading behaviour of scientists and its influence in their writing strategies has brought a new situation which justifies an update of the state of the process. In the second half of the 1990s numerous ethnographic studies showed that, despite the increase in the use of electronic information, scientists still relied heavily on paper journals (e.g. Schriver 1997; O'Hara & Sellen 1997; Tomney & Burton 1998; Sellen & Harper 2001). However, in the last ten years, electronic publishing (and reading) has gradually gained a significant position in the scientific communication process. Most major scientific journals publish electronic versions of print articles on Internet Web sites, and multiple studies report on the progress toward acceptance in the academic community (e.g. De Groote 2001; Liu 2005). As expected, print journals and conferences have been replaced by Web access to digital resources. Parallelly, there is now a growing body of research that examines how professionals with subject expertise use and interact

with online journal literature, which provide interesting data on disciplinary differences observed in the use of online papers (e.g. Rowlands 2007; Hemminger et al. 2007; Olander 2008) and reveal biomedical researchers as pioneers and most active in the use of the internet, with figures showing access to the electronic versions more than ten times as often as the print versions (Morse & Clintworth 2000).

Especially illuminating are the works of information scientists Tenopir and King (2009), who have studied the changes in reading and usage behaviour of scientists for 30 years. Some of their key findings include: (a) scientists rely more and more on electronic library subscriptions at the cost of library print and personal subscriptions, and (b) scientists read more journal articles and book chapters than ever before, but in not much more time (280 articles per year in 2005 against 150 in 1977, only 24 more reading hours yearly, an average of 31 minutes per article, down from 48 in 1977).

While all these works support the conclusion that e-journal design is influenced by scientists' reading and writing processes, they do not offer clues on specific changes which should be considered by article writers to facilitate the reader's finding the important information when searching, browsing and reading tasks are basically made online.

4. Methods

In the design of the methodology, I have followed Swales (1990) and Bhatia's (2004, 2008) proposal of a multi-perspective and multi-dimensional model for the analysis of discourse as genre, where social, professional and textual spaces are considered. This allows for the integration of ethnographic and corpus-based approaches into a single framework, as in Hyland (2000, 2004) and Connor (2000), where analyses of interviews were included to better understand the broader social contexts of the discourse.

Accordingly, the present research was developed in two phases:

1. A series of interviews was programmed in order to verify to what extent scientists' reading behaviour is determined by the use of electronic journals. Interviews were preferred to questionnaires because direct contact with the researchers allows for the *in situ* inclusion of new questions to clarify aspects of their answers. As in previous works on scientists' composition and reading behaviour (Bazerman 1985; Rymer 1988; Myers 1990; Berkenkotter & Huckin 1995, chap. 2), this approach involves a limitation in the number of consultants, which in the present study is restricted to 10.

2. A small corpus (20) of RAs (*vid*. Appendix) in the selected areas was manually analyzed to detect whether online journals offer specific features that may have contributed to the increase in scientists' reading efficiency in the last decade.

4.1 Interviews with consultants: Procedure

One-hour individual interviews were conducted with ten faculty colleagues from the Departments of Biochemistry, Pathology, and Genetics and Embryology, all of them active senior researchers publishing internationally. The procedures established in Berkenkotter and Huckin's (1995:30) study were followed so the interviews took place in a relaxing atmosphere in the consultants' offices. They were informed that the purpose was to learn about their professional reading and writing habits. The introductory question was simple: 'How do you select journal articles, online or on paper?' They were then asked to describe the steps they follow in their search, selection and reading. In most cases, a detailed account of the procedure followed was substituted by a "hands on" demonstration on the computer.

4.2 Results of the interviews

All consultants informed me that they make the initial selection of RAs online. Four of them even admitted that they had not actually read, nor even seen, a *paper* article in months. From their descriptions I could infer that apparently their behaviour is identical to that shown by the seven scientists in Berkenkotter and Huckin's study (1995: 30). The ten consultants of the present work first scan the table contents, then the names of authors and the title. In most cases the reading stops there. When the article holds a certain interest for them, they read the abstract and/or have a look at tables, figures and legends. If these contain data related to their own research, they jump to the results section. If the results are not convincing, they turn to the discussion. Only one declared reading the introduction, but only the last paragraph. They hardly ever read the methodology section and the supporting information - a section situated at the end of the article, recently incorporated to online journals in Biology and Life Sciences where authors can include additional information to further clarify different aspects of their research by way of dataset, figures, tables, stereograms, multimedia files-. Finally, if the topic, although interesting, is only transversally related to their own research, they concentrate on the summary or even read more in detail the adapted parallel texts that online publications offer at a more adequate level for less informed readers.

The main difference observed in the reading attitude and habits of the researchers interviewed in the present work when compared to those described in Berkenkotter and Huckin's study lies in an apparently minor detail. In their description of

the interviews with the consultants, Berkenkotter and Huckin (1995: 2) remarked that "their desks contained piles of journals, some still unopened". In my experience, however, journal papers do not invade the scientists' territories. The 10 scientists I interviewed either scan on the screen the parts they have selected (online versions offer links to the different sections separately) or download and/or print the section(s) of their interest. As a consequence, they hardly ever read the complete article, following the discourse linearly from beginning to end. When asked about the possibility of missing important information as a result of this selective, non-linear reading, they all agreed that this is impossible, since "the main findings are *advertised* constantly throughout the article". They also declared being aware of the demand for reiteration when describing their own research, although they could not offer an explanation for this implied or tacit agreement. While they admitted having difficulties in finding "different ways to offer the same information in English", they insisted on mimicking the repetition strategy they have observed in already published papers, which forces them to find linguistic alternatives to present the relevant findings of their own research repeatedly but differently, while maintaining the lack of ambiguity which is inherent to scientific discourse.

In short, the information obtained through the interviews revealed an outstanding feature of RAs in biomedical areas: the constant repetition of the findings through all the sections of the paper. In this respect, the inclusion of an ethnographic approach in the process of the study proved to be extremely useful as it allowed me to observe the genre under study in action. The insightful experiences of members of the disciplinary community of practice showed me to what extent disciplinary conventions that govern the use of language are "most often implicitly understood and unconsciously followed by the participants taking part in the communicative situation in which the genre in question is used" (Bhatia 2004: 166). As suggested in Bhatia's (2004) description of a four space-model (textual, tactical, professional, social) for analyzing written discourse, these interacting views are not mutually exclusive, but essentially complementary to each other. Thus, the second part of the study - which involved textual analysis - was undertaken in order to verify the real scope of the presence of repetitive statements summarizing the results of the research.

4.3 Corpus selection and approach to the analysis

Several factors were considered in the selection of study material

– The articles should be published in journals covering a wide range of topics, which could be of interest for researchers working in different subareas within the general fields of Biology and Life Sciences.

- The selected journals should be indexed in the Journal Citation Reports and occupy a high post in the corresponding Impact Factor List.
- The articles should be accessible online.

Accessibility online is possible either via open access journals or through subscriptions that offer pre-print versions of the paper articles some months before publication. Both alternatives were considered to evaluate possible differences.

The consultants were asked for advice in the selection. The coincidental choice among consultants was *Plos Biology*. This publication is an open access peer-reviewed prestigious journal (n.1 of a total of 76 in the impact list of Biology and 8 out of 283 in Biochemistry and Molecular Biology) that features works of exceptional importance in all areas of biological science, including works at the interface with other disciplines, such as chemistry and medicine. A paper version of the journal is not available, which ensures that scanning, browsing and selection of information is made online.

As for the journal published both online and in print, *Biology of Reproduction* was selected, because it was frequently consulted by the informants as the most highly cited journal in the field of Reproductive Biology (n.5 out of 26 in the Impact List).

Ten articles published between January 2007 and December 2008 were randomly selected from each journal. Due to the international composition of most scientific teams, the origin of the authors was considered as irrelevant.

A three-step analysis was designed to detect the presence of summarizing statements of the findings throughout the article:

1. Initial reading was made online mimicking the scientist's behaviour as described in the interviews. Two hypotheses were considered: (a) the article contains relevant information for the reader and (b) the study is only transversally related to the reader's interests. This double approach to reading and the objective established required that all elements of the article accessible under the same reference be considered sections susceptible to containing the information searched. Based on this distinction, two reading paths were followed, reproducing the steps observed in the scientists for each situation:
   - Title → abstract → results → discussion and
   - Title → summary → adapted texts (where available).

   All these sections were carefully read and analyzed as independent cognitive modules.
2. A linear front to back traditional reading of each whole article was made covering the sections which had been excluded in the previous approach. Building on previous studies that have identified move types in each RA section, I tried to detect those statements that could be indicative of main findings

presentation. The detailed structural analysis of biochemistry RA moves and steps offered by Kanoksilapatham (2007) was used as a leading axis. Alternative moves detected in similar studies on RA and Abstracts (see references in Introduction) were also considered.

3. All the statements which summarized the results and conclusions of the research were extracted and contrasted to verify their equivalence in terms of amount and level of information. Due to the high level of specialization required to understand the scientific concepts expressed, advice was sought from the consultants to test the adequacy of the equivalences. Statements which referred to partial results were also considered when according to my informants' opinion, implied information could be easily inferred by a reader specialist in the field. Examples of this are offered below:
    1. a. *Protein Aggregation Propensity and Protein Instability are Synergistic Risk Factors for fAL*
       (partial result in article n.8 *Plos Biology* corpus).
       b. *Protein Aggregation and Protein Instability Govern Familial Amyotrophic Lateral Sclerosis Patient Survival*
       (title summarizing findings of the same article)
    2. a. *Mice Lacking c-Secretase or All Notch Proteins in the Skin Develop Severe B-LPD after Birth*
       (partial result in article n. 7 *Plos Biology* corpus)
       b. *RBP-j – Independent (but c-Secretase – Dependent) Notch Signaling in the Skin Contributes to Longevity*
       (partial result in article n. 7 *Plos Biology* corpus)
       c. *Notch-Deficient Skin Induces a Lethal Systemic B-Lymphoproliferative Disorder by Secreting TSLP, a Sentinel for Epidermal Integrity*
       (title summarizing findings of the same article)

5. Results and interpretations of findings

5.1 Plos biology

Traditionally, biomedical articles follow the requirements and conventions established by the International Committee of Medical Journal Editors (ICMJE). The intention of the ICMJE is to help authors and editors in their mutual task of creating and distributing accurate, clear, easily accessible reports of biomedical studies by providing ethical principles in the conduct and reporting of research, and recommendations relating to specific elements of editing and writing. In this respect, ICMJE (2009), recommends the Introduction – Methods – Results – Discussion

(IMRD) model, and provides under each of these headings information on how to structure content. According to these requirements, the body of the article is normally preceded by an abstract or summary.

The analysis of *Plos Biology* articles, however, revealed some basic differences between the traditional and extensively followed model and that adopted by this prestigious journal. The first immediate finding was the elevated number of new sections that this open access publication offers. Apart from the traditional sections, articles in *Plos Biology* include:

- a brief running head, which appears as a heading on each page of the paper.
- a non-technical summary of the work - distinct from the scientific abstract - to make findings accessible to an audience of both scientists and non-scientists, and
- a separate brief statement ('blurb') to be displayed on the web site below the title. The blurb should be about 20 to 30 words long and should entice people to read the manuscript. It should not be redundant with the title and concentrate on the news value of the research findings.
- Some articles are accompanied by a synopsis written for a general audience to provide non-experts with insight into the significance of the published work. A "primer" accompanies other articles. Primers provide a concise introduction into an important aspect of biology highlighted in the corresponding research article. Although both primers and synopses are usually written by other authors, these sections have also been included in the analysis because they are integral parts of the scientists' reading strategies, as revealed in the interviews.

A change in the traditional organizational features of experimental research articles was the second important observation derived from the analysis: the Materials and Methods section appears after the Discussion, followed by the Supporting Information section, both offered in a smaller type. This is in line with previous research (e.g. Bazerman 1988; Berkenkotter & Huckin 1995) with methods taking a secondary position. Once again my informants were consulted about the possible reasons for this relegation. They declared that due to the constant and deep technological advancements in the field of science both methods and materials have become too specialized to be understood by the general scientific reader, so a description of the methodology used in a research work is relevant only for those scientists working on a very closely related experimental study and, in these cases, the detailed information needed is so extensive that must be complemented with visual material provided in the Supporting Information section.

Another distinctive feature of electronic articles is the possibility of using links to access every individual section, which allows the reader to collect only those parts of the article with particular value for his/her specific interests and download/

print them separately, as entities in their own, dislocated from the article as a whole. This "loss of contact" between the component sections of the paper may explain the need for repeatedly highlighting the overall research outcome in all the sections to ensure that the main message reaches every reader.

The textual analysis carried out reveals an average of 15 statements summarizing the results per article. With the exception of Materials and Methods, each section included at least one statement condensing the main findings of the research, as Table 1 shows.

These elevated numbers exceed all expectations based on previous studies on moves within RAs. The new presentational features have increased the possibilities of anticipating the findings, but authors seem to believe that this is a compulsory move that should be included under every heading. Titles and running heads are not restricted to announcing the topic of the research; they also synthesize the findings. Blurbs, abstracts and author summaries logically comply with this obligatory move, offering occasionally several reformulated versions of the results. All introductions provide a statement of the findings in the last paragraph, but quite often previous paragraphs contain similar sentences with practically identical information. Presentation of results in the Results section is emphasized by way of headings that highlight the main findings. As expected, according to previous descriptions of Discussion moves (e.g. Hopkins & Dudley Evans 1988: 18; Holmes 1997: 324; Kanoksilapatham 2005: 76–77), this section includes at least one statement of selected findings followed by a generalization which summarizes the claims made in the study. The interesting aspect of this reiteration is that authors

Table 1. Presence of statements summarizing the findings in *Plos Biology* articles*

	T	RH	B	AS	Ab	I	Re	D/C	P/S	Total
(1)	1	1	1	1	1	1	3	3	1	13
(2)	1	1	1	2	1	3	2	1	3	15
(3)	1	0	1	2	2	1	4	2	2	15
(4)	1	0	1	1	1	2	2	3	1	12
(5)	1	1	1	1	1	1	1	1	3	11
(6)	1	1	1	2	2	2	2	4	1	16
(7)	1	1	1	2	2	2	10	5	1	25
(8)	1	1	1	1	1	1	9	1	2	18
(9)	1	1	1	1	1	1	3	3	-	12
(10)	1	1	1	3	2	1	1	2	2	14

*Numbers in brackets on the left identify articles as listed in Appendix. T: Title; RH: Running Head; B: Blurb; A.S: Author's Summary; Ab: Abstract; I: Introduction; R: Results; D/C: Discussion or Conclusion; P/S Primer or Synopsis. Numbers below abbreviations refer to occurrences in the corresponding article sections. (-) means the section is not available.

try to avoid syntactic and semantic repetitions while ensuring homogeneity and equivalence regarding scientific information. These paraphrases, however, do not allow for a "fuzziness of registers" (Hatim 1990: 51), a common feature in mass media discourse and popularizations understood as "translations of scientific research for a lay audience" (Paul 2004: 32). Only some of the titles of primers and synopses make concessions that could be deemed alien to the social purposes of experimental research articles, using eye-catching statements apparently addressed to attracting the interest of the general reader by exploiting the conventions of simple, informal and conversational speech, which may indicate a kind of "democratization" of the scientific community, as in the following examples:

1. *Who Needs Sex (or Males) Anyway?* (article 2)
2. *Male or female: The answer depends on when you ask* (article 4)
3. *Understanding the Web of Life: The Birds, the Bees, and Sex with Aliens* (article 6)

However, even in these cases, my informants admit that the underlying message is actually addressed to a specialist reader, since a lay audience would not be able to infer the information that these attention-attracting statements contain.

Unfortunately, length limitation of the present study does not allow for a detailed linguistic analysis of the statements, but an example is offered (Table 2) to illustrate the efforts made by the authors to present almost identical information in different ways. Statements have been broken up to facilitate the presentation of parallelisms between equivalent elements and are preceded by indication of the section where they appear and the number of words in each statement.

The news value of this article, as confirmed by the consultants, is the demonstration that a group of asexual animals (*Bdelloid Rotifers*) has diversified into distinct species, a property commonly thought to be restricted to interbreeding individuals.

As Table 2 shows, this message is conveyed in 15 different fashions, subtly adapting the sentence construction to the objectives of each section. The title and running head follow the typical journalistic fashion increasingly found in scientific articles, placing the finding as subject of the sentence. Only in the Results section the summarizing statement presents this same order. In all other cases it is the subject under study that occupies the prominent position. Statements in the purely presentational sections contain a complementary phrase defining the object of study, which is presented as either the subject of the sentence or an appositional phrase:

– "*a famous asexual group*" (blurb)
– "*a classical asexual clade*" (abstract)
– "*a famous group of asexual animals*" (author's summary)

Table 2. Summarizing statements in the article "Independent Evolving Species in Asexual Rotifers". *Plos Biology* 5 (4) 914-921. Roman numerals identify elements with equivalent information

Location	Words	Statements
RUNNING HEAD	4	Speciation (I) in Asexual (II) Rotifers (III).
TITLE	7	Independently Evolving Species (I) in Asexual (II) Bdelloid Rotifers (III).
BLURB	31	New analysis in genetic and morphological data reveal that (IV) the bdelloid rotifers (III), a famous asexual group (II) have diversified into (V) distinct species (I) and that sex is not a necessary condition for speciation (VI).
ABSTRACT	30	Here we show that (IV) a classical asexual clade (II), the bdelloid rotifers (III), has diversified into (V) distinct evolutionary species (I). This would challenge the view that sex is necessary for diversification into species (VI).
AUTHOR'S SUMMARY	25	We show that (IV) a famous group of asexual animals (II), the bdelloid rotifers (III), has diversified into (V) distinct species (I) broadly equivalent to those found in sexual groups (VII).
AUTHOR'S SUMMARY	12	The results show that (IV) sex is not a necessary condition for speciation (VI)
INTRODUCTION 3rd paragraph	23	Here we demonstrate that (IV) a classical asexual clade (II), the bdelloid rotifers (III), has diversified into (V) independently evolving and distinct entities arguably equivalent to species (I).
INTRODUCTION 7th paragraph	30	Our results demonstrate that (IV) bdelloids (III) have diversified into (V) not only genetic clusters, indicative of independent evolution (I), but also into entities experiencing divergent selection on feeding morphology, indicative of niche divergence (I).
INTRODUCTION last paragraph	20	In short (IV), bdelloids (III) have diversified into (V) entities equivalent to sexual species in all respects (I) except that individuals do not interbreed (II).
RESULTS 4th paragraph	19	Our results indicate that (IV) independently evolving species (I) are present in bdelloids (III) but at a lower level than taxonomic species (VII).
RESULTS 6th paragraph	22	Our analyses show that (IV) this traditional species, which are found living together on single louse individuals (III), are evolutionary independent and distinct species (I).
DISCUSSION	39	We conclude that (IV) bdelloids (III) display the same qualitative pattern of genetic and morphological clusters indicative of diversification into (V) independently evolving species (I), as found in sexual clades (VII). This refutes the idea that sex is necessary for diversification into evolutionary species (VI).
TITLE PRIMER	6	Who Needs Sex (or Males) Anyway?
BODY PRIMER 3rd paragraph	21	The researchers show that (IV), despite an ancient aversion for interbreeding (II), bdelloids (III) display (V) evolutionary patterns (I) similar to those in sexually reproducing taxa (VII).
BODY PRIMER 8th paragraph	13	The researchers show that (IV) this distinct monophyletic genetic clusters (III) represent (V) independently evolving species (I).

and are followed by reformulated versions of the conclusion:

- "*sex is not a necessary condition for speciation*" (blurb and author's summary)
- "[*This would challenge the view that*] *sex is necessary for diversification into species*" (abstract)
- "[*This would refute the idea that*] *sex is necessary for diversification into evolutionary species*" (discussion)

The categorical claims made in the title, abstract and blurb are diminished in other sections introducing information which weakens the assertions:

[*bdelloids*] have diversified into:

- "*independently evolving species*" (title)
- "*distinct species*" (blurb)
- "*distinct evolutionary species*" (abstract)
- "*species broadly equivalent to those found in sexual groups*" (author's summary)
- "*independently evolving and distinct entities arguably equivalent to species*" (introduction)
- "*entities equivalent to sexual species* (...) *except that individuals do not interbreed*" (introduction)
- "*independently evolving species* (...) *but at a lower level than taxonomic species*" (results)
- "*evolutionary patterns similar to those seen in sexually producing taxa*" (primer)

The writer's ability to avoid repetitions is also shown in the variety of metatextual elements preceding the summarizing statements. (*vid*. Table 2)

The most important finding, however, is that all 15 statements, although apparently different in terms of amount of information (number of words ranges from 4 to 39) contain the necessary elements to enable the informed reader to infer the conclusions of the research, according to my specialist informants, who were consulted to help interpret the results. The highly specialized concepts and terminology used in the study material made it necessary to seek for the collaboration of expert members of this discourse community, which once more demonstrated the need for integrating professional expertise in the analysis of specialized discourse (Bhatia 2004)

5.2 Biology of reproduction

The electronic version of this journal is practically identical to the existing paper version, showing the traditional IMRD structure where the article sections are offered in the usual order (abstract, introduction, materials and methods, results, discussion). The only difference is that articles in the online format incorporate a

one-sentence summary of the manuscript's significance, limited to 250 characters, which was also considered in the analysis.

The fact that the online article did not differ structurally from the paper version suggested that the presentation of the information would coincide with that found in printed articles, whose readers are expected to follow the discourse linearly, beginning at the top of the first page and following the line of the article until the bottom of the final page. The three-step analysis described in the previous section of the present study was carried out to contrast the findings with those obtained in the analysis of *Plos Biology* articles. The results are shown in Table 3.

While the much lower occurrence of summarizing statements can be partly explained by the absence of new presentational sections, careful reading and analysis shows that this is also due to a more conservative approach to the inner structure (moves) of the traditional sections, which is highly coincidental with findings presented in previous studies on rhetorical structure (Kanoksilapatham 2007). Thus, unlike those in *Plos Biology*, most introductions lack specific reference to findings and when offered, appear only in the last paragraph. In fact, Guidelines for Authors explicitly warn against including a summary or discussion in the Introduction that is redundant to other sections of the manuscript. The Methodology section is found after the Introduction, thus following the IMRD macrostructural model which parallels the steps given in any research process. Presentation of specific results in the Results section is not usually followed by generalizing statements that could help the reader to infer the conclusions. Exceptions to this are

Table 3. Presence of statements summarizing the findings in *Biology of Reproduction* articles*

	T	WS	Ab	I	Re	D	Total
(1)	1	1	1	0	1	2	6
(2)	1	1	2	0	0	2	6
(3)	1	1	1	1	0	2	6
(4)	1	1	1	0	0	4	7
(5)	1	1	1	0	0	1	4
(6)	1	1	1	0	0	2	5
(7)	1	1	1	1	3	2	9
(8)	0	1	1	0	0	1	3
(9)	1	1	1	1	1	3	8
(10)	1	1	2	1	1	2	8

*Numbers in brackets on the left identify articles as listed in Appendix. T: Title; WS: Web Summary; Ab: Abstract; I: Introduction; R: Results; D: Discussion. Numbers below abbreviations refer to occurrences in the corresponding article sections.

articles 1, 7, 9 and 10 where the Results section includes headings highlighting the main findings of the study. In the Discussion, the compulsory move *Consolidating results* (Kanoksilapatham 2007: 82) is used to further explain the results announced in the Results section and to make overt claims and generalizations. This produces an average of 6 occasions per article where findings are summarized, a figure that is more coincidental with previous descriptions of RAs structure and functions, which are usually based on the analysis of paper articles where "unnecessary" repetitions are avoided and each section has its distinctive moves corresponding to the logical sequence followed by scientists in the development of their research work.

The different approaches to reading described in the methodology showed that informed readers know where to find the precise pertinent information they need, skipping those parts of the article they consider irrelevant or unnecessary for the stage of the research process in which they are when consulting the article. When readers approach the text with the only objective of learning how a particular paper fits within the broader spectrum of his or her own research, or to get an idea from fields they have only a basic knowledge, they simply need a general statement of the objective of the study and the author's claims. In both reading typologies users approach the article only partially, browsing the text with a particular goal in mind. Thus, as Berkenkotter and Huckin (1995: 32) observed, a discrepancy exists between this specialist, selective reading schema and the writer-based text schema where the reader has to search for information instead of reading top-down.

When comparing the results obtained in the analysis of *Plos Biology* articles with those published in *Biology of Reproduction* we observe that the former, open access publication has modified the genre conventions to better accommodate the needs of specialist readers. The new presentational sections synthesize the main message of the text (the important findings), thus facilitating a top-down approach to reading. Furthermore, the appearance of statements of results in all sections ensures the writer the promotion of the article's news value, since any reader, not even the more superficial or less informed can escape receiving the message that the writer wants to publicize.

As shown above (Table 3), this reiteration is not so abundant in *Biology of Reproduction* articles. Only one new presentational section (Web Summary) promoting the results of the research is included in the online version, and the moves containing information directly related to the main findings are only situated in the corresponding parts of the sections as established in the IMRD framework. Consequently, although the *reader* has sufficient clues to find the information they require, the *writer* runs the risk of not reaching every reader with the information he or she wants to highlight. The fact that in all cases the titles of the articles analyzed convey the news value of the research does not guarantee reception of the

author's main message unless the article is read in paper format or the reader decides to print and read the whole text. In conversations with my consultants I learned that in many cases scientists approach a research paper through a link in any section of another thematically related article. This direct connection between sections of different papers greatly facilitates the scientists' search for specific information but endangers the objective of research promotion for writers whose articles have been accessed in sections where findings have not been highlighted.

Thus, considering that scientists prefer the electronic environment in their search for scientific information (Tenopir & King 2009), the results of this analysis show that articles written for journals published exclusively online present organizational and informational features much better adapted to both readers and writers' needs than those which are mere electronic clones of print journals, which indicates different stages in the evolution towards a new, inherently modular model in the transmission of scientific information.

6. Conclusion

Recent years have seen a dramatic shift from a largely print-based towards electronic communication of scientific scholarly information. While much of this was announced during the computing revolution of the 1980s and 1990s, the recent widespread adoption of Web-based electronic journals has been the primary driver for change.

The standard IMRD division, which reflects the commonly accepted procedure for a complete linear presentation of a finalized research (Kircz 1998; Sollaci & Pereira 2004), has been shaped in the tradition of print publishing. Linear reading, however, is not the typical approach used by scientists in their field of expertise, as observed in previous studies (Bazerman 1985; Berkenkotter & Huckin 1995) and revealed in the interviews reported here.

The results obtained in the present study suggest that the structure of scientific RAs is progressively being adapted to online users' needs. Aware of the scientists' reading practices, both editors and writers contribute to ensure that, whatever section of the text is scanned, and regardless of the reasons for approaching the article, the reader obtains the most newsworthy information, as if each of the sections could stand alone. This is particularly so in open access publications that do not offer an alternative print product, where proliferation of presentational sections facilitates quick and to the point reading, and traditional sections always include statements of the main findings.

This new form of scientific information presentation could be described as "a coherent set of linked modules [...] each being in themselves (small) texts

emphasizing aspects of the message that together establish a complete message from author to reader" (Kircz 1998: 217).

I would suggest that repetitions of findings are the consequence of a transitional period where the scientific article is still considered as a unitary genre while their users hardly ever read it as such. Writers have to combine observation of the generic conventions of the standard, linear type of research articles with an ability to meet the needs of busy, highly selective online readers. They seem to have found the solution to this problem by repeating their findings in whatever part of the article the reader may have direct access and, at the same time, paraphrasing the statements in attention to the editors and those who prefer a sequential reading of the entire paper. As a result of this, scientists (especially non-native users of English) are faced with the difficult task of finding new ways to say the same things while maintaining the clarity of the message and avoiding informal or ambiguous language.

Finally, I would also like to suggest that this phenomenon be taken into account in the design of scientific academic English courses. This would involve the inclusion of summarizing strategies and practice in rephrasing techniques that contemplate syntactic conversions and semantic equivalences through the use of context-dependent synonyms, hypernyms and hyponyms. As the variety of possible readers approaching an open access article may range from the highly specialized researcher to the occasional browser, practice on certain rhetorical strategies, such as synecdoche, metaphor or metonymy could also be taught as helpful tools to attractively reformulate the reiterative statements if they are cautiously used to preserve the univocal message that science should convey.

Acknowledgements

I am indebted to my colleagues, the research scientists Dr. Pedro Muniesa and Dr. Jesús de la Osada for their cooperation, comments and guidance in the present report.

References

Anderson, K. & Maclean, J. 1997. A genre analysis study of 80 medical abstracts. *Edinburgh Working Papers in Applied Linguistics* 8: 1–23.

Anthony, L. 1999. Writing research article introductions in software engineering: How accurate is a standard model? *IEEE Transactions on Professional Communication* 42: 38–46.

Ayers, G. 2008. The evolutionary nature of genre: An investigation of the short texts accompanying research articles in the scientific journal Nature. *English for Specific Purposes* 27(1): 22–41.

Bazerman, C. 1985. Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written Communication* 2: 3–23.

Bazerman, C. 1988. *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison WI: University of Wisconsin Press.

Bazerman, C. 1994. Systems of genres and the enactment of social intentions. In *Genre and the New Rhetoric,* A. Freedman & P. Medway (eds), 79–101. London: Taylor and Francis.

Berkenkotter, C. & Huckin, T.N. 1993. Rethinking genre from a sociocognitive perspective. *Written Communication* 10: 475–509.

Berkenkotter, C. & Huckin, T.N. 1995. *Genre Knowledge in Disciplinary Communication: Cognition, Culture, Power*. Hillsdale NJ: Lawrence Erlbaum Associates.

Bhatia, V. 1993. *Analysing Genre: Language Use in Professional Settings*. London: Longman.

Bhatia, V. 1997. Genre mixing in academic introductions. *English for Specific Purposes* 16: 181–195.

Bhatia, V. 2004. *Worlds of Written Discourse: A Genre-based View*. London: Continuum.

Bhatia, V. 2008. Towards critical genre analysis. In *Advances in Discourse Studies*, V. Bhatia, J. Flowerdew & R.H. Jones (eds), 166–177. London: Routledge.

Brett, P. 1994. A genre analysis of the results section of sociology articles. *English for Specific Purposes* 13: 47–59.

Bruce, I. 2008. Cognitive genre structures in methods sections of research articles: A corpus study. *Journal of English for Academic Purposes* 7: 38–54.

Bruce, I. 2009. Results sections in sociology and organic chemistry articles: A genre analysis. *English for Specific Purposes* 28(2): 105–124.

Connor, U. 2000. Variation in rhetorical moves in grant proposals of US humanists and scientists. *Text* 20(1): 1–28.

Crookes, G. 1986. Towards a validated analysis of scientific text structure. *Applied Linguistics* 7: 57–70.

Cross, C. & Oppenheim, C. 2006. A genre analysis of scientific abstracts. *Journal of Documentation* 62(4): 428–446.

De Groote, S.L. & Dorsch, J.L. 2001. Online journals: Impact on print journal usage. *Journal of the Medical Library Association* 89(4): 372–8.

Dubois, B.L. 1997. The biomedical discussion section in context. *Advances in Discourse Processes Series*. Greenwich CT: Ablex.

Dudley-Evans, T. & Henderson, W. 1990. The organisation of article introductions: Evidence of change in economics writing. *ELT Documents* 134: 67–78. London: Modern English Publications and The British Council.

Fairclough, N. 1992. *Discourse and Social Change*. London: Polity Press.

Fairclough, N. 1993. Critical discourse analysis and the marketization of public discourse: The universities. *Discourse and Society* 4(2): 133–168.

Fortanet, I., Posteguillo, S., Palmer, J.C. & Coll, J.F. 1998. Disciplinary variations in the writing of research articles in English. In *Genre Studies in English for Academic Purposes*, I. Fortanet, S. Posteguillo, J.C. Palmer & J.F. Coll (eds), Filología 9: 59–78. Castellón de la Plana: Universitat Jaume I, Summa.

Gross, A.G., Harmon, J. & Reidy, M. 2002. *Communicating Science: The Scientific Article from the Seventeenth Century to the Present.* New York NY: Open University Press.

Hackos, J. & Redish, J. 1998. *User and Task Analysis for Interface Design*. New York NY: Wiley.

Harter, S.P. 1998. Scholarly communication and electronic journals: An impact study. *Journal of the American Society for Information Science* 49(6): 507–516.

Hartley, J. 1999. From structured abstracts to structured articles: A modest proposal. *Journal of Technical Writing and Communication* 29: 255–270.

Hatim, B. & Mason, I. 1990. *Discourse and the Translator*. London: Longman.

Hemminger, B., Lu, D., Vaughan, K.T. & Adams, S.J. 2007. Information seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology* 58(14): 2205–2225.

Holmes, R. 1997. Genre analysis and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes* 16: 321–337.

Hopkins, A. & Dudley-Evans, T. 1988. A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes* 7: 113–121.

Huckin, T. 1987. Surprise value in scientific discourse. Paper presented at the 38th Annual Meeting of the Conference on College Composition and Communication, Atlanta, GA, March 19–21.

Huckin, T. 2006. Abstracting from abstracts. In *Academic Writing in Context*, M. Hewings (ed.) 93–103. London: Continuum.

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

Hyland, K. 2004. Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing* 13(2): 133–151.

International Committee of Medical Journal Editors (ICMJE). 2009. Uniform requirements for manuscripts submitted to biomedical journals. <http://www.icmje.org/>

Jaime-Sisó, M. 2009. Titles or headlines? Anticipating conclusions in biomedical RA titles as a persuasive journalistic strategy to attract busy readers. *Miscelánea: A Journal of English and American Studies.* 39: 29–54.

Kanoksilapatham, B. 2005. Rhetorical structure of biochemistry research articles. *English for Specific Purposes* 24(3): 269–292.

Kanoksilapatham, B. 2007. Rhetorical moves in biochemistry research articles. In *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure* [Studies in Corpus Linguistics 28], D. Biber, U. Connor & T.A. Upton (eds), 73–119. Amsterdam: John Benjamins.

Kircz, J.G. 1998. Modularity: The next form of scientific information. *Journal of Documentation* 54(2): 210–235.

Li, J.L. & Ge, G.C. 2009. Genre analysis: Structural and linguistic evolution of the English-medium medical research article (1985–2004). *English for Specific Purposes* 28: 93–104.

Lim, J.M.H. 2006. Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes* 25(3): 282–309.

Liu, Z. 2005. Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation* 61(6): 700–712.

Miller, C.R. 1984. Genre as social action. *Quarterly Journal of Speech* 70: 151–67.

Morse, D. & Clintworth, W.A. 2000. Comparing patterns of print and electronic journals use in a health science library. *Science and Technology Librarianship*. <http://www.istl.org/00-fall/refereed.html>

Myers, G. 1990. *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. Madison WI: University of Wisconsin Press.

Nwogu, K. 1997. The medical research paper: Structure and functions. *English for Specific Purposes* 16(2): 119–138.

O'Hara, K. & Sellen, A. 1997. A comparison of reading paper and on-line documents. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI 1997,335–342. New York NY: ACM Press.

Olander, B. 2008. Scholarly communication in transition: Computer scientists' information behaviour over twenty years. *Information Research* 13(4) paper 372.

Ozturk, I. 2007. The textual organization of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes* 26: 25–38.

Paul, D. 2004. Spreading chaos: The role of popularizations in the diffusion of scientific ideas. *Written Communication* 21(1): 32–68.

Peng, J. 1987. Organizational features in chemical engineering research articles. *ELR Journal* 1: 79–116.

Posteguillo, S. 1999. The schematic structure of computer science research articles. *English for Specific Purposes* 18(2): 139–160.

Rowlands, I. 2007. Electronic journals and user behavior: A review of recent research. *Library and Information Science Research* 29(3): 369–396.

Rymer, J. 1988. The scientific composing process: How eminent scientists write articles. In *Writing in the Academic Disciplines*, D. Jolliffe (ed.), 211–250. Norwood NJ: Ablex.

Salager-Meyer, F. 1990. Discoursal flaws in medical English abstracts: A genre analysis per research- and text-type. *Text* 10(4): 365–384.

Samraj, B. 2002a. Introductions in research articles: Variations across disciplines. *English for Specific Purposes* 21: 1–17.

Samraj, B. 2002b. Disciplinary variation in abstracts: The case of wildlife behaviour and conservation biology. In *Academic Discourse*, J. Flowerdew (ed.), 40–56. Harlow: Longman.

Samraj, B. 2005. An exploration of genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes* 24: 141–156.

Santos, M.P. 1996. The textual organization of research paper abstracts in applied linguistics. *Text* 16(4): 481–499.

Schriver, K. 1997. *Dynamics in Document Design*. New York NY: Wiley.

Sellen, A. & Harper, R. 2001. *The Myth of the Paperless Office*. Cambridge MA: The MIT Press.

Skelton, J. 1994. Analysis of the structure of original research papers: An aid to writing original papers for publication. *British Journal of General Practice* 44: 455–459.

Sollaci, L.B. & Pereira, M.G. 2004. The introduction, methods, results, and discussion (IMRAD) structure: A fifty-year survey. *Journal of the Medical Library Association* 92: 364–371.

Swales, J. & Najjar, H. 1987. The writing of research article introductions. *Written Communication* 4: 175–192.

Swales, J. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: CUP.

Swales, J. & Feak, C.B. 1994. *Academic Writing for Graduate Students*. Ann Arbor MI: University of Michigan Press.

Swales, J. & Feak, C.B. 2000. *English in Today's Research World: A writing guide*. Ann Arbor MI: University of Michigan Press.

Swales, J. 2004. *Research Genres: Exploration ad applications*. Cambridge: CUP.

Tenopir, C., King, D.W. & Edwards, S. 2009. Electronic journals and changes in scholarly article seeking and reading patterns. *Aslib Proceeding: New Information Perspectives* 61(1): 5–32.

Thompson, D.K. 1993. Arguing for experimental facts in science: A study of research article results sections in biochemistry. *Written Communication* 8: 106–128.

Tomney, H. & Burton, P.F. 1998. Electronic journals: A study of usage and attitudes among academics. *Journal of Information Science* 24(6): 419–429.

Williams, I.A. 1999. Results sections in medical research articles: Analysis of rhetorical categories for pedagogical purposes. *English for Specific Purposes* 18: 347–366.
Yang, R. & Allison, D. 2003. Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes* 22: 365–38.

Appendix: Articles in the corpus

Plos Biology

1. FMRP Mediates mGluR5-Dependent Translation of Amyloid Precursor Protein. Vol. 5(3) March 2007.
2. Independently Evolving Species in Asexual Bdelloid Rotifers. Vol. 5(4) April 2007.
3. Thermal Stress and Coral Cover as Drivers of Coral Disease Outbreaks. Vol. 5(6) June 2007.
4. Indirect Effects of Ploidy Suggest X Chromosome Dose, Not the X: A Ratio, Signals Sex in Drosophila. Vol. 5(4) December 2007.
5. Solid-State, Dye-Labeled DNA Detects Volatile Compounds in the Vapor Phase. Vol. 6(1) January 2008.
6. Invasive Mutualists Erode Native Pollination Webs Vol. 6(2) February 2008.
7. Notch-Deficient Skin Induces a Lethal Systemic B-Lymphoproliferative Disorder by Secreting TSLP, a Sentinel for Epidermal Integrity. Vol. 6(5) May 2008.
8. Protein Aggregation and Protein Instability Govern Familial Amyotrophic Lateral Sclerosis Patient Survival. Vol. 6(7) July 2008.
9. Rice XB15, a Protein Phosphatase 2C, Negatively Regulates Cell Death and XA21-Mediated Innate Immunity. Vol. 6(9) September 2008.
10. Brain IGF–1 Receptors Control Mammalian Growth and Lifespan through a Neuroendocrine Mechanism Vol. 6(10) October 2008.

Biology of Reproduction

1. Functional HY-Specific CD8 T Cells Are Found in a High Proportion of Women Following Pregnancy with a Male Fetus. Vol. 76: 96–101. January 2007.
2. CYP2E1-Catalyzed Oxidation Contributes to the Sperm Toxicity of 1-Bromopropane in Mice. Vol. 76: 496–505. March 2007.
3. Redox Regulation of Sperm Surface Thiols Modulates Adhesion to the Fallopian Tube Epithelium. Vol. 76: 728–735. April 2007.
4. Autocrine Prolactin Inhibits Human Uterine Decidualization: A Novel Role for Prolactin. Vol. 76: 777–783. May 2007.

5. Catechol-O-Methyltransferase and Methoxyestradiols Participate in the Intraoviductal Nongenomic Pathway Through Which Estradiol Accelerates Egg Transport in Cycling Rats. Vol. 76: 934–941. December 2007.
6. Follicle-Stimulating Hormone Induces Spermatogenesis Mediated by Androgen Production in Japanese Eel, Anguilla japonica. Vol. 77: 970–977. December 2007.
7. Galectin 15 (LGALS15): A Gene Uniquely Expressed in the Uteri of Sheep and Goats that Functions in Trophoblast Attachment. Vol. 77: 1027–1036. December 2007.
8. Effects of Diets Enriched in Omega–3 and Omega–6 Polyunsaturated Fatty Acids on Offspring Sex-Ratio and Maternal Behavior in Mice. Vol. 78: 211–217. February 2008.
9. Meiotic Arrest in Human Oocytes Is Maintained by a Gs Signalling Pathway. Vol. 78: 667–672. April 2008.
10. Abortion in Mice with Excessive Erythrocytosis Is Due to Impaired Arteriogenesis of the Uterine Arcade. Vol. 78: 1049–105. June 2008.

The use of adverbial hedges in EAP students' oral performance

A cross-language analysis

Pascual Pérez-Paredes, Purificación Sánchez Hernández and Pilar Aguado Jiménez
Universidad de Murcia

This paper addresses the natural, non-elicited occurrence of adverbial hedges in the production of Spanish learners of EAP and British students of Modern Languages. The two sets of corpus data we discuss in this research are made up of interviews which were conducted following the same methodology and mirroring the tasks in the LINDSEI oral corpus (de Cock 1998). Linguistically, our research builds on Biber (1988) and Biber et al.'s (1999) Multidimensional Analysis of language, which maintains that there is a tendency for linguistic features of morpho-syntactic and semantic nature to cluster together around dimensions of use.
Results show important differences in the frequency of use as well as in the categories of adverbial hedges in these two comparable communities of use. Our research shows that, overall, both groups of speakers use adverbial hedges in a statistically significant and different way.

Keywords: Corpus linguistics, adverbial hedges, contrastive analysis

1. Introduction

It is not uncommon to hear English Language (EL) teachers affirm that hedging is a signal of advanced competence. Authors such as Ferris (1994), Hyland (1994), Wishnoff (2000) and Archibald (2001) point out that hedging is, among a few other features, an indicator of proficiency in L2 writing, although not a predictor of a high holistic score. As we will see in the literature review, the use of hedging appears, then, as a sophisticated feature of the language of learners of English. However, very little is known about the use of hedging by native speakers in communicative situations which resemble, or are exactly the same as, those where language

learners usually show their competence and skills (Nikula 1997; Markkanen & Schröder 1997). Would a native speaker hedge in a different way if confronted with a communicative situation which is typical of a foreign language teaching setting? To what extent can we affirm that non-natives and natives hedge differently if they do not face the same communicative demands or use the same register?

This paper addresses these questions by looking at the use of adverbial hedges in two groups of speakers who are confronted with exactly the same interview format, and who are required to express their opinions and views within the framework of an interview.

2. The scope of hedging in corpus-based research

The majority of the existing corpus-based research on hedging has addressed the written medium, with a marked emphasis on the study of academic discourse (Hyland 1998; Ädel & Reppen 2008). A first group of research efforts is focused on the examination of native discourse, either involving a single language or the comparison of several languages (Fitzpatrick 2007; Sanderson 2008). A second set of efforts deals with the contrast between native and non-native languages, usually English (Hunston 2002; Connor & Upton 2004).

Salager (1994) analysed the qualitative and quantitative presence of hedges in 15 articles comprising both research papers and case reports extracted from medical English written discourse. She studied hedges in the sections of introduction, discussion/comment, method and results and classified the hedging devices into shields, approximators, emotionally-charged intensifiers and compound hedges. Hyland (1995: 39) similarly studied hedges in four such parts in 26 biology research articles and concluded that "hedging represents a major rhetorical gap that L2 students have to cross before they can gain membership of a discourse community and pursue their chosen careers". The author reported that the ability to hedge statements appropriately is essential to effective communication and proposes that hedges should be given a higher priority in both our teaching and research agendas. Silver (2003) examined a single stance device, the adverbial *evidently*, in a corpus composed of history and economic articles written by native speakers. He contrasted the 26 overall occurrences in his corpus against the three functions of the adverb established in the Longman Corpus of Spoken and Written English, which are to a large extent shared by many other hedges: the position or function of the adverbial as a hedge or a booster; the way in which the adverbial intervenes in the overall 'meta-textual' strategy; the way in which the adverbial intervenes in the overall 'meta-pragmatic' strategy (Silver 2003: 364). His findings confirm that "hedges are employed to achieve a single primary objective: to

overcome the inherent negatability of statements and to gain the reader's acceptance of a knowledge claim" (Silver 2003: 371). They are, along with boosters, essentially argumentation devices which help the writer regulate her/his attention more to the proposition or to the reader, emphasizing or diminishing the truth value or writer accountability.

Vold (2006) carried out a comparison of epistemic modality markers in 450 English, French and Norwegian research articles (French: *sembler*, *paraitre*; *pourrait*; Norwegian: *kan*, *synes*, *sannsynligvis*; English: *may*, *assume*, *suggest* (Vold 2006: 69)). These pertained to linguistics and medicine and were extracted from the KIAP (Cultural Identity in Academic Prose) corpus. One of the purposes of her research was to investigate which factor influenced the use of hedges or, more precisely, the use of the selected epistemic modality markers in research articles. Inferential statistics yielded a significantly higher proportion of hedges in Norwegian and English articles in contrast to the French, a result which did not seem to be influenced by the discipline in question. However, differences emerged as to the actual type of markers used in either discipline, especially in the French and the English Corpus, where many of the markers were associated almost exclusively to one of the disciplines. The author's gender did not seem to be relevant for the frequency of epistemic modality markers used.

A contrastive analysis of the expression of indirectness between native and non-native written language can be found in Hinkel (1997). He studied 30 native and 120 non-native speaker university essays written in English and focused on classes of indirect strategies and devices such as rhetorical strategies and markers, lexical and referential markers, within which hedges are located, syntactic markers and other structures. Results showed that certain indirect strategies are significantly used in non-native speakers' essays as opposed to the native essays, whilst other indirect devices do not follow this pattern. Hinkel (2003) undertook a second contrastive analysis between native and non-native use of adverbials in written academic English. This time results lead him to conclude that "the more common certain types of adverb clauses in conversational discourse, the greater the likelihood of their high frequency rates in L2 academic essays" (Hinkel 2003: 1049). Milton (2001) highlights the difficulties EFL students show when learning hedging, probably due to cultural subtleties. In the particular case of Hong Kong students of English, many of the adverbs that are used as hedges in written assignments are misused.

Despite the above final remark on the saliency of linguistic features in written communication, the study of hedging in spoken discourse is noticeably far scarcer. Fung and Carter (2007) compare discourse markers in two different corpora: a pedagogic subcorpus from the CANCODE spoken British English corpus (460,055 words) and a corpus composed of classroom interactional discourse by

Hong Kong non-native English students (14,157 words). The students who contributed towards the corpus deployed an abundant use of referentially functional discourse markers (*and, but, because, OK, so*, p. 425) at the expense of other types of markers (*yeah, really, say, sort of, I see, you see, well, right*, p. 410), whereas the discourse markers produced by native speakers reflected an ampler range of pragmatic functions. Finally, Precht (2008) studied gender patterns in the expression of stance in a 900,000 word corpus consisting of American informal conversations. Multiple analyses of variance (ANOVAs) show no significant differences between sexes in terms of hedges and other stance devices, whilst expletives did reveal a significant difference between men and women. However, of the 27 hedges analysed in the article, 8 show significant differences between men and women. Men have significantly higher frequencies than women for *about, basically, like + adj/noun, something like, pretty*, and women have significantly higher frequencies than men for three (*almost, maybe, well*) (Precht 2008: 99).

Farr and O'Keefe (2002) investigated the socio-cultural context as a factor in explaining why speakers hedge in discourse. They look at the modal verb *would* in two institutionalised settings: radio phone-ins on national radio, subcorpus A, and post-observation teacher training interactions in a university setting, subcorpus B. Subcorpus A consists of about 55,000 words, while subcorpus B is composed of nearly 52,000 words. A first look at the data indicated that there were no significant differences in the frequency of use. However, after generating independent concordances, it was shown that three significant patterns clearly emerged: pronoun + *would*, *would* in questions, and *would* in verb phrase structures. Closer research reveals that much of the use of *would* is context-specific, giving rise to down-toning, concluding that hedges have a broader pragmatic function for the speakers in their data.

3. Hedging, the LGSWE and the Multidimensional Analysis-related research tradition

There is an important wealth of research that has examined hedging and related linguistic phenomena in the light of Multidimensional Analysis (Biber et al. 1999; Biber 2006; Cortes 2004). This research tradition seeks to interpret linguistic data in the light of variation across registers or, as Biber (1988) put it, different dimensions of use.

In this tradition, the analysis of hedging is commonly found within the realm of stance, which examines the expressions of emotion, attitude, certainty and doubt in language. Stance expressions can convey many different kinds of personal feelings and assessments, including attitudes that a speaker has about information,

how certain they are about its veracity, how they obtained access to the information and what perspective they are taking (Biber 2006: 87). Stance is related to register and dialect; person marking and part of speech are also important components of stance expression and can be expressed to differing extents through grammatical devices. Two common grammatical devices used to mark stance are adverbials and complement clause construction (Conrad & Biber 2000; Precht 2000).

According to the LGSWE, "adverbs can be used to realize all three types of stance: epistemic, attitude and style" (Biber et al. 1999: 557). Adverbs can be used to show different levels of certainty or doubt and also to convey imprecision. Adverbs such as *sort of* and *kind of* are also called hedges. Hedges such as *essentially*, *sort of*, and *virtually* indicate fuzzy restrictions (Lakoff 1972). The original definition has been widened in part due to the observation that certain verbs and syntactic constructions convey hedged performatives ("*I suppose/guess/think that Harry is coming; Won't you open the door?*" (Markkanen & Schröder 1997: 5). Hedging can be described as a strategy speakers use to mitigate and soften the force of their utterances (Nikula 1997) and has received much attention as a relevant part of text analysis, especially in written registers (Östman 1981; Schiffrin 1987).

Biber (2006: 103–106) analysed the expression of stance through adverbs of certainty (e.g. *actually*, *in fact*), likelihood (e.g. *possibly*, *probably*), attitude (e.g. *amazingly*, *curiously*) and style (e.g. *generally*, *typically*) and found that these adverbs are more common in the spoken register than in the written one. Within the spoken register, certainty adverbs are the most frequently used, followed by likelihood adverbs. Style adverbs are less frequent, while attitudinal adverbs are the least common overall.

Within the realm of university discourse (Biber 2006), the group of adverbial hedges has a positive loading of 0.55 within Dimension 1 (Oral vs. Literate discourse). In Dimension 2 (Procedural vs. Content-focused discourse), adverbial hedges do not show a significant factor loading, and are not included within this dimension. This same behavior appears in Dimension 3 (Reconstructed account of events). However, likelihood adverbials show a positive loading of 0.35 in Dimension 4 (Teacher-centered stance).

These results are in line with Poos and Simpson (2002), who focused their research on the use of *kind of* and *sort of* as prototypical examples of hedging devices in academic spoken English. They consider these two prototypical hedges to hypothesize that their frequency of use would correlate more strongly with academic division than with speaker gender. Both authors propose that there is a higher rate of hedges in the fields of humanities and social sciences compared with natural sciences. Thus, they conclude that academic discipline is a stronger predictor of frequency of hedges than gender. Their work also includes a pragmatic analysis and a close analysis of one speaker, showing that *kind of* and *sort of* are

multifunctional and serve a variety of overlapping purposes in spoken interaction, as reducing the force of an utterance or conveying inexactitude (e.g. *we're talking about sort of racial cultures*...), mitigating criticisms (e.g. ...*you can't let your argument kind of disappear*...), accommodating to interlocutors who are not familiar with the jargon (e.g. ...*along with that success has come kind of Faustian bargain*...), and others.

Precht (2003) identified stance moods in British and American English conversation. The author uses factor analysis to identify co-occurrence patterns. This multidimensional analysis is considered by Precht as a valuable tool in identifying complex relationships in language. For her study, she used part of the spoken section of the LGSWE, so the contexts considered in her study were conversation among related and non-related adults, conversation among family members, and conversation at work. Applying factor analysis and optimizing results, Precht obtained three factors: informal affect, boulomaic planning versus small talk, and hedged opinion. This last dimension was primarily made up of hedges (adverbial, verbal and modal). She concludes that American English conversations among adults tend to use much more affect (shown in factor 1) compared with British-English conversations (factor 3).

Cortes (2004: 407–8) focused on four-word lexical bundles in a student writing corpus and a corpus of published academic writing covering history and biology texts. From her study, she developed a taxonomy of lexical bundles organized into four categories: referential bundles (time markers and place markers, e.g. *at the time of*, *the size of the*), text organizers (contrast/comparison bundles, inferential bundles, focus bundles, and framing bundles, e.g. *as a result of*, *in the absence of*), and *stance bundles* (epistemic-impersonal/probable-possible, e.g. *may be due to*, *are likely to be*). The fourth category is devoted to functionally classifying all bundles. Stance bundles do not occur in the history section of the corpus, and the author states that most of these bundles are mainly used as hedges, introducing a degree of tentativeness or to hedge the effect of an assertion.

Biber et al. (2004) presented a multidimensional analysis of conversation, using the British English subcorpus of conversation from the LGSWE (about 4,000,000 words). He found three dimensions: information-focused vs. interactive discourse, stance vs. content-focused discourse, and narrative-focused discourse. In the particular case of dimension 2, stance vs. content-focused discourse, most of the linguistic features show positive loads in the expression of stance, and among them, general hedges are found (all adverbials with a hedging function, such as *at about*, *something like*, *almost*). Using cluster analysis based on the three dimension scores, Biber (2004) proposed different conversation types: informational vs. context-focused, informational vs. stance-focused, interactive vs. context-focused, narrative, unmarked interactive and unmarked context-focused. The second type

is especially marked by the use of stance markers, and that their primary purpose is to express stance in relation to a particular topic. Biber concludes that the results of this analysis are very similar to the ones obtained in his 1988 study and, at the same time, closely related to analyses of spoken and written registers in other languages, i.e. Somali and Korean.

4. Methodology

In our research we set out to analyse the occurrence of adverbial hedges in the production of Spanish learners of English for Academic Purposes (EAP), hereafter C1, and British students of Modern Languages, hereafter C2. In order to examine natural, non-elicited occurrence of adverbial hedges, we decided to gather spoken data which (1) could be comparable, that is, speakers of different mother tongues, C1 and C2, going through the same interview format, and (2) reflected a variety of communicative situations.

In reference corpora, a wide array of different genres and texts contribute to the spoken component as these corpora are built to represent as faithfully as possible the spoken medium as a whole. In the case of the British National Corpus (BNC) and the Longman Spoken and Written English Corpus, the spoken component involved recruits who recorded all their conversations unobtrusively over two or three days. In EFL and ESL contexts, however, it is difficult, if not completely impossible, to gather data which is not elicited by means of an interview. Apart from ethical reasons, for learners who lead their lives in their L1 countries it would set up artificial communicative situations which would do very little in favour of the type of data analysis we pursue. It is not surprising to find then that the interview has been the most widely used elicitation technique in learner data compilation (Tono 2003).

Conducted by English native speakers, the interviews followed the format of the Louvain International Database of Spoken English Interlanguage (LINDSEI) corpus, (de Cock 1998) and thus were structured in three parts. First, speakers were given three topics for discussion, namely, an experience that has taught the speaker an important lesson, a country that has impressed the contributor, or a film or play that has attracted the attention of the speaker. Then, a small part of the interview was devoted to interpersonal communication. Finally, students were given four pictures which represented a story and were asked to describe them and offer an account of what was going on there[1].

1. Further details can be found at http://www.uclouvain.be/en-cecl-lindsei.html

The reasons behind the use of the LINDSEI elicitation framework are that (1) this interview format presents researchers with opportunities to record language behaviour in different communicative situations, and (2) the LINDSEI interview format will favour future comparability with forthcoming LINDSEI learner data.

59 English for Academic Purposes (EAP) students contributed to our corpus of Spanish informants, C1, whose mean age was 19.6 years. The average number of years studying English prior to university was 8.8, and 45.8% of them had travelled to English-speaking countries, with an average stay in months of 1.9. All these students were enrolled in *Lengua Inglesa I*, a compulsory subject in the Degree in English Studies offered at the Universidad de Murcia, Spain. The total number of words, considering only the informants' production, not the interviewers', is 45,558[2], and the mean word count is 772.16 per contributor.

28 students of Modern Languages (ML) contributed to our corpus of British speakers of English, C2, all of them native speakers of English, with a mean age of 22.25 years. In C2, the total number of words considered is 21,509, and the mean word count is 768.17 per contributor. The interviews were recorded at the Manchester Metropolitan University during the second term of the 2005–2006 academic year. Data processing was the same as in C1.

In order to analyse adverbial hedges, we decided to look at the group of adverbial hedges in Biber (1988), and so the occurrences of *almost, maybe, sort of* and *kind of* were selected and subsequently classified into the four types of categories suggested by Biber et al. (1999: 557–558): 1) Imprecision word choice; 2) Approximators and quantifiers; 3) Uncertainty; and 4) Degree diminishing. We analysed each occurrence of the hedges under study in both corpora and classified them into one of the above categories. The occurrences of the selected hedges in C2 all fitted in one of the four categories. However, in C1 there were some occurrences which proved difficult to classify, especially *maybe*. At this point we decided to create a new category called 0 category for one of the occurrences of *maybe*, which did not fit in any of the four categories.

5. Results

In the EAP learner corpus, C1, we found 58 occurrences of the four adverbial hedges analysed: *kind of*, *sort of*, *maybe* and *almost*. The following occurrences are taken from C1:

2. The total count is higher if we include the interviewer's questions, backchannelling and comments. For the purposes of this research we have decided to include only the tokens produced by the students.

yeah because our teacher was [**kind of**] American accent because her daughter lived in the United States
(CAOS-E Corpus 1)
the area that I come from.. [**sort of**] middle class area you know .. [**sort of**] .. But .. em .. it seems now the fashion is to become very rough
(CAOS-E Corpus 1)
to know something more about different countries and er a new country well er more than one year [**maybe**] one year and a half because because my mother
(CAOS-E Corpus 1)
felt ashamed .. so I spoke Spanish [**almost**] all the all the time in my house in my room my I stayed with
(CAOS-E Corpus 1)

In C2 we found 70 occurrences of four different adverbial hedges, namely, *kind of*, *sort of*, *maybe* and *almost*. The following are taken from C2:

accepted sort of norm in a small French .. [**sort of**].. extremely Catholic .. er village and er people coming in from
(CAOS-E Corpus 2)
mixing things up a bit erm .. so yeah it gets you [**kind of**] like thinking .. erm .. favourite actor 's probably erm
(CAOS-E Corpus 2)
even though he didn t look like the evil man he portrayed.. it was [**almost**] evil the way he did it
(CAOS-E Corpus 2)\
and the school was sort of open .. in the morning from [**maybe**] eight o'clock .. till about twelve or one there 's they
(CAOS-E Corpus 2)

Table 1 shows the descriptive data for C1, while Table 2 shows the weight of the adverbial hedges found in terms of percentage.

Table 1. Descriptive data for C1

Number of adverbial hedges	58
Informants	59
Total number of words	45,558
Frequency of adverbial hedges vs. total number of words	0.0012
Standard deviation	1.2400

Table 2. Adverbial hedges found in C1

Tokens	
Almost	18.96%
Maybe	70.68%
Sort of	06.89%
Kind of	03.44%

From the data in Table 2 it can be observed that Spanish students tend to use one-word adverbial hedges, nearly 90% do, in contrast with multiword adverbial hedges.

Table 3 shows the descriptive data for C2, while Table 4 shows the weight of the adverbial hedges found in terms of percentage.

Table 3. Descriptive data for C2

Number of Adverbial hedges	70
Number of informants	28
Total number of words	21,509
Frequency of adverbial hedges vs. total number of words	0.0032
Standard deviation	2.5500

Table 4. Adverbial hedges found in C2

Tokens	
Almost	01.42%
Maybe	20.00%
Sort of	34.28%
Kind of	44.28%

Contrary to the trend observed in the Spanish students, English native speakers show a preference for multiword adverbial hedges, with almost 80% of the total count, which seems to resemble natural spoken interaction. Table 5 compares findings in both corpora.

Although the number of students is higher in C1 than in C2, the ratio of words per informant is similar in both corpora (772.1 in C1 and 768.2 in C2). The frequency of use and the preferences in each group vary to a great extent. In C1, 23 informants use hedges, which represents 34% of the total number of informants, whereas in C2, 21 informants use hedges, which represents 75% of the total of participants. Consequently, even with a similar number of words per student in

Table 5. Comparison between the descriptive data for C1 and C2

C1 (Spanish EAP students)	C2 (British ML students)
Informants: 59	Informants: 28
Running words: 45,558 (informant contribution only)	Running words: 21,509 (informant contribution only)
Mean of words per informant: 772.1	Mean of words per informant: 768.2
Number of informants who use adverbial hedges: 23 (34% of the informants)	Number of informants who use adverbial hedges: 21 (75% of the informants)
Frequency of tokens as adverbial hedges with respect to the total number of words: 0.0012	Frequency of tokens as adverbial hedges with respect to the total number of words: 0.0031
Ratio of adverbial hedges with respect to each informant of the Corpus: 0.38	Ratio of adverbial hedges with respect to each informant of the Corpus: 1.10

both corpora, native students use almost three times more hedges (0.0031) than non-native students (0.0012), and accordingly the mean of adverbial hedges with respect to each participant is almost three times higher in C2 than in C1.

C1 and C2 informants did not exhibit the same patterns of use of the different hedges, as can be seen in Table 6, where we can appreciate that the frequency of use of *almost* and *maybe* is much higher in C1 than in C2, whereas *sort of* and *kind of* are more frequently used in C2 than in C1.

Differences also affect the frequency of use of the adverbial hedges involved. Spanish students, C1, tend to use *almost* and *maybe* in a high proportion, whereas British students, C2, use *sort of* and *kind of* instead. Biber (2006: 92) characterises these adverbs as *stance adverbs*, the 4 items being used by all students within the group of *likelihood adverbs* (*almost* is included in Biber 1988, but not in Biber 2006). Figure 1 represents the raw frequency of the four adverbial hedges analysed in both corpora:

Table 6. Comparison of hedges used in C1 and C2

Hedge	Corpus 1 (%) Spanish students	Corpus 2 (%) British students	Differences (C1 – C2)	Representation of hedges in C1 as compared with C2
Almost	0.0024	0.0004	+0.0020	More frequent
Maybe	0.0090	0.0060	+0.0030	More frequent
Sort of	0.0010	0.0100	–0.0090	Less frequent
Kind of	0.0002	0.1200	–0.1198	Less frequent

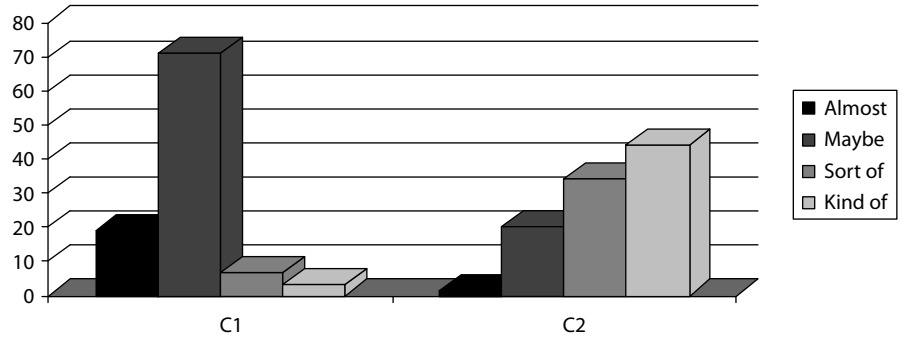


Figure 1. Adverbial hedges used in C1 and C2

Likelihood/imprecision adverbs as hedges are preferred within the group of epistemic adverbs, and also compared with attitude and style stance adverbs. So, the speakers in C1 do not seem to show in their speech any reference to their attitude towards any proposition (such as an evaluation or expectations), or to the manner of conveying their message by means of adverbial hedges.

As discussed previously, we have also looked at the different uses that these adverbial hedges have been put to. To this effect, we classified every occurrence into one the four types or categories suggested by Biber et al. (1999: 557–558): imprecision word choice, approximators and quantifiers, uncertainty and, finally, degree diminishing.

Table 7 shows the raw frequencies for the five types of adverbial hedging uses in the corpus of Spanish EAP Students:

0 category was developed as a convenient way to classify uses of *maybe* which did not accommodate the definition and scope of uncertainty in the LGSWE.

> ... the afternoons we sometimes we had trips... for activities or something like that or free time [**maybe**] mm... you may think it is a stupid reason but I started learning English when I was ten

Table 7. *Raw frequencies for* the five types of adverbial hedging uses *in C1*

	Minimum	**Maximum**	**Mean**	**SD**
Imprecision word choice	0.000	5.000	0.08512	0.650902
Approximators and quantifiers	0.000	0.022	0.00037	0.002858
Uncertainty	0.000	0.176	0.01265	0.030284
Degree diminishing	0.000	0.066	0.00409	0.011866
0 category	0.000	0.044	0.00223	0.008832
Total hedges	0.000	0.176	0.02158	0.039411

Table 8. *Raw frequencies for* the five types of adverbial hedging uses *in C2*

	Minimum	Maximum	Mean	SD
Imprecision word choice	0.000	0.558	0.07472	0.129252
Approximators and quantifiers	0.000	0.046	0.00166	0.008786
Uncertainty	0.000	0.232	0.02159	0.048154
Degree diminishing	0.000	0.046	0.00166	0.008786
0 category	0.000	0.000	0.00000	0.000000
Total hedges	0.000	0.558	0.10627	0.134424

Table 8 shows the raw frequencies for the five types of adverbial hedging uses in the corpus of British EAP Students.

Table 9 shows the distribution and percentages of adverbial hedges uses in C1 and C2.

From Table 9 it can be observed that most adverbial hedges used by C1 speakers function as uncertainty hedges. This attribution, however, cannot be made straightforwardly as *maybe* could denote doubt, and, accordingly, could have nothing to do with hedging. Our British speakers clearly show a higher preference for adverbial hedges conveying the function of imprecision word choice.

In order to find out whether these differences were statistically significant, we ran a t-test for independent samples where the means of use of adverbial hedges in each of the types or categories of use in Biber et al. (1999: 557–558) in both groups of speakers, Spanish EAP students and British ML students, were tested for significance. Results are shown in Table 10.

Table 9. Distribution and percentages of hedges according to categories of use in C1 and C2

	Almost	Maybe	Sort of	Kind of
C 1 59 informants *34% of the informants use hedges*	11 (18% of the total of hedges) Cat. 4 (degree diminishing)	41 (70.6% of the total of hedges) Cat. 3 (uncertainty) 38 Cat. 2 (approximators + quantifiers) 1 Cat. 1 (imprecision word choice) 1 Cat. 0 (no hedges) 1	5 (8.5% of the total of hedges) Cat. 1 (imprecision word choice)	1 (1.7% of the total of hedges) Cat. 1 (imprecision word choice)
C2 28 informants *75% of the informants use hedges*	1 (1.5% of the total of hedges) Cat. 4 (degree diminishing)	14 (21.8% of the total of hedges) Cat. 3 (uncertainty)	23 (35.9% of the total of hedges) Cat.1 (imprecision word choice)	26 (40.6% of the total of hedges) Cat. 1 (imprecision word choice)

Table 10. T-test for independent samples for the means of use of adverbial hedges in each of the types or categories of use

		Levene's test for equality of variances		T-test for equality of means						
		F	Sig.	t-statistic	Degrees of freedom	Sig. (two-tailed)	Mean difference	Standard error difference	95% Confidence interval of the difference	
									Max	Inferior
Imprecision word choice	Equal variances assumed	0.458	0.500	0.084	85.000	0.934	0.010398	0.124516	−0.237172	0.257969
	Equal variances not assumed			0.118	67.044	0.906	0.010398	0.088190	−0.165628	0.186425
Approximators and quantifiers	Equal variances assumed	4.400	0.039	−1.023	85.000	0.309	−0.001288	0.001259	−0.003791	0.001215
	Equal variances not assumed			−0.757	29.744	0.455	−0.001288	0.001702	−0.004765	0.002188
Uncertainty	Equal variances assumed	3.792	0.055	−1.055	85.000	0.294	−0.008936	0.008470	−0.025778	0.007905
	Equal variances not assumed			−0.901	37.473	0.373	−0.008936	0.009918	−0.029023	0.011150
Degree diminishing	Equal variances assumed	3.453	0.067	0.965	85.000	0.337	0.002432	0.002520	−0.002579	0.007443
	Equal variances not assumed			1.072	69.670	0.287	0.002432	0.002268	−0.002092	0.006956
0 category	Equal variances assumed	7.982	0.006	1.333	85.000	0.186	0.002232	0.001674	−0.001097	0.005561
	Equal variances not assumed*			1.941	58.000	0.057	0.002232	0.001150	−0.000069	0.004534
Total hedges	Equal variances assumed	26.869	0.000	−4.475	85.000	0.000	−0.084690	0.018923	−0.122315	−0.047065
	Equal variances not assumed*			−3.268	29.225	0.003	−0.084690	0.025917	−0.137678	−0.031702

Equal variance, or its absence, is indicated by means of an asterisk in every type or category analysed after the Levene test. Differences between groups are only significant when the five categories of adverbial hedging are taken as a unique set of data (Significance 0.003 < 0.005), which may point to the fact that the uniqueness which represents type 5 in C1 may somehow distort the interpretation of the data when taken as a whole.

In the rest of the individual categories of adverbial hedging use, we find no significant differences for these two particular groups of speakers. It is close to significance in category/type 5 (Significance 0.0570) and, far from significance, in descending order, for degree diminishing (Significance 0.2870), uncertainty (Significance 0.0373), approximators and quantifiers (Significance 0.4550) and, finally, imprecision in word choice (Significance 0.9060).

6. Discussion and conclusion

Adverbials are the second most important device to convey hedging in English (Hyland 1996), only second to lexical verbs, and more prominent than other areas that have received more academic attention such as modal verbs. The use of adverbial hedges in this research by the Spanish informants indicates, judging from the frequencies of use observed, shortcomings in the oral production of, at least, this group of Spanish EAP students. Our results may point to the lack of formal instruction on the part of C1 speakers in the use of hedging devices. In turn, the comparatively higher use of adverbial hedges in native language implies a lower lexical density among English speakers, who somehow avoid the specification of meaning (Leech 1998). This supports one of the main characteristics of spoken communication, in which the speaker avoids the elaboration or specification of meaning (Biber et al. 1999; Leech 1998).

Despite these apparent differences, the uses that these adverbial hedges were put to by both groups of speakers were not statistically significant when the different uses of adverbial hedging are considered individually. The differences between C1 and C2 groups were only significant when the five categories of adverbial hedging were taken as a unique set of data (Significance 0.003 < 0.005). This means that both groups of speakers did not behave differently when using the four adverbials analysed across the different uses we looked at, that is, imprecision word choice, approximators and quantifiers, uncertainty and degree diminishing. This may be explained, at least in part, by the convergence of a situation where 66% of the 28 informants in C1 did not use adverbial hedges at all plus the fact that, even when 75% of the speakers in C2 used them, the standard deviation of the mean of adverbial hedges used in the C2 sample is very high (see Table 3). Thirteen informants

in C2 do not use adverbial hedges as a way to express word imprecision; only one uses adverbial hedges as approximators or quantifiers, nineteen speakers do not express uncertainty through adverbial hedges; and only one uses adverbial hedges to express degree diminishing. This clearly explains the high standard deviation shown in Table 3. At the same time, this shows how easily corpus data can be influenced by the speakers' own idiosyncratic speech habits. Notwithstanding, this research shows that, overall, both groups of speakers use adverbial hedges in a statistically significant different way.

Given the narrower range of uses given to adverbial hedges in the contributors to C1, we can state that the rhetorical consciousness of the "hedging" device as manifested by adverbials is much lower, which may result in poor persuasive language skills and a different projection of the affective interpretations that the interlocutors are engaged in. For this reason, reinforcing these skills in non-native speakers would make students acquire a richer repertoire of use of the English language, not only in terms of accuracy and correction, but also in terms of cultural behaviour.

Although all texts compiled were obtained in an academic context, the use of the adverbial hedges analysed denotes informality, except for the case of *almost*, which is classified by Biber (1988) as a downtoner as well, being also considered by Chafe and Danielewicz (1986) as an academic hedge. *Kind of* and *sort of* have traditionally been regarded as linguistic items that are typical of informal style and non-standard language varieties (Miskovic-Lukovic 2009). Thus, in spite of the fact that these corpora were compiled in a university context, their oral/spoken mode, in some way, is stronger than the communicative situation, so the use of hedges appears as a natural consequence of oral interaction, independent of the communicative context and the formality/informality of discourse.

Maybe is widely used by speakers of C1 probably as a transference from L1, which agrees with Milton (2001) who states the difficulties EFL students show when learning hedging, probably due to cultural subtleties. An equivalent term is commonly used in Spanish. However, there can be other reasons for the "overuse" of *maybe* related to the lack of confidence of the Spanish students when they speak English. Moreover, the overlapping of *maybe* as an uncertainty hedge and *maybe* as a non-hedge adverb can be seen in many occasions in C1 speakers discourse:

> I think also it's a different lifestyle different kind of people mm... I think here [**maybe**] we can be bit more ru = rude in our manners but at the end of the day we are
> Yeah 12 years old no and I I wish I I repeat this experience but [**maybe**] the the next year the next sum no no the next the following the I I would like

These uses of *maybe* are preferred by non-native speakers in our data to those of tentativeness or vagueness. The same behaviour is observed with *almost*, whose Spanish equivalent is widely employed in conversation.

Sort of and *kind of*, considered to soften FTAs, face-threatening acts, through vagueness (Brown & Levinson 1987; Carter & Simpson 1989) are almost exclusively used by the informants of C2, which agrees with Fung and Carter (2007) who report a restricted use of markers such as *sort of* by Hong Kong learners of English. In fact, the students who used these two hedges in C1 are precisely those who, based on their final marks in the academic year, have the best command of the English language in the group, which confirms Hyland's findings that "hedging represent a major rhetorical gap that students have to cross before they can gain membership of a discourse community and pursue their chosen careers" (1995: 39).

O'Keefe, McCarthy and Carter (2007), when studying hedging in CANCODE (Cambridge and Nottingham Corpus of Discourse in English), consider *kind of* as one of the hedges preferred in North American spoken English. However, our findings in the C2 data point out that young British speakers show a preference for this hedge. It is also worth mentioning the remarkable preference of C2 speakers for the use of *kind of* and *sort of* over other adverbial hedges such as *like*, *probably*, *surely* or *perhaps,* the reason being the degree of informality associated to the spoken register in which they are interacting.

A significant finding is that both groups of participants use hedges in their speeches corresponding to the first part of the interview related to the description of a personal experience. In the second/third part of the interview, more descriptive hedges are not registered. In this sense, hedges, in general, seem to be more associated to the elicitation of personal information, opinions, views, etc., rather than of more objective, impersonal information. So, facts such as politeness and an attempt not to antagonise the listener may explain the absence of adverbial hedges in these parts of the interviews.

With the exception of Silver (2003) and Precht (2008), most of the literature reviewed in this paper explicitly includes pedagogical recommendations for the teaching of hedges and other stance strategies to non-native speakers, in the light of the results obtained. Vold (2006) highlighted the importance of the teaching of hedging strategies adjusted to the disciplinary affiliation of the students. To this, we add the importance of teaching the different uses which adverbial hedges may have.

Additionally, teachers should infer from this research the need to incorporate this aspect of oral communication in the Spanish EFL students' syllabus. Improvements in the learning of hedges will lead to a better command of the target language (Hyland 1996), so some pedagogical implications should be taken into account and put into practice. Our research shows that, even when presented with the same communication tasks, C1 speakers expressed a very high degree of uncertainty through adverbial hedges, while C2 speakers showed a lower use of this function and preferred to stress word imprecision hedging.

The limitations of our research are primarily concerned with the sizes of our samples (n = 59 and n = 28) which, while being comparable to those of most learner language studies, are still too small to allow us to generalize our findings to the population of Spanish speakers of English or the British students of Modern languages.

Our study confirms, partially, the findings of Biber (2006) who, using multidimensional analysis, uncovered the existence of a dimension in spoken communication which displays a tension between information and stance. The interview texts in our corpus may be instances of this continuum, and their different parts a demonstration that the concurrence of medium/genre has a profound effect on the nature of linguistic evidence that linguists draw on.

As part of future work, a deeper analysis of the use of adverbial hedges in the C2 corpus will be carried out to test whether the traditional views on hedging (Salager-Meyer 1994; Hyland 1998) hold true in spoken contexts.

Acknowledgements

We thank the Fundación Séneca, the Agencia Regional de Ciencia y Tecnología, and the Región de Murcia for their support. This research is part of 00580/PI/04 *Estudio longitudinal y comparativo de la producción oral (oralidad) de aprendices de inglés como lengua extranjera a partir del análisis de un corpus lingüístico oral recopilado y mediante procedimientos de análisis multidimensional.*

References

Ädel, A. & R. Reppen (eds). 2008. *Corpora and Discourse: The Challenges of Different Settings* [Studies in Corpus Linguistics 31]. Amsterdam: John Benjamins.

Archibald, A. 2001. Targeting L2 writing proficiencies: Instruction and areas of change in students. Writing over time. *IJES* 1(2): 153–174.

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

Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E. & Urzua, A. 2004. Representing language use in the university: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus. *TOEFL Monograph Series MS 25.*

Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: CUP.

Biber, D. 2004. Conversation text types: A multi-dimensional analysis. Paper presented at JADT 2004: 7es Journées internationales d'analyse statistique des données textuelles.

Biber, D. 2006. *University Language: A Corpus-based Study on Spoken and Written Registers* [Studies in Corpus Linguistics 23]. Amsterdam: John Benjamins.

Brown, P. & Levinson, S. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Carter, R. & Simpson, P. 1989. *Language, Discourse and Literature: An Introductory Reader in Discourse Stylistics*. London: Unwin Hyman.

Chafe, W. & Danielewicz, J. 1986. Properties of spoken and written language. In *Comprehending Oral and Written Language*, R. Horowitz & S.J. Samuels (eds). Orlando FA: Academic Press.

Connor, U. & Upton, T.A. (eds). 2004. *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi.

Conrad, S. & Biber, D. 2000. Adverbial marking of stance in speech and writing. In *Evaluation in Text: Authorial Stance and the Construction of Discourse*, S. Hunston & G. Thompson (eds), 56–73. Oxford: OUP.

Cortes, V. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23(4): 397–423.

De Cock, S. 1998. Corpora of learner speech and writing and ELT. In *Germanic and Baltic Linguistic Studies and Translation: Proceedings of the International Conference held at the University of Vilnius*, A. Usoniene (ed.), 56–66. Vilnius: Homo Liber.

Farr, F. & O'Keefe. A. 2002. *Would* as a hedging device in an Irish context: An intra-varietal comparison of institutionalised spoken interaction. In *Using Corpora to Explore Linguistic Variation* [Studies in Corpus Linguistics 9], R. Reppen, M. Fitzmaurice & D. Biber (eds), 25–48. Amsterdam: John Benjamins.

Ferris, D.R. 1994. Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly* 28(2): 414–420.

Fitzpatrick, E. (ed.). 2007. *Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse*. Amsterdam: Rodopi.

Fung, L. & Carter, R. 2007. Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics* 28(3): 410–439.

Hinkel, E. 1997. Indirectness in L1 and L2 academic writing. *Journal of Pragmatics* 27: 361–386.

Hinkel, E. 2003. Adverbial markers and tone in L1 and L2 students' writing. *Journal of Pragmatics* 35: 1049–1068.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: CUP.

Hyland, K. 1994. Hedging in academic writing and EAF textbooks. *English for Specific Purposes* 13(3): 239–256.

Hyland, K. 1995. The author in the text: Hedging scientific writing. *Hong Kong Papers in Linguistics and Language Teaching* 18: 33–42.

Hyland K. 1996. Nurturing hedges in the ESP curriculum. *System* 24(4): 477–490.

Hyland, K. 1998. *Hedging in Scientific Research Articles* [Pragmatics & Beyond New Series 54]. Amsterdam: John Benjamins.

Lakoff, G. 1972. Hedges: A study of meaning criteria and the logic of fuzzy concepts. In *Papers from the Eighth Regional Meeting, Chicago Linguistic Society*, P.M. Peranteau, J.N. Levi & G.C. Phares (eds), 183–228. Chicago IL: Chicago University Press.

Leech, G. 1998. The special grammar of conversation? *Longman Language Review* 5: 9–14.

Markkanen, R. & Schröder, H (eds). 1997. *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts*. Berlin: Walter de Gruyter.

Milton, J. 2001. Doubt and certainty in HK interlanguage. In *Elements of a Written Interlanguage: A Computational and Corpus-based Study of Institutional Influences on the Acquisition of English by Hong Kong Chinese Students*, 79–87. Hong Kong: The Hong Kong University of Science and Technology.

Miskovic-Lukovic, M. 2009. Is there a chance that I might kinda sort of take you out to dinner? The role of the pragmatic particles kind of and sort of in utterance interpretation. *Journal of Pragmatics* 41(3): 602–625.

Nikula, T. 1997. Interlanguage view on hedging. In *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts,* R. Markkanen & H. Schröder (eds), 188–207. Berlin: Walter de Gruyter.

O'Keefe, A., McCarthy, M. & Carter, R. 2007. *From Corpus to Classroom. Language Use and Language Teaching*. Cambridge: CUP.

Östman, J.O. 1981. *You Know: A Discourse-functional Approach*. Amsterdam: John Benjamins.

Poos, D. & Simpson, R. 2002. Cross-disciplinary comparisons of hedging: Some findings from the Michigan Corpus of Academic English. In *Using Corpora to Explore Linguistic Variation* [Studies in Corpus Linguistics 9], R. Reppen, S.M. Fitzmaurice & D. Biber (eds), 3–23. Amsterdam: John Benjamins.

Precht, K. 2000. Patterns of Stance in English. PhD dissertation, Northern Arizona University.

Precht, K. 2003. Evidentiality and affect in British and American conversation. *Text* 23(2): 239–257.

Precht, K. 2008. Sex similarities and differences in stance in informal American conversation. *Journal of Sociolinguistics* 12(1): 89–111.

Salager-Meyer, F. 1994. Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes* 13(2): 149–170.

Sanderson, T. 2008. *Corpus, Culture, Discourse*. Tübingen: Gunter Narr.

Schiffrin, D. 1987. *Discourse Markers*. Cambridge: CUP.

Silver, M. 2003. The stance of stance: A critical look at ways stance is expressed and modelled in academic discourse. *Journal of English for Academic Purposes* 2: 359–374.

Tono, Y. 2003. Learner corpora: Design, development and applications. In *Proceedings of the Corpus Linguistics 2003 Conference. UCREL Technical Papers 16*, D. Archer, P. Rayson, A. Wilson & T. McEnery (eds), 800–809. Lancaster: Lancaster University.

Vold, E.T. 2006. Epistemic modality markers in research articles: A cross-linguistic and cross-disciplinary study. *International Journal of Applied Linguistics* 16: 61–87.

Wishnoff, J.R. 2000. Hedging your bets: L2 learners' acquisition of pragmatic devices in academic writing and computer-mediated discourse. *Second Language Studies* 19(1): 119–148.

Integrating approaches to visual data commentary

An exploratory case study

Carmen Sancho Guinda
Universidad Politécnica de Madrid

This paper highlights the need for explicit instruction in visual data commentary within ESP environments, positing its status as a multi-skill and cross-disciplinary practice. It reviews the state of the art in linguistic and didactic research and reports on a preliminary case study conducted at the School of Aeronautical Engineering of Madrid (Universidad Politécnica) which involves samples written by fifty-seven second-year students. Findings reveal: (a) a scarce use of metadiscourse and presentational constructions with the adoption of compensatory tactics, (b) a strong preference for common-core and pre-modified superordinate nouns, and (c) certain collocational interferences with the L1. A final outline of pedagogic directions is provided following the principles of Systemic Functional Grammar and Discourse Analysis.

Keywords: Information transfer, data interpretation, ESP, transversal skill

1. Perspectives, research aims and antecedents in the study of graphic literacy

How is data commentary to be approached and taught in the ESP classroom? As a complex professional and academic *text type* in the service of argumentation and hybrid between exposition, narrative and description? As a sub-genre or embedded genre (Bhatia 1997: 191), common to different channels and disciplines but unmistakably within situated genre templates? As a flexible genre (Paltridge 2002) or a macro genre in the systemic-functional sense (Martin 1995), combining more elementary texts such as recounts, explanations, descriptions and narratives? As a versatile mixed genre (Bhatia 1997: 190) joining multifarious discursive tactics and elements? As a strategic component of certain rhetorical moves? Or – precisely the contention of this paper – as a transversal skill with multiple realisations based on a wide repertoire along a cline of choices? Be that as it may, many an

applied linguist has noted the relevant and ubiquitous role of graphic information processing in academic, scientific and technical discourses, inevitably multimodal (Lynch & Woolgar 1990: 1–2; Swales & Feak 1994/2004: 112; Myers 1997: 93–94, 2003: 4–5; Johns 1998: 183–186; Lemke 1998: 87; Miller 1998: 29; Busch-Lauer 1998: 109; Kress et al. 2001: 3, among others). Graphs, diagrams and tables, in effect, do abound in research articles, textbooks, exams, lectures, class discussions, oral presentations, company meetings, stock exchange interactions, medical diagnoses and political debates. In the academic and scientific fields they are equally found as part of the methods and results sections within research articles and illustrating, complementing, reinforcing and summarising the contents transmitted in lectures, manuals and other pedagogical genres oral or written.

To consider the verbalisation of visual input, a genre – whether macro, flexible, mixed, pure or embedded – is nonetheless debatable: firstly, for some the margin of variation as regards communicative purpose and socio-cultural contexts of occurrence (scientific, technical, occupational and academic – this latter constituted by both teachers and learners) might appear much too broad to define a single discourse community of users. Let us remember that according to Swales (1990) and Bhatia (1993, 1997) genres are meant to attain the communicative goals of specific discourse communities. Perhaps these communities could be agglutinated by a consistent and overarching persuasive aim which would even imply viewing instruction as an act of persuasion, a notion that is not new (Cros 2003; Murphy 2001). Secondly, and despite the attempts of Bertin (1977/1981), Olsen and Huckin (1991) and Swales and Feak (1994/2004) to distinguish discrete moves in the commentaries of graphic data,[1] in everyday practice no uniform sequence of rhetorical shifts is normally found to make up a recognisable macrostructure or genre integrity (Bhatia 1999: 22) helping text producers and users to identify stable features and the constraints acting on them, which suggests its generic formalities are not as socially institutionalised as they should be. Also in this respect, although Miller (1998: 37) observes that graphs, diagrams and tables infallibly contain comparison and cause and effect relationships, their ordinary arrangement in written or spoken form does not seem to follow a

1. Bertin's moves sequence (1977/1981: 184–185) comprises the following steps: 1) Recall of the hypotheses of the classing procedure, 2) Identification and naming of the groups of features discovered, drawing analogies with groupings outside the data, 3) Identification of the groups of objects by referring to those groups of characteristics, 4) Study of the special cases and exceptions, and 5) Re-questioning of the corpus of data and features chosen by eliminating anecdotal elements to build coherent groups. Olsen and Huckin (1991: 172–173) restrict their moves to two: 1) Explanation of the main points in the visual and 2) Noting of any special implications. More recently, Swales and Feak (1994/2004: 116) have posited the existence of three basic moves: 1) Location of elements and indicative summaries, 2) Highlight of statements, and 3) Discussion of implications, problems, exceptions, recommendations, etc.

clear-cut pattern. Thus, if visual data transfers and commentaries are recognised as typical forms of discourse at all it is thanks to their linguistic micro-realisations, namely to endophoric markers (Hyland 2005) of the type 'In Figure X' or 'As shown in Figure X above/below', to beaconing expressions referring to the variables, parameters and points represented (e.g. 'In the following year we can appreciate a drop in...', 'At x = 3 there is a maximum'), sometimes to a few signalling nouns (Flowerdew 2008) like 'graph/chart', '(graph) line', 'situation', 'performance', 'stagnation', 'stretch' or 'period', and naturally to the physical proximity of the text to the pictorial source.

No matter how far we depart from the concept of pure genre and how elastic our idea of genre may be, the fact is that visual data commentaries rarely occur in isolation (i.e. outside a given matrix genre such as research articles, textbooks, presentations, etc.), and prototypical models are hard to establish due to the ample range of lexico-grammatical features employed to express various degrees of subjectivity, interactivity, factuality and abstraction. If to all this we add the complications and personal imprints engendered by individual discursive preferences concerning focus (e.g. on states/conditions versus processes as textual progressions, or on agents versus spatiotemporal locatives as topic shifters) and style (metaphorical versus literal, nominal versus verbal), as well as by the socio-pragmatic conventions determined by the matrix genre norms (dependent in turn on the intended readership/audience), the culture of origin (i.e. more or less subjectivity, interactivity and formality), and the limitations imposed by the foreign language on ESP learners, then it is easily concluded that inter- and even intra-community diversity outweighs intra-community homogeneity. In relation to this sort of phenomena, Bhatia (1997: 192) and Hyland (2002: 123) coincide in pinpointing one of the key issues in genre research: the need for a shared sense of genre to accomplish understanding and maintain genre integrity, and at the same time for an expanded definition that accounts for genre mixings and embeddings and highlights its dynamic nature.

While most studies on information transfer have dealt with semiotic facets of the verbal-visual interrelations in academic and scientific genres and privilege visual de-codification by the analyst, the present article offers a pedagogical approach that concentrates on the discursive practices, resources and weaknesses of ESP students (i.e. the future – and supposedly current – practitioners) in a university engineering setting, in this case the Technical School of Aeronautical Engineering at Universidad Politécnica de Madrid. Driven by some singular findings during ESP class tasks, this exploratory case study briefly examines the role of graphic verbalisations in the school context and attaches considerable attention to the rhetorical and collocational tendencies of learners, finally resuming the question posed by Busch-Lauer (1998: 112): 'Which method should be taught to teach graphical literacy?' and proposing some didactic directions. Here such literacy is deemed to be a transversal or polyvalent skill across genres and disciplines,

involving analysis and synthesis, paraphrasing, critical thinking (Swales & Feak 1997: 67) and logical deduction. In other words, when facing visual data the user is expected to compare, explain, summarise and present them in a structured way, to differentiate relevant items from details, to draw conclusions, to reflect on the precise function and efficacy of the illustration, and usually to grasp concepts in a non-linear fashion (Johns 1998: 83) by moving back and forth between the graphics and the verbal content in a sort of primitive hypertext (Lemke 1998: 97). This multiple skill, granted the status of *essential* in graduate education by Swales and Feak (1994/2004) more than a decade ago, can be easily inserted into the 'general competences' enunciated by the Common European Framework for Languages (2001/2003: 107–108), as a blend of the study and heuristic abilities forming part of its *savoir-apprendre* or autonomous learning. We may wonder, however, to what extent it can be acquired autonomously.

Curiously enough, very few investigations, manuals and textbooks on academic and professional communication do tackle the translation of visuals into prose. Together with the works by Bertin (1977/1981), Olsen and Huckin (1991) and Swales and Feak (1994/2004, 1997), mentioned above, we can cite those by Fletcher and Hargreaves (1980), Kerridge (1988/1992), and Busch-Lauer (1998). Interestingly, Fletcher and Hargreaves design full verbalisation tasks for peer work as argumentative strategies and provide interpretation models, premises and concise repertoires of useful expressions. Kerridge (1988/1992) gives more weight to the acquisition of rhetorical and lexico-grammatical repertoires: of markers of purpose, topic shift, cause and effect and chronological and contrastive progressions, or the specific vocabulary of trends, including approximators and intensifiers and softeners. By contrast, Busch-Lauer (1998) opens up a window on contrastive rhetoric, especially into the interlingual/cultural differences in the text interface and to the critical contextualised learning by means of authentic tasks, and Bertin (1977/1981: 184–185) and Olsen and Huckin (1991: 172–173) address ethical concerns relative to word-image integration, insisting that to avoid misunderstanding the objectives, the limitations, assumptions and implications of non-verbal materials must never be taken for granted. Finally, Swales and Feak (1994/2004: 112–146) focus on the impact of modality, evidentiality, tense and voice in the qualification of claims and underscore the importance of pragmacultural factors, textuality (i.e. the explicit connection among statements) and critical intelligence (1997: 66–67, 70, 72).

A myriad of other titles within the ESP tradition prioritise instead the opposite task, the conversion of texts into graphs, diagrams and tables (Jordan 1982; Jordan & Nixson 1986/1988; Anderson 1998; Bombardó et al. 2007/2008), either by completing them through verbal instructions – oral or written – or by drafting them from numerical data. Most textbooks give guidelines for an efficient construction

and use of visual aids, dwelling on the advantages and drawbacks of their variants, attending to questions of format, function, layout and typography, facilitating flawed or inappropriate samples for discussion and recommending an awareness of the cultural background of readerships and audiences. In Bertin's line (1977/1981), several sources even touch upon an ethical focus about the responsibilities of the visual communicator for representing the information accurately and honestly (Turk & Kirkman 1982/1996; Anderson 1998; Pickett et al. 2001). Yet the treatment of the text-visual interface is oversimplified and reduced to short answers to comprehension questions (Elsworth 1982; Jordan & Nixson 1986/1988; Anderson 1998), to titling graphs and labelling their parts (Jordan & Nixson 1986/1988), and to condensed classifications of endophorics and synoptic clues for reader guidance at visual interpretation (Anderson 1998). The stylistic aspects most often minded, if any, are the alternative inclusion of explanations and the explicit formulation of conclusions the reader/listener is supposed to draw (Anderson 1998), sentence length and complexity, word order conventions and the use of jargon (Turk & Kirkman 1982/1986), or redundancy avoidance (Bombardó et al. 2007/2008).

On balance then, if Myers (1997: 103) remarks that from secondary school onwards students learn to read visuals along with the written language of the specific subject without any ad hoc training, we must realise that the same should apply to the obverse (i.e. to the transference of visual data into verbal texts) but with three extra difficulties: (a) its command depends in addition on the customary prescriptions for good writing and critical thought, (b) socially agreed standards of data commentary (i.e. generic integrity samples) for pedagogical purposes do not abound, and (c) it is required as an indispensable skill at later stages – at university or during career development – to cope with class discussions and professional presentations, among other genres. It is therefore left to the ESP instructor to systematically teach the techniques and repertoires necessary for the correct verbalisation of visuals, uniting stylistic, communicative and ethical criteria.

2. The study: Context and methodology

To inquire into the role of visuals in the academic routine of aeronautical students, an initial ethnographical approach was made via face-to-face unstructured interviews with staff from three departments in which the use of graphs is essential and occupies from 30 to 70% of class notes: Chemistry, Mechanics and Economics. A major difference between them is that whereas in Chemistry and Economics (subjects taught in the first and last years) all charts are compiled in the class notes booklet published by the school reprography services, the Mechanics teachers (who lecture throughout the first three years) may give additional handouts in

class and create graphs impromptu on the blackboard to explain, demonstrate or exemplify focal points. Another distinctive feature as to the relationship between verbal and non-verbal materials is that in the chemistry and economics booklets visuals tend to illustrate or encapsulate the text, while in mechanics the images habitually present information parallel or complementary to the verbalised content. Consequently, from the above it can be gathered that interpretations and paraphrases of visual data are scarce in all three disciplines, and that mechanics demands a greater deal of effort from the early years owing to the immediacy of class graphs and their impossibility of replication and to the lack of verbal support of those in the printed notes.

Furthermore, the interviews revealed that exam tasks in the three areas are restricted to drawing, marking and indicating, to answering short questions linked to the trends shown in the graph, or to attached theoretical aspects derived from its understanding. This is partly so because (all the teachers interviewed chemists, physicists and economists alike) do admit that to correct paragraph-long answers in examinations is frustrating and time-consuming: they judge students to be 'poor verbalisers' whose unskilled expression is detrimental to proving their knowledge under pressure, hence the emphasis on comprehension. Comprehension is also key to the predominant task involving visuals in engineering: problem solving, which permits a marginal degree of verbalisation, virtually imperceptible in learners (more keen on producing long strings of equations as response) but patent in the 'idiolect' of some teachers, who may intersperse few sentences or even a lean paragraph in the numerical texts publicly displayed as exam correction models on the departmental notice boards. This whole situation confronts us with a vicious circle: a lack of expertise from students due to a lack of practice and vice versa. Quite a challenge for ESP practitioners and the team-teaching methodologies fostered by the European Framework.

This challenge should help LSP trainees encounter frequent verbalising tasks and reflect on the types of meaning inherent to every communicative transaction. Drawing on the useful tripartite taxonomy set forth by Systemic Functional Linguistics (Halliday 1985/1987 and henceforth SFL), the pedagogical use of its semantic metafunctions as broad classifying categories (i.e. ideational or content-based, interpersonal and textual or rhetorical/organisational) may be productive in this sense, as I try to show in the model for data commentary at the end of this chapter. It not only takes into account the expression of views on the selected data but also the interaction with an audience/readership, taking on roles and conveying attitudes and judgments, and the organisation of the message itself to meet the informative expectations of each discipline and within it of its varied genres. For example, a graph may be illustrative in an economics class paper but stand by itself in a chemistry or physics lab report, giving rise to derived or 'satellite' texts, such

as brief comments, chemical formulae or equations. The first step to fight 'poor verbalisations' is then to state these disciplinary and situational outlooks.

In accordance with this, a systemic model may equally get students acquainted with the bi-functionality of ideational and textual devices, which can also lead to interpersonal consequences. It may make them notice that, at the textual level, rhetorical ordering may exert a persuasive effect by anticipating or postponing the conclusion or choosing the type of progression most suitable to their arguments (i.e. chronological, spatial, contrastive or thematic), that abstraction and vagueness, two ideational constituents, dissolve agency, or that explicitness (pervasive in the three metafunctions) denotes commitment (Stubbs 1986). It may, moreover, hone their perception of the differences between their mother tongue and the foreign language they are learning.

In a subsequent phase of the study, fifty-seven students of Aeronautical Engineering were asked to perform two tasks (see Appendix) oriented to qualitatively investigate their tendencies in visual data transfer and commentary: the first one consisted in writing a free paragraph in English (and also voluntarily in Spanish) to verbalise the trend shown on a line graph. Twenty of the informants opted for peer work (being four the maximum number of members per group allowed) and the rest wrote their accounts individually. A total of forty-six samples were obtained and manually examined: eight of them collective and with an appended Spanish version, and the rest individual and in English. The second task, motivated by class observation, comprised a question and a simple activity to identify their individual collocational preferences when qualifying claims by means of the adjectives 'drastic' and 'dramatic'. Both words had been persistently shunned by the learners during the class workshops, even though they were included in a tailor-made glossary for graph verbalisation. In view of this behaviour, the fifty-seven informants were administered a matching exercise (see Appendix) with the first fifty bases of 'drastic' and 'dramatic' in the British National Corpus (BNC) Baby and requested to find suitable collocates. All the participants were native Spanish speakers aged 20–22, with an intermediate level of proficiency in the English language (on average close to the CEF B1-B2 standards) and taking the sixty-hour and English-medium elective 'Technical English'.

3. Findings

The verbalisations collected at the aeronautical school, in accordance with Busch-Lauer's (1998) survey with medical students, avoid explicit commentary and exhibit a deficient metadiscursive repertoire. To adopt Swales and Feak's terms (1997: 65, 70), students' commentaries lack 'textuality' and are confined to mere

'mode transfers' devoid of interpretation[2]: five subjects, in fact, submitted their accounts as outlines or lists of noun phrase descriptors labelling each of the graph stretches. The paragraphed samples, on the contrary, showed varying degrees of metadiscursive awareness. This basically consisted in pronominal and chronological constructions presenting the succession of movements through the represented trend (see Table 1).

Metadiscourse repertoires ranged from an introductory purposive sentence setting the context, which was found in nine cases (see Examples 1–2 on next page), to what might be perceived as manifest reader guidance or a solidarity strategy through pronouns ('you', inclusive 'we' and impersonal 'one') but actually functioning as a covert presentational structure to bring in topic shifts in the evolution of the trend (Examples 3–6). Differently put, these affiliating devices engage the reader while introducing a new graphic movement into the verbal text. Surprisingly, here they are low-frequency features, unlike in the compositions

Table 1. Presentational structures as topic shifters

Structural nature	**Tokens**	**Rate**
(360 trend movements described)		
Pronominal structures		
(Pron + Modal V + V of perception + Trend change)		
with 'you'	1	0.2%
with 'we'	4	1.1%
with 'one'	1	0.2%
Subtotal	**6**	**1.5%**
Chronological structures		
(Time adjunct + Existential V/Alternative V + Trend change)		
with context-bound time markers	185	51.3%
with context-independent time markers	12	3.3%
Subtotal	**197**	**54.7%**
Other introductory formulas		
(Endophorics, relative and adverbial clauses, V-ing clauses, trend change in subject position + 'take place'/'occur'/'come', etc.)		
Subtotal	**157**	**43.7%**

2. It should be remembered that 'mode/data/information transfer' and 'data commentary' are not fully synonymous terms. As Swales and Feak (1997: 64–65) largely stress, every data commentary entails a transfer, but not all transfers do necessarily presuppose evaluative comments. The expression *information transfer* was introduced by Allen and Widdowson in ESP pedagogy as early as 1974.

and research articles in English written by native Spanish speakers, respectively by university students (Neff et al. 2003) and engineering teachers (Sancho Guinda 2003), where they are overused.

(1) 'The graph shows the evolution of sales from 1990 to 1996.'
(2) 'The graphic shows the trend of sales between 1990 and 1996.'
(3) 'There is a stagnation and later, one could see a constant rise in C.' (...)
(4) 'You can appreciate a descent in sales.' (...)
(5) 'We can find a constant performance.' (...)
(6) *'Podemos apreciar que en 1993 las ventas sufrieron unas turbulencias...'* (...) ('We can see that in 1993 sales suffered turbulences'...)

The opening 'panoramic' sentences in (1) and (2) mislead us into expecting a consecutive overall evaluation of the facts depicted, either by anticipating a conclusion and the reporter's stance on the evolution of the variables and/or exposing the limitations of the methodology, a move very much praised by Swales and Feak (1997: 71) but thwarted straight off in my samples with statements paraphrasing each graph stretch in a redundant relationship. These concrete statements follow unanimously the chronological progression dictated by the time variable on the x axis and ruled by temporal markers, which also work as 'veiled presentationals' to make up for a too choppy syntax abusing coordination and juxtaposition. True, context-bound chronological signposters of the type Time period + Existential verb/Alternative presentational structures are extremely abundant (slightly over 54%) among all the trend changes described (e.g. 'In.../'In the year...', 'After...', 'Between...', 'From...to...', 'During/For...', 'At the beginning/end of...', etc. + 'there is/are/was/were' or 'is/are', 'appears/appeared', 'take(s)/took place', 'we can see/it can be seen ...'). They can be regarded as fixed collocational patterns and prospective informative chunks (Langacker 2008: 460–461), since they supply the co-text for other collocations to occur and most often thematise the variable 'sales' in the subject position. These constructions are also used in Spanish (see Examples 7–9):

(7) 'In the first year sales decreases till they flatten out during the next year. Since then, they begin to grow, pass through a fluctuation performance and reach a maximum. After that sales experiment a dramatic fall and a small decline in 1995.'
(8) 'From 1990 to 1991 sales suffered a weak decline and then the situation became stable. In 1992 sales grew in a steady way until 1993 when the situation turned unstable for a period of one year. After that, sales raised again but near 1995 a dramatic slump took place and nowadays the recovery is very slow.'

(9) *'Las ventas empezaron con una tendencia decreciente, que terminó en 1992. Entonces comenzaron a crecer, experimentando un periodo inestable entre 1993 y 1994, hasta que alcanzaron su máximo entre 1994 y 1995. En ese año, las ventas cayeron bruscamente, y no fue hasta 1996 cuando empezaron a recuperarse.'*
(English version by student: 'The sales started with a decreasing tendency, that ended in 1992.Then they began to increase, experiencing an unstable period between 1993 and 1994, until they reached their maximum between 1994 and 1995. In that year the sales dropped heavily and it was not before 1996 that they began to recover.')

Context-independent connectors (e.g. 'first', 'then', 'later', 'next', 'afterwards', 'finally') are far less resorted to (amounting only to 3.3% of connective tokens) as they do not lend themselves so much to an economic presentational function: their time demarcations are too vague and do not exempt the writer from itemising more accurate periods (e.g. 'Finally, from 1996 sales started to recover'). In any case, the presentational verbs preceded by both context-bound and independent time markers (temporal specificity apart) display a relative but disperse variation. The most widely used ones are 'there + be' (30 tokens), 'reach' (12 instances) and the simplifications 'sales have a.../sales are' + adjective (12 occurrences), followed by the intransitive group formed by 'happen', 'come', 'take place' and 'occur' (16 occurrences in total) and the inchoatives 'began/start to' + verb (totalling 17 cases). Special mention deserve more creative options such as the metaphorical 'sales' + 'experience'/ 'experiment'/ 'see'/ 'live'/ 'pass through'/ 'suffer' (26 tokens), or the result indicators 'result in', 'end in', 'lead to', 'culminate in', 'turn/become into' + noun/adjective, 'arrive at', 'give way to' and '*desembocar*' (14 uses).

Last, external viewpoint markers (Hyland's 2005 endophorics) complete the repertoire and bear evidential meaning (12 tokens): 'The graph shows a...'/'As can be seen in the graph...' and the tandem Personal pronoun + Verb of perception (e.g. 'we/you/one can see/appreciate/find...'), already commented. All of these items tend to be symmetrically present in the L1 texts and, strikingly, the superfluous repetition of the topical variable ('sales'/*ventas*) and the disjointed syntax, reminiscent of children's narratives, are also detected in the Spanish versions (see Examples 10 and 11 from different writers; my emphasis), in spite of the greater pronominal flexibility of the Spanish language to incorporate agency in verbal suffixes and reintroduce antecedents in number and gender-marked pronouns and determiners. In contrast, reports in the L2 show a more frequent reiteration of complement forms (e.g. 'in/of sales', see Example 11), which generates a similar imbalance between given and new information (Hoey 1991/1996: 134–135) that hinders comprehension:

(10) '*La gráfica nos muestra* las ventas *durante 1990 y 1996. Durante 1990* las ventas *decrecieron. El año siguiente no variaron. De 1992 hasta el final de 1994,* las ventas *aumentan de manera uniforme salvo en 1993, donde fueron inestables. Al final de 1994 alcanzan la cima.* Las ventas *disminuyen desde la cima hasta 1996, primero rápidamente y después más lentamente. En 1996* las ventas *mejoran.*'

(11) 'In the 1990 you can appreciate a descent *in sales*. The next year *the sales* didn't suffer any change, but in 1992 *the sales* increased until a level that was superior than the initial one in 1990. 1993 is an unstabilic year *in sales,* with strong increases and decreases. In 1994 it was reached the maximum level in the analyzed period, but at the end of this year *the sales* suffered the worst fall. In 1995 *sales* rised again.'

So far we have seen how scanty metadiscursive repertoires based exclusively on chronology and the dispersion of alternative presentational constructions may generate a repetitious style and induce writers to recycle time markers into handy topic shifters, when nevertheless a focus combining overt comparison and cause and effect relationships could be viable. In passing we have also mentioned that the samples scrutinised lack interpretation, and at this point it must be clarified that they do contain evaluation but implicit in their lexical choices. Lexis is a proven evaluative device (Thompson & Hunston 2000: 6, 14, 21–22) and may embrace metaphors (e.g. 'sales suffered turbulences', 'sales sank', *'florecimento financiero'*), hyperbole (e.g. 'the worst crisis in the company', *'un máximo histórico'*), very occasionally disjuncts (e.g. 'unfortunately') or even 'plot-building', sketchy storylines in which fictional and anonymous company members struggle to get over a critical financial period (e.g. 'since then they are trying to overcome the crisis'). It is noteworthy that to implement this lexical evaluation aeronautical students markedly choose generic or superordinate common-core terms (see Table 2) – that is, nominal or verbal bases such as to 'increase/decrease', 'grow', 'fall' and their corresponding nouns and adjectives instead of more domain-specific vocabulary deeply rooted in the description of trends and mathematical functions (e.g. 'peak', 'maximum', 'gradient', 'slump', 'plummet', 'linear', 'sinusoidal function'), though some of them derive from the general language.

According to Bliss (2001: 16), university students are in general asked by faculty to produce two types of writing: descriptive and persuasive, both vital to and coexisting in data commentary. The evaluative lexis referred to in the previous lines embodies the convergence of the three SFL metafunctions to fulfil this double aim: metaphors, for example, qualify data ideationally whilst appealing to emotion in order to convince readers of the validity of the data sorting and ultimately of the possible arguments it helps put forward. Their adjectival premodification, as that

of any other noun under which data are designated, also hints at the writer's text reflexivity and assumptions about the register used (academic in this case) and about his/her alignment with the practices of the community. It corroborates that Academic English is in the main believed to be precise, elaborate, highly nominalised and lexically dense (Swales & Feak 1994/2004: 22–24; Hyland 2006: 13–14). A second remarkable tendency is the students' inclination towards the phrasal structure Adjective + Noun over Verb + Adverbial/Prepositional phrase, which makes their style more nominal than actional (see Table 3 for a more detailed structural breakdown). This is another instance of the convergent SFL metafunctions that may be fruitfully exploited to raise students' awareness.

To conclude, a further collocational pattern was disclosed with regard to the use of 'drastic' and 'dramatic' as qualifying devices. By and large, results ran counter to the BNC list and may be summarised in three central points (see Table 4).

1. A reinforcement of the heavy affective load of bases with a strong negative connotation (e.g. 'DRAMATIC' + 'fall', 'collapse', 'deterioration', 'failure').

Table 2. Semantic nature of the nouns and verbs employed in all samples

	Rate	Tokens	Examples
Generic	80.8%	273	increase, fall, *recuperarse, caer*
Specific	19.2%	65	gradient, peak, slump, *máximo*, boom

Table 3. Structural nature of the nouns and verbs used in all samples

Structural nature	Rate	Tokens	Examples
Unmodified single verb	21.5%	78	increased, improve, *suben, se normaliza*
Unmodified verb + D_o	7.4%	27	reached the top, *alcanzar su valor máximo*
Unmodified noun	12.1%	44	recovery, peak, *recuperación, máximo*
Modified noun (Adj+N/N+Adj)	47.7%	173	very noticeable increases, moderate dip, *descenso leve*
Modified verb (V+Adv/PrepPh/Adj)	14.6%	41	dropped heavily, *permanecieron invariables*
Subtotal of modified terms	**57.4%**	**214**	big drop, slight decrease, *pequeña mayoría, importante caída*
Subtotal of unmodified terms	**42.6%**	**159**	dropped, growth, peak *crecimiento, descenso, bajadas*
TOTALS	**100%**	**373**	

2. A reproduction of highly idiomatic collocations from the L1 journalistic register, which involves 'neutral' words (e.g. 'DRASTIC' + 'change', 'reduction', 'action') and intermittently positively-loaded ones (e.g. 'solution').
3. A disregard of base synonymy, thus generating asymmetric results for synonyms like 'change' and 'shift' or 'reduction' and 'cut'.

Table 4. Collocational tendencies of informants with *DRASTIC* and *DRAMATIC*
n = 57 informants

Base	***DRASTIC***		***DRAMATIC***	
	Rate	**Tokens**	**Rate**	**Tokens**
Action	73.6%	42	33.3%	19
Break	35.0%	20	49.1%	28
Case	21.0%	12	49.1%	28
Change	85.9%	49	35.0%	20
Collapse	22.8%	13	42.1%	24
Cut	57.8%	33	26.3%	15
Decline	28.0%	16	50.8%	29
Deterioration	24.5%	14	64.9%	37
Display	26.3%	15	24.5%	14
Effect	33.3%	19	63.1%	36
Evidence	24.5%	14	45.6%	26
Expansion	52.6%	30	24.5%	14
Failure	17.5%	10	78.9%	45
Fall	47.3%	27	71.9%	41
Growth	47.3%	27	38.5%	22
Improvement	59.6%	34	12.2%	7
Measure	64.9%	37	12.2%	7
Order	66.6%	38	8.7%	5
Performance	24.5%	14	50.8%	29
Policy	52.6%	30	14.0%	8
Punishment	56.1%	32	38.5%	22
Reduction	82.4%	47	35.0%	20
Remedy	68.4%	39	19.2%	11
Rise	59.6%	34	28.0%	16
Shift	38.5%	22	19.2%	11
Solution	84.2%	48	24.5%	14
Success	33.3%	19	35.0%	20
Violation	7.0%	4	66.6%	38

Table 5. Cognitive associations of informants involving *DRASTIC* and *DRAMATIC*
n = 57 informants

Collocate	Upward movement connotation		Downward movement connotation		Horizontal connotation	
	Rate	Tokens	Rate	Tokens	Rate	Tokens
DRASTIC	84.2%	48	38.6%	22	26,3%	15
DRAMATIC	22.8%	13	75.4%	43	29.9%	17

The second matching task discovered the conceptual associations underlying these patterns: students preferably related 'drastic' with ascending processes and abrupt interruptions along a horizontal axis and 'dramatic' with affect and downward verticality (see Table 5), since the emotional (negative) connotation of the word prevented them from applying it to the description of growth or rising variables. This suggests that indeed there is still much collocational research to be done with the aid of Cognitive Linguistics and Cultural Pragmatics.

4. Final remarks: A critical systemic teaching of graph commentary in ESP contexts

In light of the difficulties and findings reported in this first-approach case study, the verbalisation of visual data points to suggestive lines of research and virtually untapped classroom approaches enabling technical teachers to instruct and test their students in more critical, creative and realistic ways. For instance, they may achieve gradual training in data commentary by inviting them to make predictions and justify their identifications of the visuals with certain behaviours related to materials (e.g. alloys in Chemistry), physical phenomena (e.g. types of movement in Mechanics) or economic trends. They may linger on them as part of the problem-solution progression typical of case studies, and use them as tools for periodic assessment as class exercises, teamwork discussions or seminar activities. These may prove convenient for ESP practitioners as well, who can benefit from an integration of contrastive rhetoric and disciplinary discourses, encompassing the notions of genre, strategic move and micro-linguistic realisations – for example at a collocational level.

Here we have learnt that aeronautical engineering students need explicit training in interpretive data commentary, particularly in the handling of metadiscursive items and presentational structures, so necessary to weave textual coherence and cohesion. We have also been informed that they deploy compensatory presentational strategies with familiar elements additionally imposed by the

variables involved (i.e. chronological expressions as prospective transition markers or subjectification, as in 'sales suffered a drop'), which indicates their rhetorical and grammatical gaps account for their choices. To bridge those gaps I propose below a tentative inventory of decisions concerning data interpretation, revolving around the three Hallidyan metafunctions along a continuum with an embedded cline for each sub-option, often overlapping, involving metadiscourse and underlining the dynamic nature of any communicative act. With this functional orientation learners may gain a more complete picture of language in use, conceiving meaning as a system of intertwined choices operating simultaneously and the text as its delivery agent. Graph commentaries may be then presented as multi-skill and cross-disciplinary practices that entail a series of decisions and repercussions on the ideational, interpersonal and textual planes of language. In the late 1990s Swales and Feak (1997) and Miller (1998) already noticed the import of the free-will component and this threefold action of data rendering.

Continuum of data commentary choices

1. Ideational component
    a. Focus
        i. State/condition (nominal) vs. process (verbal)
        (e.g. 'The graph shows a moderate drop' vs. 'Sales dropped moderately')
    b. Abstraction
        i. Common-core vs. domain-specific terms (e.g. 'improve' vs. 'upturn')
        ii. Generic (superordinate) vs. specific (e.g. 'rise' vs. 'soar')
        iii. Vagueness vs. accuracy
            1. Quantifying refinement (e.g. 'somewhat/far/way/a great deal/a lot bigger' vs. 'twice as big')
            2. Indication of the degree and speed of change (e.g. 'a drop' vs. 'a significant sudden drop')
    c. Factuality
        i. Present vs. past tenses + durative vs. perfective aspects – nuances of habit, change, isolated fact, completion: (e.g. 'Sales drop' vs. 'Sales are dropping' vs. 'Sales dropped' vs. 'Sales have dropped' vs. 'Sales have been dropping')
        ii. Conditionals (hypothesis, prediction) (e.g. 'If sales keep falling...')
2. Interpersonal component
    a. Subjectivity
        i. Interpretation vs. plain description/transfer

1. Neutrality vs. grammatical emphasis (e.g. 'In 1995...' vs. 'It was not until 1995...')
2. Metaphorical vs. literal expression (e.g. 'rise dramatically' vs. 'rocket')
3. 'Affect-free' vs. value-laden terms – often metaphorical (e.g. 'a quick and very deep fall' vs. 'a collapse')
4. Modalisation vs. non-modalisation of quantifications and qualifications (e.g. hedges, boosters, evidentials, endophorics)

b. Interactivity (interpersonal or interactional metadiscourse, Hyland 2005)
   i. Non-personalisation vs. personalisation through pronominal reference (e.g. 'you', 'one', inclusive 'we')
   ii. Reader/listener guidance (e.g. transition markers, topic shifters, code glosses, endophorics, engagement markers)

3. Textual component
   a. Dominance (verbal text larger than visuals or vice versa)
   b. Interface *visual/verbal*
      i. Redundancy/paraphrase
      ii. Additive/complementary
      iii. Inclusive/illustrative
   c. Possible moves
      i. Panoramic purpose of the visual representation (e.g. 'This graph shows the evolution of sales during the period...')
      ii. Overall framing evaluation of data (e.g. 'progressive/regressive', etc.: 'There has been a regressive trend in sales during the year')
      iii. Detailed evaluation of data (grouping + examination of group features + inter-group comparison + study of special cases)
      iv. Discussion of implications in broader contexts
   d. Progression (textual or interactive metadiscourse, Hyland 2005)
      i. Chronological (e.g. time and stage sequencers)
         1. Cause and effect (e.g. cause and effect markers)
      ii. Spatial (e.g. locative expressions)
      iii. Contrastive (e.g. markers of comparison and contrast)
      iv. Topical or thematic (e.g. signaling nouns & verbs denoting states, conditions and movements)
      v. Inductive vs. deductive or specific vs. general and vice versa
      vi. Problem → solution

This continuum should of course be inserted in the larger frameworks of national and disciplinary cultures, which shape variable degrees of interventionism in the interpretation of data and favour or discourage their climatic or anti-climatic

presentation. Rhetorical differences depend on the practices and expectations (often tacit) of social relationships within communities, which mostly have to do with idiosyncratic face issues (e.g. see the work of Connor 1996 and Panetta 2001) and the argumentative strategies intrinsic to the field of knowledge at stake and its particular genres. For example, Swales and Feak (1997: 66) let us know that Asian students tend to refrain from commenting on their visual aids as a sign of respect, since explaining the obvious would constitute a face-threatening act for the reader or interlocutor. Likewise, Hyland (2004, 2009) reminds us of the differences between the scientific and humanistic discourses (e.g. more or less empiricism and conversational style and different types of argument) and delves into some key genres and their most outstanding features: the balance between criticism and praise in book reviews and of credibility and promotion in abstracts, the prudence, solidarity and priority marking in scientific letters and research articles, and the exhibition of expert identities characteristic of textbooks. Learners should be aware, in sum, that the cluster of decisions sorted out in the continuum provided here is not, as Hyland (2004: 11) holds with reference to the common practices of academic discourse, a mere matter of personal stylistic preference but of community-recognised ways of expressing stance and managing communicative interactions, and that those communal conventions are governed by the subtle interplay of culture, discipline and genre.

Acknowledgements

I want to express my deepest thanks to the teaching staff at the Madrid School of Aeronautical Engineering (Universidad Politécnica de Madrid) for their stimulating conversations and useful materials, especially to Professors González Prolongo and Salom Coll (Chemistry Department), Ramis Abril and García-Pelayo Novo (Physics Department), and González García (Department of Business Administration and Statistics). Thanks too to Ms. Jiménez Ruiz (Chemistry Department) and all the second-year students who volunteered to participate in this research, as well as to the reviewers of this paper for their helpful comments.

References

Allen, J.P.B. & Widdowson, H.G. 1974. *English in Physical Science*. Oxford: OUP.

Anderson, P.V. 1998. *Technical Communication: A Reader-centered Approach*. Orlando FL: Harcourt Brace College Publishers.

Bertin, J. 1981 [1977]. *Graphics and Graphic Information Processing*. Berlin: Walter de Gruyter.

Bhatia, V.K. 1993. *Analysing Genre: Language Use in Professional Settings.* London: Longman.
Bhatia, V.K. 1997. Genre-mixing in academic introductions. *English for Specific Purposes* 16(3): 181–195.
Bhatia, V.K. 1999. Integrating products, processes, purposes and participants in professional writing. In *Writing: Texts, Processes and Practices,* C.N. Candlin & K. Hyland (eds), 21–39. London: Longman.
Bliss, A. 2001. Rhetorical structures for multilingual and multicultural students. In *Contrastive Rhetoric Revisited and Redefined,* C.G. Panetta (ed.), 15–30. Mahwah NJ: Lawrence Erlbaum Associates.
Bombardó, C., Aguilar, M. & Barahona, C. 2008 [2007]. *Technical Writing: A Guide for Effective Communication.* Barcelona: Edicions UPC.
British National Corpus Baby. <http://www.natcorp.ox.ac.uk>
Busch-Lauer, I.A. 1998. Non-verbal elements and data commentary in English medical texts. In *Genre Studies in English for Academic Purposes,* I. Fortanet, S. Posteguillo, J.C. Palmer & J.F. Coll (eds), 109–132. Castelló: Publicacions de la Universitat Jaume I.
Connor, U. 1996. *Contrastive Rhetoric: Cross-cultural Aspects of Second-language Writing.* Cambridge: CUP.
Council of Europe/Conseil de l'Europe. 2003 [2001]. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: CUP.
Cros, A. 2003. *Convencer en clase: Argumentación y discurso docente.* Barcelona: Ariel Lingüística.
Elsworth, S. 1982. *Count Me In: Understanding Numbers in English.* Harlow: Longman.
Fletcher, M. & Hargreaves, R. 1980. *Defining and Verbalising.* London: Evans.
Flowerdew, J. 2008. From noun to clause to discourse: Reconciling clausal and textual properties of signalling nouns. Plenary lecture given at the *INTERLAE International Conference – Interpersonality in Written Academic Discourse: Perspectives across Languages and Cultures*, Jaca, 11–13 December 2008.
Halliday, M.A.K. 1987 [1985]. *An Introduction to Functional Grammar.* London: Edward Arnold.
Hoey, M. 1996 [1991]. *Patterns of Lexis in Text.* Oxford: OUP.
Hyland, K. 2002. Genre: Language, context and literacy. *Annual Review of Applied Linguistics* 22: 113–135.
Hyland, K. 2004. *Disciplinary Discourses: Social Interactions in Academic Writing.* Ann Arbor MI: University of Michigan Press.
Hyland, K. 2005. *Metadiscourse. Exploring Interaction in Writing.* London: Continuum.
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.
Johns, A. 1998. The visual and the verbal: A case study in macroeconomics. *English for Specific Purposes* 17(2): 183–197.
Jordan, R.R. 1982. *Figures in Language: Describe and Draw. A Workbook of Communicative Activities.* Glasgow: Collins ELT.
Jordan, R.R. & Nixson, F.I. 1988 [1986]. *Language for Economics: Integrated Study Skills and Advanced Language Practice.* London: Collins ELT.
Kerridge, D. 1992[1988]. *Presenting Facts and Figures*. London: Longman.
Kress, G., Jewitt, C., Ogborn, J. & Tsatsarelis, C. 2001. *Multimodal Teaching and Learning: The Rhetorics of the Science Classroom.* London: Continuum.
Langacker, R.W. 2008. *Cognitive Grammar: A Basic Introduction.* Oxford: OUP.

Lemke, J.L. 1998. Multiplying meaning: Visual and verbal semiotics in scientific text. In *Reading Science: Critical and Functional Perspectives on Discourses of Science,* J. Martin & R. Veel (eds), 87–113. London: Routledge.

Lynch, M. & Woolgar, S. (eds). 1990 [1988]. *Representation in Scientific Practice.* Cambridge MA: The MIT Press.

Martin, J.R. 1995. Text and clause: Fractal resonance. *Text* 15: 5–42.

Miller, T. 1998. Visual persuasion: A comparison of visuals in academic press and the popular press. *English for Specific Purposes* 17(1): 29–46.

Murphy, K.P. 2001. Teaching as persuasion: A new metaphor for a new decade. *Theory into Practice* 40(4): 224–227.

Myers, G. 1997. Words and pictures in a Biology textbook. In *Functional Approaches to Written Text: Classroom Applications,* T. Miller (ed.), 93–104. Washington DC: United States Information Agency.

Myers, G. 2003. Words, pictures and facts in academic discourse. *Ibérica* 6: 3–13.

Neff, J., Dafouz, E., Herrera, H., Martínez, F., Rica, P. & Sancho, C. 2003. Contrasting learner corpora: The use of modal and reporting verbs in the expression of writer stance. In *Extending the Scope of Corpus-based Research: New Applications, New Challenges,* S. Granger & S. Petch-Tyson (eds), 211–230. Amsterdam: Rodopi.

Olsen, L.A. & Huckin, T. 1991. *Technical Writing and Professional Communication.* New York NY: McGraw-Hill.

Paltridge, B. 2002. Genre, text type, and the English for Academic Purposes (EAP) classroom. In *Genre in the Classroom,* A.M. Johns (ed.), 69–88. Mahwah NJ: Lawrence Erlbaum Associates.

Panetta, C.G. (ed.). 2001. *Contrastive Rhetoric Revisited and Redefined.* Mahwah NJ: Lawrence Erlbaum Associates.

Pickett, N.A., Laster, A.A. & Staples, K.E. 2001. *Technical English: Writing, Reading and Speaking.* London: Longman.

Sancho Guinda, C. 2003. Contractual role of modality as convergence strategy in technical research articles. *RESLA* 16: 219–233.

Stubbs, M. 1986. A matter of prolonged field work: Notes toward a modal grammar of English. *Applied Linguistics* 7: 1–25.

Swales, J. 1990. *Genre Analysis: English in Academic and Research Settings.* Cambridge: CUP.

Swales, J. & Feak, C.B. 2004[1994]. *Academic Writing for Graduate Students: Essential Tasks and Skills.* Ann Arbor MI: University of Michigan Press.

Swales, J. & Feak, C.B. 1997. From information transfer to data commentary. In *Functional Approaches to Written Text: Classroom Applications,* T. Miller (ed.), 64–76. Washington DC: United States Information Agency.

Thompson, G. & Hunston, S. 2000. Evaluation: An introduction. In *Evaluation in Text: Authorial Stance and the Construction of Discourse,* S. Hunston & G. Thompson (eds), 1–27. Oxford: OUP.

Turk, C. & Kirkman, J. 1996 [1982]. *Effective Writing.* London: E & FN Spon.

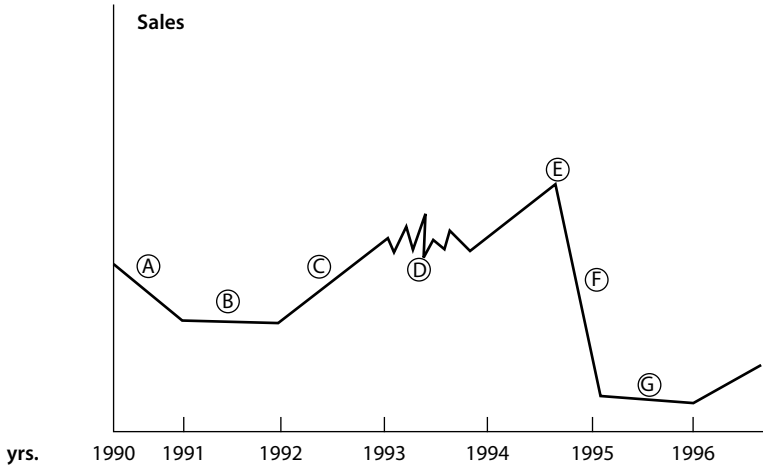
Appendix

Task 1

Description of graphic information in English

Prof. Sancho Guinda
Dpt. of Applied Linguistics
ETSI Aeronáuticos-UPM

Verbalise the graph content in a single paragraph



Task 2

Questionnaire on graphic information descriptors in English

Prof. Sancho Guinda
Dpt. of Applied Linguistics
ETSI Aeronáuticos-UPM

A. Match now *DRASTIC* and *DRAMATIC* with the words below provided (mark only A or B). You may match both terms with the same word.

1. Display _____
2. Collapse _____
3. Failure _____
4. Cut _____
5. Violation _____
6. Solution _____
7. Measure _____
8. Deterioration _____

9. Punishment _____
10. Order _____
11. Reduction _____
12. Improvement _____
13. Change _____
14. Evidence _____
15. Effect _____
16. Performance _____
17. Expansion _____
18. Shift _____
19. Fall _____
20. Break _____
21. Decline _____
22. Rise _____
23. Case _____
24. Policy _____
25. Action _____
26. Remedy _____
27. Success _____
28. Growth _____

B. With what image would you associate the terms *DRASTIC* and *DRAMATIC*? Indicate numbers (more than one association per term is allowed).

Image 1. 

Image 2. 

Image 3.

DRASTIC _____

DRAMATIC _____

SECTION TWO

Research based on meta-analysis and applications in LSP

Some dichotomies in genre analysis for Languages for Specific Purposes

John Flowerdew
City University of Hong Kong

Genres are staged, structured, communicative events, motivated by various communicative purposes, and performed by members of specific discourse communities (Swales, 1990; Bhatia, 1993, 2004; Berkenkotter and Huckin, 1995; Johns, 1997). Since its inception with the two seminal works on the topic by Swales (1990) and Bhatia (1993), genre analysis has taken pride of place in much of the ongoing research in languages for specific purposes (LSP). The goal of much of this research is pedagogic, the understanding being that good genre descriptions can feed into pedagogy in the form of syllabus and materials design. This chapter considers four dichotomies for genre research. These dichotomies are: 1. Individual genres vs genre networks; 2. Written vs spoken genres; 3. Macro vs micro levels of analysis; and 4. Move structure vs. lexico-grammar. Each of these dichotomies has important implications for LSP pedagogy, not just the last, and these are highlighted in the course of the chapter.

Keywords: Genre; genre analysis; genre networks; move structure; lexico-grammar; languages for specific purposes; LSP.

1. Introduction

Since its inception with the two seminal works on the topic by Swales (1990) and Bhatia (1993), genre analysis has taken pride of place in much of the ongoing research in languages for specific purposes (LSP). One only has to look at almost any issue of *English for Specific Purposes Journal*, the leading journal in the field, to see how predominant this approach to the analysis of specialist language is. The goal of much of this research is pedagogic, the understanding being that good genre descriptions can feed into pedagogy in the form of syllabus and materials design (e.g. Bhatia 1993; Flowerdew 1993; Swales & Feak 2000, 2004; Paltridge 2001; Johns 2002; Hyland 2004). Hyon (1996) refers to three schools of genre analysis: the ESP school, the Systemic Functional school and the New Rhetoric school. My point of

departure and main focus in this chapter, given that the theme of this volume is LSP, is the ESP school, but I will make reference to the other approaches in the course of this chapter. Given this orientation, I will begin with a working definition of genre from an ESP perspective, as follows. Genres are staged, structured, communicative events, motivated by various communicative purposes, and performed by members of specific discourse communities (Swales 1990; Bhatia 1993, 2004; Berkenkotter & Huckin 1995; Johns 1997). This is rather dense and concise, so let me expand somewhat with a list of characteristics, as provided by Bhatia (2004: 22–3):

1. Genres are recognizable communicative events, characterized by a set of communicative purposes identified and mutually understood by members of the professional or academic community in which they regularly occur.
2. Genres are highly structure and conventionalized constructs, with constraints on allowable contributions not only in terms of the intentions one would like to give expression to and the shapes they often take, but also in terms of the lexico-grammatical resources one can employ to give discoursal values to such formal features.
3. Established members of a particular professional community will have a much better knowledge and understanding of the use and exploitation of genres than those who are apprentices, new members or outsiders.
4. Although genres are viewed as conventionalized constructs, expert members of the disciplinary and professional communities often exploit generic resources to express not only "private" but also organizational intentions within the constructs of "socially recognized communicate purposes".
5. Genres are reflections of disciplinary and organizational cultures [...]
6. All disciplinary and professional genres have integrity of their own, which is often identified with reference to a combination of textual, discursive and contextual factors.

I will return to several of these aspects of genre throughout this chapter.

One of the most influential features of genre analysis for ESP has been generic moves (Swales 1990; Bhatia 1993), referred to as schematic structure by the Sydney School (Martin 1984). This is what is meant in describing genres as "staged" and "structured" in my definition earlier and is covered in point 3 by Bhatia above. Basically, the idea is that, in performing a genre, speakers or writers go through a more of less predictable set of communicative acts (described as moves and steps by Swales and Bhatia) which have recognizable (although usually variable) linguistic forms. Let us look at some published examples of genres analyzed according to their move structure. Probably the best known model and the one which started it all off was Swales's *Create a Research Space* (CARS) pattern (Swales 1990, 2004) for the introductions to research articles. Swales (1990) specified the

Establishing a territory
  Claiming centrality and/or
  Making topic generalization(s) and/or
  Reviewing items of previous research
Establishing a niche
  Counter-claiming or
  Indicating a gap or
  Questions raising or
  Continuing a tradition
Occupying the niche
  Outline purposes or
  Announcing present research
  Announcing principal findings
  Indicating RA structure

Figure 1. Move structure for introductions to academic research articles (Swales, 1990)

preceding series of moves and component steps for the introductions to academic research articles (Figure 1).

Similarly, Bhatia (1993) proposed the following structure of moves and steps for sales letters in business communication: 1. Establishing credentials; 2. Introducing the offer (Offering the product or service, and/or Essential detailing of the offer and/or Indicating value of the offer); 3. Offering incentives; 4. Enclosing documents; 5. Soliciting response; 6. Using pressure tactics; 7. Ending politely.

Another, less influential (by far) move structure was presented in a paper I myself wrote with Alina Wan (Flowerdew &Wan 2006), to describe tax computation letters: 1. Opening salutation; 2. Subject; 3. Actions taken; 4. Discussion of issues; 5. Solicit action; 6. Express availability; 7. Closing salutation.

And here is another one I did with Tony Dudley-Evans on the generic structure of editorial letters (Flowerdew & Dudley-Evans, 2002) (Figure 2).

Hyland (2007) has a very accessible, simple example for the very stereotypical genre of the dissertation acknowledgement, showing how move structure maps out onto an actual text (Figure 3). I should stress that this is extremely formulaic and that most genres are much more complex than this. But it is very suitable for showing how form can map onto function.

So, this provides you with a quick idea of the sort of thing that you can do regarding move structure in genre analysis. Some of these move structures I will come back to later.

In spite of the great amount of work that has been and continues to be done using the ESP approach to genre analysis, there are a range of theoretical issues which remain problematic. In this chapter I will address four of these issues in

1. Prepare reader for decision*
1. refer to submission* and/or
2. apologise for delay and/or
3. interpret referees' reports

2. Convey decision*
either
1. accept or
2. accept as a research note or
3. offer resubmission or
4. reject (± mitigate; ± justify)

3. Make recommendations for revision #
either
1. refer reader to reviewers' recommendations and/or
2. make editorial recommendations

4. Sign off *
either
1. confirm decision and/or
2. mitigate bad news and/or
3. apologise for delay and/or
4. refer to enclosure and/or
5. refer to personal matters and/or
6. present a deadline and/or
7. suggest further contact and/or
8. give encouragement

* obligatory # optional 1

Figure 2. Move structure for editorial letters (Flowerdew & Dudley-Evans, 2002)

Move	Example
1. Reflecting move	The most rewarding achievement in my life, as I approach middle age, is the completion of my doctoral dissertation
2. Thanking move	
2.1. Presenting paticipants	During the time of writing I received support and help from many people.
2.2. Thanks for academic help	I am profoundly indebted to my supervisor, Dr Robert Chau who assisted me in each step to complete the thesis
2.3. Thanks for resources	I am grateful to The Epsom Foundation whose research travel grant made the field work possible and to the library staff who tracked down elusive texts for me.
2.4. Thanks for moral support	Finally, thanks go to my wife who has been an important source of emotional support.
3. Announcing move	However, despite all this help, I am the only person responsible for errors in the thesis.

Figure 3. Move structure for dissertation acknowledgements (Hyland, 2007)

terms of dichotomies. These dichotomies are not mutually exclusive pairings, but are presented as a useful way of thinking about important issues. The four dichotomies are as follows: 1. Individual genres vs. genre networks; 2. Written vs. spoken genres; 3. Macro vs. micro levels of analysis; 4. Move structure vs. lexicogrammar. I should point out that, in presenting these dichotomies, my purpose is not to discredit genre theory, but to highlight some areas where this very productive theory might benefit further theoretical and empirical research, with a view to further enhancing its productivity for praxis.

2. Dichotomy 1: Individual genres vs. genre networks

By the term network, I refer to how an individual instance of a genre can relate to other genres (Swales 2004 attaches a rather different meaning to the term, see below). Devitt (1991) talks about the *genre set*, a term which she uses to refer to a range of text genres which a professional group uses in the course of their daily routine e.g. a conference presentation, a poster, and a research article in the case of academics. Bazerman (1994: 97) talks about *systems of genres*. A genre system, for Bazerman, is a full set of texts from a particular genre which constitute a complete interaction (e.g. a complete exchange of letters). Raisanen (2002) refers to genres sets and genre systems, but she also exemplifies *genre chains*, which are chronologically related sequences of genres in a given interaction. The following is a simplified version of the genre chain for a conference paper, as illustrated by Raisonen (2002), showing how other genres precede and follow the conference paper itself: 1. Call for abstracts > 2. Conference Abstract > 3. Review Process [Acceptance] > 4. Instructions > 5. Conference paper draft > 6. Review process [Acceptance] > 7. Revised conference paper > 8. Review Process > 9. Published conference paper > 10. Oral presentation.

All three of these different, but related, concepts are important for LSP pedagogy. The genre set allows the learner to see the similarities and differences in move structure and linguistic realisation patterns across different genres in a particular field. The genre system allows the learner to see the similarities and variations in move structure and linguistic realisation patterns within one particular interaction. The genre chain also focuses on one interaction as it develops over time, but through different genres. Working with genre networks is, of course, closer to real life than dealing with individual instances of genres and may be closer to the target activities of an LSP programme than dealing with individual genres in isolation.

In an LSP context, working at this level highlights the role of intertextuality – how there are references in one text to other texts (Kristeva 1980; Bakhtin

1981, 1984). Intertextuality may take various forms. Fairclough (1992: 117) distinguishes between *manifest intertextuality* – quotation, citation, and paraphrase – and *constitutive intertextuality* – (generic) features which do not leave an obvious trace from the source. All of these aspects of intertextuality are important for LSP. There is clearly work to be done by the LSP course designer, teacher and learner in raising to consciousness these features of genre, especially the more "hidden" constitutive variety, which may not be obvious. While the concepts of genre set, genre system and genre chain allow the course designer, teacher and learner to identify intertextuality between and across genres and instances of genres, much current genre-based pedagogy focuses on individual texts and this notion of intertextuality is lost.

A further notion we can classify under the umbrella of genre network, in addition to genre set, genre system, and genre chain, is that of *disciplinary genre* (Bhatia 2004: 54), which includes all those genres associated with a profession (not just those involved in a particular individual's sphere of activity (genres system) or specific activity (genre set and genre chain). Swales (2004: 22) refers to this as *genre network* – "the totality of genres available for a particular sector (such as the research world), as seen from any chosen synchronic moment". Disciplinary genre and genre network (in Swales's sense of the term) refer to a more abstract concept than the preceding three, in so far as it may not relate to the life world of individuals. But it is significant in so far as it can identify all of those genres which an individual *might* engage in, and which might, therefore, serve as an organising principle for an LSP course.

3. Dichotomy 2: Written vs. spoken genres

Swales (1990) and Bhatia (1993) have both focused on written genres, although they do mention more spoken genres in their later work (Bhatia 2004; Swales 2004), and most of the work I briefly reviewed above also relates to written genres. Swales, indeed, in his 1990 book, claimed, on the grounds that it is too unpredictable, that casual conversation was not amenable to genre analysis (pp. 58–61). At about the same time as Swales and Bhatia were developing their theories, based on written genres, interesting work was being done on spoken genres in the Firthian/Hallidayan/Systemic-functional tradition (especially Hasan 1978, 1979, 1985/9; developed further by Ventola 1987). But even before that, Mitchell, a student of Firth, had provided a precursor of move structuring with his model of market interactions in North Africa, as follows: 1. Opening salutation; 2. Enquiry as to object of sale; 3. Investigation as to object of sale; 4. Bargaining; 5. Conclusion. Here is Ventola's (1987) similar model for service encounters in Australia: 1. Greeting;

2. Attendance allocation; 3. Bid for service; 4. Service; 5. Resolution; 6. Goods handover; 7. Pay; 8. Closing; 9 Good-bye. And here is Ventola's (1987) model for casual conversation: 1. Greet; 2. Address; 3. Identify; 4. Approach; 5. Centring; 6. Leave-taking; 7. Good-bye.

A decade later, Eggins and Slade (1997) devoted a whole volume to the genre of conversation. They identify the following sub-genres for this macro-genre: Narrative, Anecdote, Exemplum, Recount, Observation/comment, Opinion, and Gossip. However, they recognise that what they call *chat* (also *joke-telling* and *sending-up*) is not amenable to generic analysis, chat, making up 58% for women, 47% for men and women, and 52% for men in their conversational data (Eggins & Slade 1997: 268). So, based on these figures, according to Eggins and Slade, about half of all conversation is amenable to genre analysis. The following figure (Figure 4) shows the generic staging of the sub-genres that they judged to be amenable to genre analysis, based on their data (Eggins & Slade, 1997: 268).

One might say that the debate about whether casual conversation is amenable or not to genre analysis is unimportant for LSP, given that casual conversation is not "specific", in the sense that it does not belong to "professional" or "disciplinary" genres, as specified for genre analysis by Bhatia, cited in the introduction to this chapter. However, it would be naïve to believe that this "genre set", if we can give it such a label, based on Eggins and Slade's division of it into sub-genres, can be ignored in specific purpose language pedagogy. Although specialist genres might be the most salient feature of professional interaction, as specified by Bhatia, more often that not these specialist genres are implicated in genre networks that involve a certain amount of conversation. If we take the genre chain, as exemplified for the conference paper by Raisonen (2002) above, it can be confidently inferred that a lot of informal spoken interaction will have mediated between the different stages of the chain. Conversation, indeed, might be seen as a sort of oil that lubricates the interaction of other professional genres in genre networks.

Narrative	(Abstract)^(Orientation)^(Complication)^Evaluation^Resolution^(Coda)
Anecdote	(Abstract)^(Orientation)^Remarkable Event^Reaction^(Coda)
Exemplum	(Abstract)^(Orientation)^Incident^Interpretation^(Coda)
Recount	(Abstract)^Orientation^Record of Events^(Coda)
Observation/comment	(Orientation)^Observation^Comment^(Coda)^(Completion)
Opinion	(Opinion)^Reaction^(Evidence)^(Resolution)
Gossip	Third person^Substantiating Behaviour^(Probe)^/Pejorative Evaluation^ (Defence)^(Response to Defense)^(Concession)^(WrapUp)

Figure 4. Move structure for conversational genres (Eggins and Slade, 1997)

This work by Eggins and Slade on conversational genres is also valuable in that it is accompanied by a detailed lexico-grammatical description of how the various moves in the move structures are realised, but I shall come to that below (Dichotomy 4: Move structure vs. lexico-grammar). The main point here is that this work does show that spoken language, even many aspects of casual conversation, can be analysed using genre analysis.

If casual conversation, as suggested by Swales, is the most problematic for genre description (indeed for him it is impossible), it follows that other spoken genres, such as presentations, lectures, and meetings are perhaps less problematic. There are indeed published examples for some of these. Du Bois (1980, 1997) and Ventola et al. (2002), for example, have presented genre analyses for the conference presentation, while Young (1993) has an interesting model for lectures, where instead of a staged series of moves, there is a small set of communicative acts – definition, generalisation, exemplification, summarising, and evaluation – which are recursive. Swales (2004: chapters 5–6), indeed, provides a review of the limited research available on two spoken academic genres: the dissertation defense and the research talk. So, in short, the point that I want to make is that in ESP we can use the results of spoken, as well as written genre analysis, although I should say that much more research is needed before we can apply such findings with equal confidence to that of the research article.

4. Dichotomy 3: Macro vs. micro levels of analysis

Bhatia (1993: 22–36) provides a very succinct description of his approach to genre analysis methodology, in terms of seven stages, as follows: 1. Placing the given genre-text in a situational context; 2. Surveying the existing literature; 3. Refining the situational / contextual analysis; 4. Selecting a corpus; 5. Studying the institutional context; 6. Levels of linguistic analysis; 7. Consulting with specialist informants. Most of these stages can be described as macro levels of analysis, in so far as they are concerned with the broader context. Stage 3 of this macro contextual analysis is broken down into four further stages:

1. Defining the speaker/writer of the text, the audience, their relationship and their goals;
2. Defining the historical, socio-cultural, philosophic and/or occupational placement of the community in which the discourse takes place;
3. Identifying the network of surrounding texts and linguistic traditions that form the background to this particular genre/text;

4. Identifying the topic/subject/extra-textual reality which the text is trying to represent, change or use and the relationship of the text to that reality.

What Bhatia is alluding to here are the sort of categories which are the primary targets of investigation of ethnography. Benson (1994: 181), for example, in a discussion on ethnography in the context of academic listening, lists the following aspects of a culture which are susceptible to ethnographic description:

> structures, contexts, rituals, universals, significant symbols, roles, status markers, patterns of behavior, beliefs, values, assumptions, attitudes, and even the allocation of praise and blame (together with their consequent rewards and punishments) (p. 181).

I will return to this more ethnographic macro analysis issue in a moment. First I would like to consider the more micro analysis, which is what Bhatia refers to in his sixth stage for genre analysis (levels of linguistic analysis), which is further broken down as follows: 1. Analysis of lexico-grammatical features; 2. Analysis of text patterning or textualization; 3. Structural interpretation of the text-genre.

Along with stage 3 – selecting a corpus – it is probably true to say that these three sub-stages of micro analysis are what has most captured the attention of the majority of genre analysts working in ESP, especially the third of these sub-stages, the structural interpretation of the text-genre, or move structure, at the expense, it could be argued, of the more ethnographic macro analysis. Why this should be the case is not clear. It could be that, as linguists primarily, genre analysts do not wish to leave their ivory towers and do more ethnographic work, preferring the desk-bound work of the structural and linguistic (move) analysis. It could also be the result of the emphasis on move structure in the two influential books by Swales (1990) and Bhatia (1993) and the fact that there is less exemplification of the more ethnographic dimensions in these two publications. I am not sure. What I would like to emphasise, however, is that there can be a danger for ESP if the micro-analysis is not accompanied by the more macro one. For the more contextual analysis, we need to turn to the New Rhetoric school, but here there is relatively little emphasis on actual linguistic form (Flowerdew 2002).

Let me provide two examples to illustrate how a macro analysis can be valuable for LSP. First, if we take the lecture genre again, in one (unpublished) study conducted in Australia, using ethnographic methods, the researcher found that, in lectures, the international students he was observing did not actually listen to the lecture very much at all. They spent most of their time copying from the overhead slides. So a micro analysis of the move structure of the lectures might not

have proved very useful. An analysis of the slides might have been of more value in that context. Second, if we return to a study I reported on earlier, the one I conducted with Alina Wan (Flowerdew & Wan 2006), in observing the group of tax accountants who were the focus of this study, we arrived at one very important finding: the tax accountants made tremendous use of templates and only actually needed to do original writing in one key move of the genre, the *discussion of issues* move. If a micro analysis had been conducted on a corpus of letters, without the more macro ethnographic dimension, the researchers would not have realised this. Carried over into pedagogy, the implications are clear; valuable time might have been wasted in developing pedagogic materials to teach how to write the whole letter, when in fact what was needed was a more targeted analysis of the important moves that accountants actually needed to write. Similar findings were arrived at in another study (Flowerdew & Wan, 2010), focusing on the company audit genre.

5. Dichotomy 4: Move structure vs. lexico-grammar

If the emphasis in LSP genre analysis has been, as I have suggested, on levels of linguistic analysis, as in Bhatia's stage 6 above, I would now like to suggest that within this stage there has also perhaps been a lack of balance in emphasis in that much more work has been done on move structure than lexico-grammar and text patterning. Both Swales and Bhatia stress that the lexico-grammatical realisation of the genre is an important part of the analysis, but in actual fact the amount of attention they give to it in their respective books is fairly minimal, compared to that given to move structure.

Here are two quotes from Bhatia, taken from his earlier book and his later one:

> *Genres are highly structured and conventionalized constructs, with constraints on allowable contributions not only in terms of the intentions one would like to give expression to and the shaper they often take, but also in terms of the lexico-grammatical resources one can employ to give discoursal values to such formal features ...* (emphasis added) (Bhatia 1993: 110).
>
> *Genre essentially refers to language use in a conventionalized communicative setting in order to give expression to a specific set of communicative goals of a disciplinary or social institution, which give rise to stable structural forms by imposing constraints on the use of lexico-grammatical as well as discoursal resources.* (emphasis added) (Bhatia 2004: 23).

Significantly, though, Bhatia does not focus much on lexico-grammar in his actual empirical analysis, merely providing the odd example of typical lexico-grammatical realisations in both of his books (Bhatia, 1993, 2004).

With regard to Swales, the following are examples of authentic realisations drawn from academic articles of the first step of the first move of Swales's research article introductions, "claiming centrality", as presented by Swales (1990: 144):

Recently, there has been a spate of interest in how to ...
In recent years, applied researchers have become increasingly interested in ...
The possibility ... has generated interest in ...
Recently, there has been wide interest in ...
The time development ... is a classic problem in fluid mechanics.
The explication of the relationship between ... is a classic problem of ...
Many investigators have recently turned to ...

Swales provides findings for each of the three moves identified for the introductions to research articles in his 1990 book (see also Swales & Feak 1994 for evidence of this in materials writing). However, these examples not withstanding, it is probably true to say that Swales has shown more interest in his two books on genre (Swales 1990, 2004) on move structure than on lexico-grammatical realisation.

For more interest in lexico-grammar we need to turn to the Sydney School and the work of Eggins and Slade (1997), already mentioned, for example (also Thornbury & Slade 2006). Interestingly, in their work on spoken genres, these two systemicists give quite a lot of attention to lexico-grammar, relating the moves of the different spoken genres to particular lexico-grammatical features. The following figure (Figure 5) is from Eggins and Slade (1997). It shows the typical moves and associated language features in the "story" genre in conversation.

This work, it should be noted, is of a qualitative nature and general tendencies are extrapolated from individual texts.

A probable reason for the relative lack of interest in lexico-grammar among genre analysts is that such work requires the study of a reasonably large corpus if reliable relations are to be established and this is time consuming and may not be realistic in many, if not most, LSP situations. What I want to emphasise here, though, is that more interest might be shown in lexico-grammar, in order to balance out the more extensive work on move structure. As course designers and

Stages	Moves	Language features
Abstract Establishes the point of the text Signals that a story is about to be told	1a–1c	Use of generalized *you* to state general relevance of story Attitudinal lexis *quite amazing* An attributive clause in move 1 is elaborated in moves 1b–1c
Orientation Orients the listener to what is to follow in terms of people, actions, time and place	1d–1f	Introduces wider field (charity) and then more specific fields (Christian charity, catering manager) Use of circumstances of location Use of generic statements (e.g. move 1d) Point of story established in moves 5a–5e where appraisal of whole experience is given *Every minute* and *everything* are markers that this is a general comment, there by establishing the significance of the story
Incident Outlines temporally sequenced events in order to elucidate interpretative comments or moral judgement	5f–9i	Past tense action processes Time sequence
Interpretation	9j–1l	Use of anaphoric extended text reference *that* in move 9j Refers back to whole incident Use of descriptive/attributive clauses which may be interpersonally modalized and/or the attributes interpersonally loaded

Figure 5. Move structure and typical language features in the 'story' genre (Eggins and Slade, 1997)

materials writers, exemplificatory linguistic realisation patterns of the various moves in move structure are highly important in conveying generic form to learners (bearing in mind that this information needs to be appropriately presented, making it clear that form function relations are probabilistic and not one-to-one).

6. Conclusion

So let me summarise and conclude. I have given a quick overview of what genre analysis does, especially in terms of move structure, and I have presented four dichotomies for genre research. These dichotomies are: 1. Individual genres vs. genre networks; 2. Written vs. spoken genres; 3. Macro vs. micro levels of analysis; and 4. Move structure vs. lexico-grammar. If I had more space I could discuss further dichotomies. Another eight, to double the number that I have dealt with here, might be, for example: 1. Linear ordering vs. complex embedding; 2. Obligatory vs. optional components; 3. Discrete communicative acts vs. complex intentions; 4. Top-down vs. bottom-up modes of analysis; 5. Disciplinary vs. other types of variation; 6. Public vs. occluded (Swales 1996) genres; 7. Acceptable intertextuality vs. plagiaristic practice; and 8. Training vs. learning in genre pedagogy. All of these have important implications for LSP pedagogy, not just the last. But, looked at mathematically at least, this would be equivalent to two further articles. So, clearly, in this article, I have barely scratched the surface. Finally, lest I be misunderstood, let me say that my purpose in this chapter has not been to undermine genre. As I stated at the beginning, the theory has already been very productive for LSP. In presenting this set of dichotomies, as stated, again at the beginning of the chapter, my goal has been to highlight areas where further theoretical reflection and empirical enquiry might enhance the theory and at the same time offer even more for application. Indeed, in a small way, I hope that in the course of my discussion of these issues I might already have made a small contribution to praxis.

References

Bakhtin, M.M. 1981. *The Dialogic Imagination: Four Essays,* M. Holquist (ed.), transl. C. Emerson & M. Holquist. Austin TX: University of Texas Press.

Bakhtin, M.M. 1986. *Speech Genres and Other Late Essays*, transl. V.W. McGee. Austin TX: University of Texas Press.

Bazerman, C. 1994. Systems of genres and the enactment of social intentions. In *Genre and the New Rhetoric,* A. Freedman & P. Medway (eds), 79–101. London: Taylor and Francis.

Benson, M.J. 1994. Lecture listening in an ethnographic perspective. In *Academic Listening: Research Perspectives*, J. Flowerdew (ed.), 181–199. Cambridge: CUP.

Berkenkotter, C. & Huckin, T. 1995. *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*. Hillsdale NJ: Lawrence Erlbaum Associates.

Bhatia, V.K. 1993. *Analysing Genre: Language Use in Professional Settings*. London: Longman.

Bhatia, V.K. 2004. *Worlds of Written Discourse*. London: Continuum.

Devitt, A. 1991. Intertextuality in tax accounting. In *Textual Dynamics of the Professions*, C. Bazerman & J. Paradis (eds), 336–357. Madison WI: University of Wisconsin Press.

Dubois, B.L. 1980. The use of slides in biomedical speeches. *English for Specific Purposes* 1: 45–50.
Dubois, B.L. 1987. Something on the order of around forty to forty-four: imprecise numerical expressions in biomedical slide talks. *Language in Society* 16: 527–541.
Eggins, S. & Slade, D. 1997. *Analysing Casual Conversation*. London: Cassell.
Fairclough, N. 1992. *Discourse and Social Change*. Cambridge: Polity Press.
Flowerdew, J. 1993. A process, or educational, approach to the teaching of professional genres. ELT Journal 47(4): 305–316.
Flowerdew, J. 2002. Genre in the classroom: A linguistic approach. In *Genre in the Classroom: Multiple Perspectives*, A.M. Johns (ed.), 91–102. Mahwah NJ: Lawrence Erlbaum Associates.
Flowerdew, J. & Wan, A. 2006. Genre analysis of tax computation letters: How and why tax accountants write the way they do. *English for Specific Purposes* 25(2): 133–153.
Flowerdew, J. & Wan, A. 2010. The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific Purposes* 29: 78–93.
Hasan, R. 1978. Text in the systemic functional model. In *Current Trends in Text Linguistics*, W.U. Dressler (ed.), 228–246. Berlin: De Gruyter.
Hasan, R. 1979. On the notion of text. In *Text Versus Sentence: Basic Questions of Text Linguistics* [Papers in Textlinguistics 20(2)], J.S. Petofi (ed.), 369–390. Hamburg: Helmut Buske.
Hasan, R. 1985. The nursery tale as a genre. *Nottingham Linguistic Circular* 13: 71–102. (Special issue on Systemic Linguistics).
Hyland, K. 2004. *Genre and Second Language Writing*. Ann Arbor MI: University of Michigan Press.
Hyon, S. 1996. Genre in three traditions: Implications for ESL. *TESOL. Quarterly* 30(4): 693–722.
Johns, A.M. 1997. *Text, Role and Context*. Cambridge: CUP.
Johns, A.M. (ed.). 2002. *Genre in the Classroom: Multiple Perspectives*. Mahwah NJ: Lawrence Erlbaum Associates.
Kristeva, J. 1980. Word, dialogue and the novel. In *Desire in Language: A Semiotic Approach to Literature and Art*, L. Roudiez (ed.), Transl. T. Gora, A. Jardine & L. Roudiez. New York NY: Columbia University Press.
Martin, J.R. 1984. Language, register and genre. In *Children Writing*, F. Christie, (ed), 21–29. Geelong: Deakin University Press.
Martin, J.R. 1993. A contextual theory of language. In *The Powers of Literacy. A Genre Approach to Teaching Writing*, W. Cope & M. Kalantzis (eds), 116–136. Pittsburgh PA: University of Pittsburgh Press.
Paltridge, B. 2001. *Genre in the Classroom*. Ann Arbor MI: University of Michigan Press.
Raisanen, C. 2002. The conference forum: A system of interrelated genres and discursive practices. In *The Language of Conferencing*, E. Ventola, C. Shalom & S. Thompson (eds). 69–93. Frankfurt: Peter Lang.
Swales, J.M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: CUP.
Swales, J.M. 1996. Occluded genres in the academy: The case of the submission letter. In *Academic Writing: Intercultural and Textual Issues*, E. Ventola & A. Mauranen (eds), 45–58. Amsterdam: John Benjamins.
Swales, J.M. & Feak, C.B. 1994. *Academic Writing for Graduate Students: A Course for Nonnative Speakers of English.* Ann Arbor MI: University of Michigan Press.
Swales, J.M. & Feak, C.B. 2000. *English in Today's Research World: A Writing Guide.* Ann Arbor: University of Michigan Press.

Thornbury, S., & Slade, D. 2006. *Conversation: From Description to Pedagogy*. Cambridge University Press.

Ventola, E. 1987. *The Structure of Social Interaction: A Systemic Approach to the Semiotics of Service Encounters*. London: Pinter.

Young, L. (1994). University lectures macro-structures and micro-features. In J. Flowerdew (Ed.), *Academic listening: research perspectives* (pp. 159–179). Cambridge: Cambridge University Press.

English for legal purposes and domain-specific cultural awareness

The 'continental paradox', definition, causes and evolution

Shaeda Isani
Université de Grenoble (France)

In a departure from mainstream language-orientated English for Legal Purposes (ELP) didactic approaches, this article focuses on target legal culture. It identifies the 'Continental paradox' which relates to the appropriation of common law legal culture as source legal culture by the lay public of civil law countries, and evokes the didactic repercussions of the phenomenon. The author identifies the 3-phase cognitive process which underlies the acquisition of the ersatz legal culture in the light of Konrad Lorenz's imprinting theories and Stephen Krashen's acquisition theories. Focusing on a French environment, she then probes into the media-related, academic and didactic reasons which sustain such misconceptions, notably amongst law students themselves, but concludes that changing public expectations regarding transparency in the justice system may usher in remedial counter-discourse.

Keywords: ESP, ELP, Anglo-American entertainment media, legal film thrillers, European legal culture, lay public misconceptions, Konrad Lorenz, Stephen Krashen

The study proposes to analyse a phenomenon often encountered by ELP teachers working in European civil law academic environments who find themselves confronted with a public of learners lacking essential knowledge of their own legal culture. The discussion seeks to situate the phenomenon within a broader systemic framework and analyse the academic and media factors which lead to its creation and, perhaps more importantly, to its perpetuation.

The research presented derives from a cross-disciplinary approach which overlaps three disciplinary *loci*, law (legal institutions, professionals, procedures,

etc.), ethnology (intersecting perceptions of different legal cultures), and English for Specific Purposes (ESP) didactics related to teaching ELP to European students of law. It proposes a paradigmatic shift in ELP studies by moving from the language-centred, text-based approach traditionally privileged in ELP didactics and research, towards a more holistic approach which relates specialised discourse to the professional and/or disciplinary culture of its genesis. The aim is to cultivate greater learner awareness of target (common law) legal culture, thereby correcting an imbalance which produces law students who are familiar with the principal characteristics of legal textual genres – terminology, doublets, syntax, pro-forms, pronominal adverbs, etc. – but remain largely ignorant of the professional culture that fashions such text, not to mention other aspects of the target legal system.

1. Legal culture

As an expression and reflection of the values by which a particular society seeks to regulate its present and future generations, and given its historically endogenous vocation, law is probably one of the most deeply culture-embedded areas of human activity. Even in today's era of cultural relativism, and in spite of the increasing erosion of the perception of law as a monolingual, monocultural and monojuridical entity (Gémar and Kasirer 2005) it continues to be viewed as a predominantly culture-specific domain, notably in countries like France which have a comparatively recent tradition of immigration and cultures in contact.

If the debate with regard to the convergences and divergences of distant legal cultures is of current topicality as evidenced by the controversies arising from attempts to regulate aspects of Islamic law in countries such as the United Kingdom, Canada or France, the central focus of this enquiry concerns two close cultures which nevertheless possess a justice system and legal culture which have been historically considered as antonymic: common law as opposed to civil law. This enquiry does not concern the law *per se*, as an object of legislation, text and interpretation for and by professionals but relates to the notion of *popular* legal culture, which Friedman (1989: 1579) defines as "ideas, attitudes, values and opinions about law held by people in a society."

Casanovas (1999) defines three areas of legal culture according to whether it focuses on knowledge, behaviour or organisation and whether it is "internal" or "external", a concept further developed by Greenfield, Osborn and Robson (2001: 4) for whom "external legal culture" refers to popular legal culture, and "internal legal culture" to a narrower, profession-based perspective. Extrapolating from the concepts of 'High culture' (i.e., relating to the Fine Arts) and 'Low culture' (i.e., relating to a more anthropological study of how humans live and work together), a

further defining paradigm specific to the subject-domain of this study may be defined in terms of a distinction between 'High *Legal* Culture' and 'Low *Legal* Culture', the former referring to more abstract areas of legal culture such as values, aspirations, philosophy and thinking, and the latter to an ethnological and pragmatic approach regarding the organisation and functioning of the law community, as defined by David Nelken (2004):

> Legal culture, in its most general sense, is one way of describing relatively stable patterns of legally oriented social behaviour and attitudes. The identifying elements of legal culture range from facts about institutions such as the number and role of lawyers or the way judges are appointed and controlled, to various forms of behaviour as litigation or prison rates, and, at the other extreme, more nebulous aspects of ideas, aspirations and mentalities. Like culture itself, legal culture is about who we are not just what we do. (Nelken 2004) <http://repositories.cdlib.org/csls/lss/20>

To this must also be added a more commonly shared popular dimension of legal culture, that of national myths and legends, heroes and villains, a dimension to which the famous trials and protagonists that have fuelled a nation's past and present imagination lend themselves particularly well and contribute to shaping a national discourse based on a system of cultural references shared by the lay public and professionals of law alike.

2. Identifying the Continental paradox

What then is 'the Continental paradox'? Robin Lakoff was amongst the first to document the phenomenon in her book *Talking Power: The Politics of Language* (1990). During a conference at a Barcelona university, she observed that while students had difficulty discussing trial procedure relevant to their own country, they were able to discuss American trial procedure "with almost as much accuracy as Americans" (1990: 85). In other words, Spanish students possessed greater knowledge of American legal culture than of their own. The same phenomenon has been analysed with reference to France (Isani 2001, 2006) and Germany (Machura & Ulbrich 2001), and may also be presumed to be prevalent in other civil law countries as well, a cross-cultural dimension reflected in its designation as 'the Continental paradox'.

The singularity of the Continental paradox lies in that it is not a simple question of ignorance but arises, as demonstrated *infra*, from a more complex situation of cultural substitution caused by the fact that ignorance of the endogenous legal culture (civil law) runs parallel to awareness of an exogenous legal culture (common law), leading the latter to be appropriated as a form of ersatz or surrogate culture

to compensate for the absence of the former. So deeply entrenched in the public's mind is this ersatz legal culture that anecdotal evidence gathered from members of the French Bar reveals that French lawyers find it necessary to brief clients prior to court appearances, warning them that there will be no jury,[1] that counsel will not leap up to shout *"Objection, Votre Honneur"*, and that the judge is not to be addressed as *"Votre Honneur"* but as *"Monsieur le Président"*.

3. Some ELP didactic considerations

Though the phenomenon analysed is symptomatic of the European lay public at large, in the didactic perspective of this discussion, our observations focus mainly on law students studying ELP in a French academic environment.

The profile of the law student differs from that of the general public with regard to familiarity with source legal culture. As future professionals of law in the process of acquiring a capital of specialist knowledge, their lay status presents a more complex picture. It would appear natural to assume that if first-year law students may be equated with the general public, each further year should logically diminish the student's ignorance in his specialism domain. This, however, is not always so regarding source legal culture and it is not infrequent to find even third-year law students whose knowledge of their source legal culture – as opposed to knowledge of the law itself – is often as flawed as that of the general public in certain areas (Isani 2001, 2006).

The learners chosen to provide data to support empirical findings (Lakoff 1990) are third-year students taking a double degree course in law (at the Law Faculty) and two advanced level foreign languages (at the Arts Faculty). Curriculum design introduces ELP relatively late in the curriculum and consequently, to enable teachers to calibrate input, students are administered a simple test at the beginning of each academic year designed to identify common misconceptions regarding source legal culture. The findings of this survey, though intended for pedagogic purposes, substantiate empirical evidence gathered by teachers and observers regarding this line of enquiry.

The following sampling from tests administered in September 2008 provides some typical occurrences of such cultural inversion:

– Questioned as to the appropriate form of address required for French judges, 39% of the respondents answered correctly *("Monsieur le Président"),* 35% answered incorrectly ("*Votre honneur*"), and 28% said they did not know.

1. Juries only sit in the Assizes in France.

- In response to whether "*Objection, votre honneur!*" was appropriate in a French court, over 78% answered in the affirmative, a multiple error since neither the practice, formula nor title exists in French legal culture.
- When asked to ascribe a given definition (that of the adversarial system) to the civil or common law system, 35% made a correct identification, while 56% made an erroneous identification attributing the definition to the civil law system.
- For 7% of the respondents, the inquisitorial system belongs to the common law system, while for 22% adversarial procedure belongs to the civil law system.

These findings are corroborated with empirical evidence in other fields of ELP teaching as described elsewhere (Isani 2006) when students working in a self-learning environment on a selection of legal films successfully identified such elements as lawyer TV commercials, conditional fees and absence of dedicated court dress for counsel as being specific to American legal culture, but described examination and cross-examination, "Objection, Your Honour", the 'Miranda Rights', lay juries sitting in civil cases, punitive damages, search warrants and the negative image of the lawyer's profession, for example, as being part of French legal culture.

The existence of the Continental paradox has didactic repercussions relative to both teaching and learning. For the teacher, one of the main consequences is the obvious impediment created by the absence of any reliable presumption regarding learner specialist-domain knowledge. A lesser documented impediment to learning concerns behaviour that imprinting theories associated largely with Konrad Lorenz (see *infra*) describe as "bizarrely maladaptive" when subjects "imprint the wrong image" (Lorenz 1970). One manifestation of "bizarrely maladaptive" behaviour is the disbelief manifested by learners when confronted with the erroneous nature of long-established perceptions regarding their legal culture. Learners displayed tension, destabilisation and varying degrees of resistance to remedial input. Although other factors undoubtedly contributed to the "bizarrely maladaptive" behaviour, such reactions are understandable when analysed in the light of the affective factor which underlies all pedagogic situations, notably the pride of being right and the embarrassment of being wrong, all the more so with regard to what the learner considers his area of specialisation. The situation highlights the delicate situation created for the ELP teacher, the supposed "empathiser" (Robinson 1991: 81), whose very first pedagogic task consists of dismantling an edifice the learner has, consciously or not, been building up over time.

4. Creation of an ersatz legal culture: Cognitive processes

What are the cognitive processes underlying the creation of the ersatz legal culture which defines the Continental paradox? We propose to analyse them as a 3-phase sequential process defined in terms of vacuum, exposure and appropriation.

i. *Vacuum*

The first phase concerns the vacuum generated by the absence of legal culture. What begs the question, however, is that such a state of 'innocence' should at all exist in an era of mass intelligence and a social environment of ubiquitous, plethoric and easily accessible information.

Part of the explanation lies in the fact that European law institutions and professionals remain anachronistically hermetic and remote. In this respect, a first consideration is that, in spite of the very public vocation of the law, the general public has little first-hand knowledge of law courts as compared with such other public institutions as schools, town halls, doctors' surgeries and even hospitals. Likewise, the lay public is seldom called upon to interact directly with the professionals of law as is the case with other more visible professionals as doctors, dentists, school teachers and even policemen for example, all social actors the public interact with from infancy.

Another well-known factor often cited as a cause of alienation between the public and the institutions and professionals of law is the obscure and arcane character of the legal world. The notoriously incomprehensible language of the law (in spite of efforts towards "plain English"), the antiquated dress code of its professionals designed to differentiate and mask them from the very public they are meant to serve, the archaic rites and rituals of court proceedings and finally, the negative reputation of lawyers and law courts, are all elements which converge to keep the public at a distance and nurture the vacuum.

And finally, European legal institutions remain shrouded in obscurity due to the limited and highly regulated access allowed to journalistic media, as opposed to the United States, for example, where the spirit of the First Amendment's guarantee of freedom of the press, makes live media coverage of courtroom proceedings available at the click of the remote control button. European courts of law, in contrast, remain institutions to which the lay public may accede physically and in person but not through the vicarious facility of televised mass media.

ii. *Exposure*

Exposure is the second phase of the cognitive processes underlying the creation of an ersatz culture, and in this context the imprinting theories of 1973 Nobel Prize ethologist, Konrad Lorenz, prove particularly relevant.

– *Imprinting theories*

Interested in genetic and environmental influences on behaviour, Lorenz's (1970) experiments with newly-hatched goslings showed that, when confronted with the absence of a mother figure at birth, goslings adopted the first moving person they were exposed to as the mother figure to be followed, often Konrad Lorenz himself:

> A gosling must print an image of its mother during a critical period just after hatching. During this period it will imprint an image of any agent to which it is exposed. [...] From then on, only that agent will elicit following behaviour. If the gosling imprints the wrong image, then its following behaviour will be bizarrely maladaptive. (Lorenz 1970: 245–246)

Lorenz's imprinting theories highlight the crucial role that the first visual images to be imprinted on a virgin mind play in durably shaping perceptive schemata and subsequent behaviour patterns. Transposed to our field of enquiry, imprinting theories allow us to posit that given the *tabula rasa* generated by the absence of a source legal culture, the lay public instinctively adopts and appropriates as its own, the first legal culture it is exposed to, thus creating an ersatz culture to fill the vacuum.

In our contemporary era of aural and visual literacy dominated by non-print media, the cinema and television are the privileged purveyors of the moving images which imprint the initial schemata that shape perceptions, as Shulamit Almog and Ely Aharonson underline with reference to the domain of justice:

> Such [moving] images form the framework through which we filter and process reality. As part of a general framework, each of us has accumulated an enormous reservoir of audio and visual capital, comprised of hundreds and thousands of images, that influences our perceptions of justice. In our age, cinema has been one of the primary suppliers of this symbolic capital. It is therefore possible to speak of a certain cinematization of our thinking about justice. (Almog and Aharonson 2004: 3)

Given its hegemonic status, it is predictably Anglo-American law-related cinematic discourse which becomes the source of the early exposure which imprints its subliminal message on the receptive virgin minds of European, if not global, audiences – which observation leads to the question of why law-related fiction remains a predominantly Anglo-American cinematic genre.

– *Anglo-American law-related entertainment media*

The power of law-related entertainment media to shape perceptions with regard to how a population views and values its justice system is a dynamic area of scholarship, particularly in the area of law and cinema studies (Robson 2009). In terms of didactic informed perspectives, three broad areas may be defined with reference to

this field of scholarship: studies relating to law-related films as pedagogic supports in the teaching of law itself (Meyer 1993; Nevins 2000; Greenfield, Osborn & Robson 2001; etc.); enquiry dominated by American academics and professionals based on an endogenous and intracultural approach focusing on how American law-related films shape perceptions regarding the American justice system and its professionals (Leonard 1988; Strickland 1997; Asimow 2000; etc.); and finally, a more European and intercultural approach which studies how Anglo-American law-related entertainment media, a culture-specific genre representing the American (and, to a lesser degree, British) justice system shapes perceptions held by a non Anglo-American public, not only with regard to Anglo-American legal culture but, more perversely and importantly, with regard to their *own* – in other words, how representations forged by and of an inherently *common law* legal culture can generate, inform and shape perceptions held by a lay public belonging to a *civil law* culture with regard to their source legal culture (Isani 2001; Machura & Ulbrich 2001; Garapon 2005; Villez 2005; Guéry 2007).

Since the world of justice in common law countries is as incomprehensible for the lay public as it is in civil law ones, the question arises as to why the lay public in these countries is more familiar with its legal culture than Europeans are with theirs. A key element of the answer lies with the entertainment media, whether films, television series or specialised television channels, a field in which the global ubiquity of the American entertainment media industry gives it hegemonic status, as the following data provided by the European Audiovisual Observatory in May 2009 confirm with regard to cinema attendance in the European Union.

European films	28.4%
European inc/US production	6.8%
US	63.2%
Others	1.6%

Source: <http://www.obs.coe.int/about/oea/pr/mif 2009_cinema.html>

Figure 1. Admissions in the European Union according to country of origin (2008)

The legal thriller is probably one of the most popular cinematic genres today and courtroom drama and lawyer films are its most profuse sub-genres. What is it that makes these two genres a primarily Anglo-American stronghold? The answer, as practitioners of law and ELP teachers know, lies in the culture-specific nature of their courtroom procedures: the (counsel-dominated) adversarial system of common law countries as opposed to the (judge-dominated) inquisitorial system of European or civil law countries, defined as follows by the *Concise Oxford Dictionary of Law.*

> Accusatorial (or adversary) procedure is a system of criminal justice in which conclusions as to liability are reached by the process of prosecution and defence. It is the primary duty of the prosecutor and defence to press their respective viewpoints while the judge acts as an impartial umpire, who allows the facts to emerge from this procedure. (OED 1992: 7)
>
> Inquisitorial procedure is a system of criminal justice, in force in some European countries but not in England, in which the truth is revealed by an inquiry into the facts conducted by a judge. In this system it is the judge who takes the initiative in conducting the case, rather than the prosecution or defence, his role is to lead the investigation, examine the evidence, and interrogate the witnesses. (OED 1992: 208)

If all trials are performances, the adversarial trial is a particularly suspenseful one. It allows for verbal duelling between two opposing counsel – what Lakoff (1990: 86) calls "the pyrotechnics and razzle-dazzle of both sides" – played out in front of a Greek chorus in the form of the legendary lay jury and an audience composed of the press and public, along the classic lines of examination and cross-examination, regularly punctuated by "Objection, Your Honour!", and terminating with the climatic *catharsis* of the verdict.

Comparing the German and Anglo-American systems, Machura and Ulbrich (2001) describe the adversarial system in metaphoric terms of combat:

> American [and British] legal procedure [are] more hospitable to scenes of intense drama, featuring classical confrontations between two antagonists and conflicts between good and evil which have to be resolved. Such conflicts can be made a great deal more powerful in an adversary procedure than in German criminal procedure [...]. In American procedure the parties wrestle to establish their individual version of the facts of the case as well as their legal views. Where the courtroom is seen as a battleground, fierce conflict becomes the norm. (Machura and Ulbrich 2001: 125)

The inquisitorial system, in contrast, with its contained presentation of facts addressed to the Bench, appears lacklustre in comparison, as Lakoff points out:

> Adversarial encounters are thrilling, inquisitorial generally tedious. It takes two sides to make a discussion interesting to an outsider. That's why we have courtroom dramas. The American courtroom is intrinsically interesting: two people are fighting tooth and nail, using every rhetorical device dreamed of by Aristotle, along with all the technical expertise money can buy. Add the desire to see justice done and truth revealed, and you have all that is needed for gripping drama. Countries with inquisitorial systems have no tradition of court room drama cinema. Hence, nobody other than professionals knows or much cares what happens inside a courtroom. (Lakoff 1990: 86–87)

In addition to adversarial courtroom procedures, Anglo-American legal culture boasts another feature which renders it particularly conducive to dramatic fictionalisation, i.e., the very culture-specific species of lawyer called the investigative lawyer whom fiction has transformed into the lawyer-cum-detective as epitomized by America's most famous fictional attorney, Perry Mason:

> Lawyers have traditionally, and accurately, been associated with legal scholarship and libraries, dusty offices and minutiae of esoteric judicial procedure. [...] Erle Stanley Gardner moved the professional legal advocate from an officer of the court, an important but still subsidiary figure in drama whose central characters were the judge and jury, to center stage as the activist master of ceremonies, a lion tamer who takes control of a dangerous situation and imposes order much as fictional investigators and detectives have always done. (Macdonald & Macdonald 2005: 47–48)

In contrast to this image of the Anglo-American lawyer as a man of action, the European lawyer, whose role is not to discover evidence but to present the case to the judge on the basis of evidence received from *le juge d'instruction,*[2] appears a deskbound paper-pusher – a somewhat unpromising professional to cast in the fictional role of leading protagonist.

The dynamics of adversarial trial procedure and the dashing role of the investigative lawyer are two common-law culture-specific attributes which make courtroom drama an essentially Anglo-American sub-genre. Given the unchallenged dominance of American films all over the world, it is small wonder then that the American adversarial system of courtroom procedure, as portrayed and perceived through filmic text, is subconsciously adopted to fill the vacuum of source legal culture, as Christian Guéry, professional judge and author, observes.

> The multiplicity and success of American television and cinema fiction on our national channels has led to public disinformation [...]. American fiction has accomplished the feat of making us familiar with cultural codes which are alien to our own legal culture. (Guéry 2007: 9)[3]

iii. *Appropriation*

To satisfy the need to fill the vacuum, such primary exposure ineluctably leads to appropriation. What it is important to highlight here is the mode of this

2. Variously translated into English as examining judge or investigating judge.

3. Loosely translated from: *La mutiplicité et le succès des fictions télévisuelles ou cinématographiques américaines sur nos chaînes nationales a entraîné une désinformation du public [...] Les fictions américaines ont réussi l'exploit de nous rendre familiers de codes culturels qui sont pourtant étrangers à notre véritable culture juridique.*

appropriation.[4] In this context, linguist Stephen Krashen's (1981) seminal distinction between *acquisition* and *learning* with reference to foreign language learning is highly relevant, a distinction summed up by David Crystal (1987: 368): "In several approaches, 'acquisition' and 'learning' are carefully distinguished: the former is restricted to what takes place in 'natural' learning situations; the latter to what takes place in classrooms when following a structured course with a teacher". The defining characteristics of the two modes may be schematically presented as follows.

Acquisition mode →	Subconscious
	Implicit
	Informal
	Feeling-based
	Random
	Uncontrolled input
	Low mediation
Learning mode →	Conscious
	Explicit
	Formal
	Rule based
	Orderly: simple to complex
	Controlled input
	High mediation

Figure 2. Characteristics of Krashen's Acquisition/Learning Theories

The construction of a legal culture on the basis of random exposure to cinematic input in informal and uncontrolled situations corresponds to Krashen's acquisition model and results in an end product which, due to its unmediated and unstructured mode of acquisition, is necessarily fragmentary and flawed.

Flawed perceptions of legal culture, inherent to the mode of random appropriation, are not in any way specific to a foreign lay public. Even the American and British lay public, which derives knowledge of its justice system through the same

4. Due to the mother/child affect underlying Lorenz's imprinting theories based on how 'orphan' goslings 'adopt' the first available visual object as the mother figure, an initial inclination was to identify this final phase as adoption. However, the term appropriation seemed more apposite due to its connoted nuance of making use of 'property' considered as rightfully belonging to another. We realise that the term, especially with regard to such compounds as culture appropriation and language appropriation, is not neutral owing to its acceptation in cultural and colonial studies where it refers to theories pertaining to colonised/coloniser culture paradigms. Nevertheless, a milder acceptation of the term is attested according to which culture appropriation refers to widespread practices of informal cultural "borrowing" such as, for instance, the wearing of nose studs or Celtic bands by people extraneous to the culture. It is in this acceptation that the term is used here.

channels, acquires its piecemeal knowledge of the national legal culture through erroneous cinematic representations, as one English magistrate forcefully points out in *The Magistrate's Blog*, dated August 26, 2008:

> Not Again! I have just watched an excellent programme on BBC4 about the Lady Chatterley trial. But why, why, why did the judge, after the verdict, bang a gavel? For the hundredth time No Court in England and Wales Uses a Gavel. Ever. <http://www.thelawwestofealingbroadway.blogspot.com/2008/08/not.again.html>

Misconceptions generated by such cinematic misrepresentations are inevitable. The difference is simply that while the native Anglo-American layman's movie-generated flawed perceptions are rooted in and pertain to the legal culture of his country, the European layman's perceptions suffer from the additional flaw that the misconceptions are 'imported' from a legal culture extraneous to civil law culture.

5. Sustaining the Continental paradox: Absence of counter discourse

Having examined some of the reasons which generate this ersatz legal culture, the question now arises as to why such widespread misapprehensions remain so firmly entrenched. Part of the answer resides in the fact that inadequate counter discourse, whether cinematic, academic or journalistic, fails to provide a remedial framework of checks and balances.

5.1 Continental cinematic legal thrillers

It might be argued that if American entertainment media generate erroneous perceptions of source legal culture, then the cinema industry of the audience's own country must surely provide the necessary corrective counter discourse through the cinematic legal thrillers produced by and for endogenous audiences.

This argument fails to take into account the inconvenient fact that the inquisitorial system of civil law countries does not lend itself to legal thrillers, as Lakoff (1990: 87) points out when speaking of the Spanish inquisitorial system, a fact that holds true for civil law countries in general. This does not imply that cinematic legal thrillers do not exist as a genre in European civil law countries. Speaking for France, television series such as *Avocats & Associés* (France2, since 1998), *Les Cordier, juge et flic* (TF1, since 1992), *Alice Never, le juge est une femme* (TF1, since 2002) are popular French productions. However, as statistics underline, French *films judiciaires* are inferior in number to American television series since French television channels prefer to import rather than produce. Furthermore, French television producers tend to privilege the quality docudrama – *L'affaire Dreyfus*

(Arte 1994), *L'affaire Dominici* (TF1 2003), *Marie Besnard, l'empoisonneuse* (TF1 2006), *L'affaire Villemin* (Arte 2008), *L'Abolition* (France2 2009) – over popular entertainment fiction. Quality, however, does not guarantee audience ratings and if all the respondents in the 2008 student survey mentioned earlier, showed what Almog and Aharonson call "kitsch-like familiarity with [Ally McBeal] from popular culture" (2004: 12), less than 3% were able to identify the otherwise famous protagonists from their own source legal culture.

More critically, French-produced *films du prétoire* or courtroom dramas often carry within them the larvae of the Continental paradox in that, to offset the tedium of the inquisitorial trial procedure, film directors allow themselves considerable poetic license and present an imaginary hybrid model freely adapted from the Anglo-American system, a tendency which dates back to 1942 according to Christian Guéry, French judge and film critic specialised in legal thrillers (2007: 16). Egregious examples of such liberties continue today, as, for example, the film *Un Crime au Paradis* (2001) in which French courtroom procedure is presented along Anglo-American lines with lawyers leaping to their feet and shouting *"Objection, votre Honneur"*, as Nadine Picca, (2001: 25), practising lawyer and then head of the Regional Bar School, substantiates after previewing the film: "It's obvious that to make [the trial] part of the film attractive, the director grafted onto the French court system – which in reality is rather static – a mode of functioning closer to what we see in the United States." Such cinematic misrepresentations begin early. In *Que du Bonheur*, a daily prime time mini-series which targets the 7 to 77 audience bracket, one episode (TF1 25/08/08) focuses on children enacting a court scene to 'try' their parents' conflicting claims: the 8-year-old who plays the part of the judge is of course attired in an unmistakeably British judicial wig and gown.

Such filmic cultural (con)fusion is not specific to France, and Machura and Ulbrich report the same phenomenon with regard to the German cinema industry when they say, "American procedure has provided the foundation for almost all cinematic legal procedures, even in films set in a country like Germany that has a different system." (2001: 123)

In view of such incongruities, it becomes clear that, far from offering remedial counter discourse capable of holding the hegemonic discourse of American law-related entertainment media in check, European cinematic courtroom drama conversely tends to facilitate the cognitive processes of erroneous acquisition that lead up to and consolidate flawed perceptions of source legal culture. In this respect, Jonathan Rosenbaum's definition of what constitutes 'national' cinema – "a cinema that expresses something of the soul of a nation that it comes from: the lifestyle, the consciousness, the attitudes." (2002: 136) – is invested with ironical significance since by reproducing the erroneous "societal expectations" (Almog & Aharonson 2004: 7) of target European audiences, script writers and film makers

do indeed reflect "the consciousness" of a large part of the nation with regard to its legal culture.

5.2 Role of law professionals

Law professionals, practitioners and teachers are well aware of the syndrome which defines the Continental paradox but, curiously enough for professionals who may legitimately be considered as primary stakeholders with regard to such misrepresentations, they tend to adopt an attitude of resignation. Given the degree of prevalent ignorance, professionals even appear to be grateful for whatever little accuracy cinematic representations offer, displaying surprising tolerance for the transposed representations entertainment media disseminate about their profession, as the following extract from an interview with Nadine Picca, *"Bon Boulot!"* ("*Good Work!")*, indicates:

> Much like the film itself, the trial scene is not very realistic but this isn't really a problem. The fact that he took the trouble to put the Public Prosecutor, the jurors and the public where they would actually be in the *Cour d'Assizes* is enough to make his transposition very credible compared to a lot of films we see.[5]

5.3 Law school academic culture

Law school curriculum designers and teachers must assume part of the responsibility for the flawed perceptions that law students hold about their own legal culture. Continental law faculties, on the whole, remain deeply attached to the black letter tradition of the law, a phenomenon which Greenfield, Osborn and Robson attribute to the fact that "the legal academy is often isolated from the wider scholastic community and perhaps possesses an innate conservatism that militates against progression" (2001: 2). As a result, law curricula tend to focus on the acquisition of knowledge relative to the law *per se.* In the case of optional subjects designed to broaden student horizons beyond the core curriculum, most European law school teachers, steeped in the rich heritage of Europe's great legal thinkers, privilege High Legal Culture over Low Legal Culture, an area weighed down by the term 'popular', as Greenfield, Osborn and Robson suggest: "[...] Popular culture is itself a loaded term, and perhaps accounts for the lack of academic treatment it has

5. Loosely translated from : «*A l'image du film, la scène du procès n'est pas très réaliste, mais cela n'est pas vraiment le problème ! Le fait d'avoir pris le soin de mettre l'avocat général, les jurés et le public aux places qu'ils occupent réellement dans la salle d'assises suffit à rendre la transposition de Jean Becker tout à fait crédible par rapport à des tas de films qu'on a pu voir.* » *(Essentiels, 7 mars 2001: 25)*

received" (2001: 3). Consequently, unlike American universities, for example, areas of study related to law and popular culture (law and literature, law and film studies, popular legal culture, etc.) remain largely ignored by European law academe, thus leaving the issue of erroneous representations generated by the Continental paradox largely unaddressed at law school level.

Divergent approaches regarding exposure to non-institutional learning reinforce the situation. Other disciplinary cultures, such as business studies for example, integrate professional experience into the curriculum at relatively early stages. Law school culture however, accords primacy to acquisition of theoretical knowledge. It is thus less orientated towards exposure to non-institutional learning and hence a large number of law students gain first-hand experience of legal culture relatively late in their academic evolution. A striking example of this experiential deficit concerning source legal culture concerns law students enrolled in the joint law/languages programmes, who undergo a reverse experience: one of the requirements of the programme being a second-year internship abroad and, given that field visits to local courts, prisons, police stations, etc. are not common in French law school culture, a student's first real experience of a court, for example, may quite well be that of a *foreign* court. This reverse exposure serves to reinforce existing flawed perceptions imprinted by the initial Gosling syndrome and results in students returning from an internship in London, for instance, half expecting French courts to be similar to what they found in Britain.

Certainly, a number of law professionals (Garapon 2005; Guéry 2007) and academics (Isani 2001; Villez 2005) have drawn attention to the French lay public's Americanised concept of its own legal culture and raised questions about the ensuing risk of cultural alienation and the gradual emergence of a standardised global model. However, in terms of readership/audience impact, it is obvious that the remedial discourse of written journalism and academic writings is no match for the *son et lumière* impact of popular entertainment media.

5.4 ELP didactics

ELP teachers may themselves inadvertently contribute to the Continental paradox, particularly in the case of a learner-centred communicative approach in which teacher/student relations are constructed on the basis of a language/specialism tandem consisting of, simply put, the teacher providing the language input and the learner providing the specialism input.

The dynamics underlying such an approach draw on the all important question of predictability of learner subject-domain knowledge and the status of the learner defined in terms of a language/specialism equation. Pauline Robinson defines three basic parameters: "Are the students pre-experience, post-experience, or

are they studying their specialism concurrently with English?" (1991: 84). In the case of source legal culture, learner profiles correspond to the third configuration with the correlated assumption of little speciality knowledge as compared to a "post-experience" learning situation. ELP teachers therefore find themselves providing input relative to target professional culture for learners who do not possess the basis of their own legal culture – and thus further contribute to 'imprinting' an alien legal culture on virgin minds.

6. Conclusion: Winds of change?

Speaking more specifically for France, while evolution in the didactic field of ELP advances slowly but perceptibly, a number of elements are perhaps generating a climate favourable to winds of change.

Firstly, there has been a significant evolution in the attitude of professionals of French entertainment media. Faced with the 'invasion' of law-related American television series, certain French directors have increasingly chosen to represent a more authentic version of French legal culture rather than continue to perpetuate a pseudo, hybrid genre designed to conform to the public's flawed perceptions. One step in this direction has been to invest the protagonist's role in the nearest French equivalent of the investigative lawyer, *le juge d'instruction* whose functions the *Britannica Online Encyclopaedia* defines as follows:

> In France, the *juge d'instruction* is a magistrate responsible for conducting the investigative hearing that precedes a criminal trial. In this enquiry, evidence is gathered and presented, witnesses are heard and depositions taken. [...] The *juge d'instruction* has a wide range of powers. He may issue warrants allowing the authorities to search the residence of the accused and seize necessary evidence. He may issue warrants requiring people to appear as witnesses, or he may request experts to testify. In case of conflicting testimony, witnesses are confronted with each other and often with the accused.

The investigative nature of this function lends itself to the image of the legal professional on the go, very much in the role dynamics of the Anglo-American cinematic lawyer – and in sharp contrast to representations of the Bench as stately but sedentary professionals, appropriately called *la magistrature du siège* or 'seated judges'.

Admittedly, one swallow does not a spring make, and these are but timid inroads into the quantitative preponderance of American legal entertainment media and French predilection for American legal thrillers. In spite of such drawbacks, there is nevertheless a perceptible stirring of sensibilities as evidenced by a recurrent meta-discursive refrain found in popular television series today where protagonists are often heard to say, when confronted with a character's erroneous

borrowings from Anglo-American legal culture, "You've been watching too many American television series."

Another area of evolution favourable to the diminution of the Continental paradox is news media coverage of trials. European countries adopt a more conservative policy regarding mediatisation of justice than the United States where currently 44 states permit still and television cameras in court. Ever wary of O. J. Simpson-like excesses in this field, European courts maintain a ban on television coverage of trials, whether live or differed. Inevitably, however, in an age dominated by access to and excess of information, public opinion is increasingly questioning the traditional veils of secrecy that protect courts from public scrutiny. In the United Kingdom there is current debate regarding camera access to courtrooms and in 2003, senior judges agreed to consider the filming of an appeals hearing. In France, media coverage of trials has evolved more gradually. For a long time, officially accredited court reporters were the only journalists authorised to report on-going trials through text and artists' impressions. In 1981, article 308 of *le code de procedure pénale* authorised television cameras in courts during the opening and closing phases of hearings. This continues to be the case today with the exception of trials deemed to be of historical interest as defined by decree 86–74 of 15/01/1986 issued on the eve of the trial of Gestapo officer Klaus Barbie for crimes against humanity. Such films, sealed for 20 years, may subsequently be released for educational purposes to selected audiences.

In view of such strictly regulated relations between courts and the media, we note with surprise that in an unprecedented move, two senior French judges authorised proceedings (of no historical importance) in their courts to be filmed and presented for public viewing in 2004 resulting in the documentary *10ème chambre, instants d'audience* (2004) and a television episode entitled *La course à la vie* (F3, 2004) presented on the weekly programme *3600 secondes*.

Such exceptions may be seen as premonitory. The conservative policy regarding media coverage of courtroom proceedings was vociferously challenged following the 2005 miscarriage of justice, the so-called *affaire d'Outreau*. Public outrage over the prolonged remand of 18 innocent suspects on charges of paedophilia accessorily resulted in the public's discovery of its ignorance regarding its own legal culture and an ensuing claim for greater judicial transparency. In a surprising reversal of situation, the entire proceedings of the 2006 parliamentary enquiry into the *Outreau* miscarriage of justice, originally intended to be held *in camera*, were broadcast *in extenso* and live on public television.

While such openings may appear timid in comparison to American First Amendment-inspired standards, they have the significant merit of generating a growing and keener awareness of the Continental paradox and a degree of

consensus with regard to the need for a country's justice system and legal culture to be understood first and foremost by its own citizens, including law students.

References

Almog, S. & Aharonson, E. 2004. Law as film: Representing justice in the age of moving images. *Canadian Journal of Law and Technology* 3(1): 1–18. <http/::cjlt.dal.ca/vol3_no1/front_cont.html>

Asimow, M. 2000. Bad lawyers in the movies. *Nova Law Review* 24(2). <http://tarlton.law.utexas.edu/lpop/etext/nova/asimow24.htm> 8 November, 2004.

Casanovas, P. 1999. Pragmatics and legal culture: A general framework. Faculty Working Paper 159, Institut de Ciències Politiques i Socials, Universitat Autonoma de Barcelona. <http://hdl.handle.net/2072/1305> 28 June, 2009.

Crystal, D. 1987. *The Cambridge Encyclopedia of Language*. Cambridge: CUP.

Friedman, L.M. 1989. Law, lawyers, and popular culture. Yale Law Journal 98(8): 15–79.

Garapon, A. 2005. *Avant-propos. Séries télé: Visions de la justice*. Paris: Presses Universitaires de France.

Gémar, J.-C. & Kasirer, N. 2005. *Jurislinguistique: Entre langues et droit/Jurilinguistics: Between Law and Language*. Bruxelles & Montréal: Éditions juridiques Bruylant & Éditions Thémis.

Greenfield, S., Osborn, G. & Robson, P. 2001. *Film and the Law*. London: Cavendish.

Guéry, C. 2007. *Justices à l'écran*. Paris: Presses Universitaires de France.

Isani, S. 2001. Calques et emprunts culturels ou le paradoxe de la culture professionnelle cible comme vecteur de perceptions erronées dans la culture professionnelle source: Le cinéma et les professions juridiques. In *Culture & communication en milieu professionnel interculturel*, S. Isani (ed.), 105–120. Grenoble : CERELC, Université Stendhal Grenoble 3.

Isani, S. 2006. Revisiting cinematic FASP and English for legal purposes in a self-learning environment. *Cinéma & Langues de Spécialité*, *Les Cahiers de l'APLIUT* 25(1): 27–38.

Krashen, S.D. 1981. *Second Language Acquisition and Second Language Learning.* Oxford: Pergamon.

Lakoff, R.T. 1990. *Talking Power: The Politics of Language*. New York NY: Basic Books.

Leonard, D.P. 1988. From Perry Mason to Kurt Waldheim: The pursuit of justice in contemporary film and television. *Legal Studies Forum* 12(3). <http://tarlton.law.utexas.edu/lpop/etext/lsf/leonard12.htm> 3 March, 2008.

Lorenz, K. 1970. *Studies in Animal and Human Behaviour*, 1, R. Martin (transl.), 245–246. Cambridge MA: Harvard University Press.

Macdonald, A.F. & Macdonald, G. 2005. *Scott Turow: A Critical Companion*. Santa Barbara CA: Greenwood Press.

Machura, S. & Ulbrich, S. 2001. Law in film: Globalizing the Hollywood courtroom drama. Journal of Law and Society 28(1): 117–132.

Martin, E.A. 1992 [1986]. *A Concise Dictionary of Law.* Oxford: OUP.

Meyer, P.N. 1993. Visual literacy and the legal culture: Reading film as text in the law school setting. *Legal Studies Forum* 17(1). <http://tarlton.law.utexas.edu/lpop/etext/lsf/meyer17.htm> 8 December, 2004.

Nelken, D. 2004. Using the Concept of Legal Culture. <http://repositories.cdlib.org/csls/lss/20> 13 April, 2008.

Nevins, F.M. 2000. Using fiction and film as law school tools. In *Legal Education for the 21st Century*, D.B. King (ed.), 177–185. <http//:tarlton.law.utexas.edu/lpop/etext/nevinsusing.htm> 12 October, 2004.
Picca, Nadine. 2001. "Cest du bon boulot". *Essentiels* #7 mars 2001: 25.
Robinson, P. 1991. *ESP Today: A Practitioner's Guide*. Hemel Hempstead: Prentice Hall International.
Robson, P. 2009. Law, Hollywood and the European experience. Studies in Law, Politics, and Society 46: 117–146.
Rosenbaum, J. 2002. *Movie Wars*. London: Wallflower Press.
Strickland, R. 1997. The cinematic lawyer: The magic mirror and the silver screen. *Oklahoma City University Law Review* 22(1) <http://tarlton.law.utexas.edu/lpop/etext/okla/strickland.htm> 6 March, 2008.
Villez, B. 2005. *Séries télé: Visions de la justice*. Paris: Presses Universitaires de France.

Filmography

Avocats & Asociés. 1998–2010. Directed by Valérie Guignoabodet & Alain Krief. France 2.
Les Cordiers, juge et flic. 1992–2005. Directed by Alain Page. TF1.
Alice Never, le juge est une femme. 2000–. Directed by Noëlle Loriot. TF1.
Que du Bonheur. 2008–. Directed by Christophe Fort & Nathalie Roger. TF1.
L'affaire Dreyfus. 1994. Directed by Yves Boisset. Arte.
L'affaire Dominici. 2003. Directed by Pierre Boutron. TF1.
Marie Besnard, l'empoisonneuse. 2006. Directed by Christian Faure. TF1.
L'affaire villemin. 2008. Directed by Raoul Peck. Arte.
L'abolition. 2009. Directed by Jean-Daniel Verhaege. France 2.

Webliography

European Audiovisual Observatory. 2009. Press release "EU film production reached record high in 2008 as admissions to European films remained strong". <http://www.obs.coe.int/about/oea/pr/mif2009_cinema.html> 20 May, 2009.
"Juge d'instruction". *Encyclopaedia Britannica*. 2008. Encyclopaedia Britannica Online. <http://www.britannica.com/EBchecked/topic/307632/juge-dinstruction> 6 October, 2008.
The Magistrate's Blog. <http://www.thelawwestofealingbroadway.blogspot.com/2008/08/not.again.html> 1 October, 2008.

The Talking Cure

From Narrative to Academic Argument

Gillian Lazar
Middlesex University

Narrative, particularly of an oral nature, is not generally a genre given much consideration in the teaching of academic writing in the tertiary sector. Yet, in everyday discourse, oral narrative predominates. This article begins by describing the difficulties of some British university students engaged in writing academic assignments which were grounded in reflective practice. The use of oral narratives was found to be a useful bridge into academic writing for these students. Thus, I consider some features of narrative discourse, and the ways in which these can be exploited to develop academic writing skills which incorporate argument. Some examples of practical classroom techniques and materials used with students from different disciplinary backgrounds, including education and product design, are suggested.

Keywords: Academic writing, narrative, argument

1. Introduction

This article will begin by offering a brief confessional narrative, based on my experiences of teaching academic literacies/English for Academic Purposes (EAP) at a British university. Cadman (2005: 354) suggests that there is sometimes "...a scholarly reluctance in EAP to go into the living-and-breathing classroom" as writers prefer to focus on materials, methodology or conceptual concerns. The beginning of this article provides a brief account of a number of interactions I had in the "living and breathing" classroom. These interactions subsequently led me to think reflectively about the pedagogy I routinely deploy as a lecturer in English Language and Learning Support (ELLS), a unit at Middlesex University, UK.

2. A story from the university chalk-face

2.1 Orientation

Based on the figures in the Statistical Digest, the Middlesex University website records more than 25, 872 students in total at the university, with 55.8% studying for undergraduate degrees. 80.8% of this total student body is from the UK, with 19.2% from other countries. Of the university's UK students, 51.5% classify themselves as being from an ethnic minority, and 42.4% classify themselves as white. These figures are significant in that they index a student population that is extremely diverse, with many students arriving at university from non-traditional routes and consequently needing guidance in the practices of academic culture.

English Language and Learning Support (ELLS) is a university-wide free service, run by Learning Resources. It offers a confidential and un-assessed environment, where students can obtain guidance in and information about academic writing. Support is provided in the form of voluntary attendance at workshops, online writing courses delivered through discussion boards, and one-to-one tutorials. The "living and breathing" classroom discussed here will be largely confined, in fact, to interactions between two participants only. This is because the focus is on the tutorial – a 30-minute session in which the participants are an ELLS lecturer and, usually but not exclusively, a single student. Tutorials, as well as workshops and online courses, are attended by a diverse range of students, including international and home students, undergraduates and postgraduates, native and non-native speakers of English, students regularly using non-standard varieties of English, students with dyslexia and other specific learning difficulties, and students from a range of disciplinary backgrounds. As a post–1992 university, Middlesex has traditionally focused on courses with a strong vocational or professional training. Thus, at the Trent Park campus where the tutorials took place, the disciplines offered include education, performing arts, media studies, television production, product design and translation studies as well as other humanities subjects such as philosophy and English literature.

Over a number of years of providing tutorials, I became aware of an emerging pattern in this provision. Students choosing to attend tutorials were often required to write assignments with certain generic similarities even though they came from different disciplines. In the main, these students were engaged in writing essays, reports or dissertations which linked in some way to a real-world professional or vocational experience. The range of assignments included reports on work placements by publishing and media students, essays arising from classroom practice by education students, dissertations documenting the design of an original product

by product design students, and critical commentary by translation students about a text they had themselves translated. Schön's notion of reflective practice (Schön 1983, 1987) was evident in many of these assignments, whether explicitly manifest in the assignment title or not. Given that the paradigm of the reflective cycle is widespread in education (see for example Paige-Smith & Craft 2008; Loughran 1996), the titles for education assignments often included the words 'reflect on' or 'reflection'. This was generally not the case in assignments from other disciplines. Nevertheless, the notion of the reflective cycle is implicit in the assignments in that students were expected to critically analyse a real-world professional experience as the starting point for their academic writing.

2.2 Complication

Many of the students needing to write the type of assignment mentioned above arrived at the tutorials with a kind of 'stuckness'. They described themselves as feeling overwhelmed, unsure of how to start, and most significantly, concerned about how to make the assignment sufficiently 'academic'. Students often expressed frustration with the methods usually suggested in tutorials to encourage them to get started, such as mind-mapping, planning a first draft, writing an introduction, etc.

Working with these different students over a period of months, it seemed that one way out of this impasse was simply to ask the students to tell the story from their professional lives (as trainee teachers, journalists, product designers or translators) which would form the basis for their academic assignments. My role during this oral narration was to listen attentively, prompt where necessary and ask relevant questions where possible. In many cases, this oral narration seemed to unblock students who, on the whole, were able to express themselves with relative ease.

2.3 Evaluation

The fluency with which students told their brief stories or short anecdotes, suggested that this informal storytelling was entirely natural to them. As Nair (2003: 7) points out when reporting on the work of Dennett (1990):

> Human 'selves', according to Dennett are 'centres of 'narrative gravity' since humans are programmed to extrude narratives as naturally as spiders spin webs or beavers build dams. Each day of their lives human beings tell stories – to themselves and to others around themselves.

After students had 'told the story', the next stage was to ask them to write a brief account of the incident they had described. Over a number of months, it became clear that students' oral narratives conformed largely to the structure of narrative described by Labov (1972, cited in Toolan 2001: 148), and which can be laid out as follows:

1. *Abstract:* What in a nutshell, is this story about?
2. *Orientation:* Who, when, where?
3. *Complicating action:* What happened and then what happened?
4. *Evaluation:* So what? How or why is this interesting?
5. *Result or resolution:* What finally happened?
6. *Coda:* That's it, I've finished and am 'bridging' back to our present situation.

Some of the written versions seemed to dispense with the abstract and the coda, bringing them into line with the modified version of the typical sequence of the narrative genre suggested by Lock and Lockhart 1998 (cited in Hyland 2004: 33). This sequence is as follows:

1. *Orientation:* Gives information about characters' situation
2. *Complication:* Presents one or more problems for the characters to solve
3. *Evaluation:* Evaluates the major events for the characters
4. *Resolution:* Sorts out the problems for the characters.

One such narrative was provided by SR, a British home student, bilingual in English and Urdu, who was doing a Post-Graduate Certificate in Education (PGCE) in order to qualify as a primary school teacher. The title of her assignment was 'The analysis of the teaching of two foundation subjects' and she had chosen to write partly about the teaching of geography. The brief story she told me can be broken into its components as follows:

Orientation I watched three geography lessons with a Year 2 class.
Complication The pupils did not seem to be enjoying it and were not very motivated and they lost interest very quickly.
So, when I had to teach a lesson I tried to make my lesson fun and interesting, but the students got very bored in my lesson too.
Evaluation This was very disheartening.
Resolution

What is noticeable about SR's narrative was that there was no resolution, and that in the tutorial she was very disheartened by the complication. These features were discernable in many of the other narratives that students wrote down for me, suggesting that their feelings of 'stuckness' partly related to feeling overwhelmed by the complication and lack of resolution in their stories. However, by alerting SR to

the fact that it is the complication itself which should prove the basis for her academic assignment, she was able to move forward in asking the relevant questions necessary to proceed in her assignment, namely:

Why didn't the students enjoy the geography lesson?
What could be possible reasons for this?
Is there anything in the literature (including government documents, educational research journals, etc) which might provide insight into these reasons?
If so, what implications might this have for the teaching of geography in general?

The pedagogic stages followed in the tutorial with SR could thus be itemised as follows:

a. SR told a story grounded in her personal experience as a trainee teacher. This provided a starting point for thinking about the assignment, and enabled her to counteract the feelings of dysfluency engendered by having to write 'academically'.
b. This story was then written down and a simple narrative frame was applied to it.
c. SR was encouraged to focus on the complication in the narrative and to systematically problematise this complication by asking questions about it.
d. These questions then formed the basis of her reading and research before she began drafting and planning her assignment.

2.4 Towards a resolution

Andrews (1989: 3) has mentioned the "productive tension brought into play when narrative and argument are yoked together". This suggests that linking these genres together can generate a creative pedagogy. If, as we have seen, narrative generally involves sequenced events that are interrelated and a "crisis to resolution progression" (Toolan 2001: 8), what of argument? Defining argument in academic writing is no easy matter, but for the purposes of this paper, Andrews broad working definition is useful. In this definition, argument is "an arrangement of linguistic, visual and/or physical propositions in engagement with one or more other points of reference in order to change or assert a position" (Andrews 1997: 267). This definition underlines the fact that argument in academic writing involves the assertion of a position, but that this can be optionally multi-modal by using, for example, visual material. Since argument is central to academic writing, understanding the differences between the two genres of narrative and argument could enable us to utilise the productive tension between them more effectively.

First of all, the narratives told by students in ELLS tutorials arose from personal experience and were deeply embedded in a real-world context. On the other

hand, in order to write academic assignments demonstrating argument, students have to abstract and theorise. According to Laurillard (2002: 23) academic learning "...has a second-order character, as it concerns descriptions of the world". Thus, academic knowledge involves not learning about the world directly, but rather learning about others' descriptions of it. Focusing on the complication in SR's story, and encouraging her to formulate more abstract questions about it, enabled her to move from the personal story about the geography lesson to a more abstract, research-based process of information gathering, which could then form the basis for her academic assignment. Understanding that academic writing involves the martialing of abstract knowledge may be something of a cognitive breakthrough for students who arrive at university from non-traditional, non-academic routes and are struggling to make the links between personal experiences in the workplace and their academic writing.

A second distinction between narrative and academic argument is that narratives generally involve an event sequence which is linear and temporal, with elements remaining in a particular order. Andrews and Mitchell (2001: 47) have therefore argued that the planning involved in telling or writing a narrative is usually done by accretion. Conversely, the elements in an argument may be rearranged more easily so planning generally needs to precede scripting. As the literature on process writing emphasises (Tribble 1996: 39) a plan is very likely to be modified in the light of what emerges during scripting. Nevertheless, there needs to be some element of planning to begin the process. In the case of SR, she was unable to get to the planning stage until she had asked the questions that would allow her to think in a more abstract way. These questions arose from the gap occasioned by the complication in her narrative.

A third difference between narrative and academic argument is that generally, as we have seen, narratives include some form of complicating action that requires resolution. Arguably, complications are inherent in academic writing and may be dealt with by systematic problematisation, rather than resolution[1]. Once students realise that the aim of most forms of academic writing in arts, humanities and education is to systematically problematise, often presenting multiple viewpoints, rather than a single point of view, they can be encouraged to consider what forms of evidence may be used to support different viewpoints in order to develop effective argumentation. Becher and Trowler (2001: 36) point out that the nature of knowledge is discipline-specific; thus, developing academic literacy in a particular field is knowing what counts as evidence in that field, whether it is the quantitative or qualitative findings from research in Education, or the citations from primary sources such as newspapers or magazines in Media Studies.

1. I am indebted to Mike Riddle for this notion of 'systematic problematisation'.

A final difference between narrative and argument is that within universities it is academic argumentation, particularly in written discourse, which is "institutionally ratified" (Holmes 2005: 675). Holmes describes workplace narratives which despite satisfying the complex demands of identity construction at work, are not institutionally ratified as "'on-task' core business talk" (Ibid). ELLS tutorials at Middlesex University have tended to focus on creating an end-product of written academic discourse, since this is the genre that is privileged in the university. Nevertheless, if the use of oral narratives is sometimes a helpful way of leading students into academic writing, are there ways/techniques in which it could be valued and utilised, if not institutionally ratified, by teachers of academic writing? What pedagogic techniques do we have at our disposal for making oral storytelling part of the repertoire of academic writing teachers, and what might be our rationale for doing so?

3. Benefits of using oral narratives

One reason for using oral narratives with students is that they are grounded in the tacit knowledge of situated everyday experience. Granville and Dison (2005: 114) argue that "...if students are able to reflect using their own voices, they can remain anchored in their secure identities while they reach out towards new understandings and identities."

Thus, the use of oral narratives as a classroom tool could have psychological benefits in helping to validate students' personal experiences while also using these narratives as a stepping stone towards developing higher order skills, such as critical analysis, abstraction and developing an argument.

Secondly, using narrative can be a way of tapping into the implicit knowledge of genre that students bring with them. Bazerman (1997: 20) puts it like this:

> ...genre is a tool for getting at the resources the students bring with them, the genres they carry from their educations and their experiences in society, and it is a tool for framing challenges that bring students into new domains that are as yet for them unexplored...

Beginning with the known genre of the narrative may be a starting point for exploring the many different genres students may encounter in their university careers. Here are some possible pedagogic strategies for doing so.

3.1 Telling the story to the academic writing tutor

This is a simple exercise in which the student tells the story from their personal experience to the tutor in academic writing. The role of the tutor is to act as a

non-judgemental listener, pointing to the complications in the narrative which may direct the student to the lack of easy resolution in the story. In the case of SR the questions asked by the tutor could include:

- Why do you think the pupils lost interest in the geography lesson so quickly?
- Is it just geography lessons in which this happens? If so, why?
- If it happens in other lessons, in what ways may they be similar to geography lessons?
- Have you done any reading or research about this?
- If so, what does the reading and research say? How does it explain this problem? Do you agree with this?

A variation on this is when the tutor does not simply listen to the student, but makes notes or mind-maps of the student's story, so that there is a visual record of the narrative. It has been suggested that this technique may be particularly beneficial for dyslexic students, who may be strong in visual skills[2]. Once the story has been told, the tutor and student may use the visual notes as a basis for formulating the kinds of questions mentioned above.

This kind of interaction between student and tutor is described by Weissberg (2008: 21) as "scaffolded conversation". Tutor and learner work together in the zone of proximal development (ZPD) identified by the tutor as "the psychological distance between what the learner can accomplish on her own and what she can accomplish with the tutor's assistance".

Another way of conceptualising the role of the tutor in this scaffolded conversation is by conceiving of the tutor as a 'broker' enabling learning from one 'community of practice' to another (Wenger 1998: 109). In this conceptualisation, the tutor encourages the student to use the narrative to formulate the kinds of questions that will enable the students to become more proficient in the practices required for academic writing, such as moving from personalised situated experience to second-order abstract thinking. The tutor's role is to mediate for the students between their everyday experience in their workplace community of practice, and the conventions of academic writing in the university community of practice, until the student is able to do this for himself or herself.

3.2 Triangulation

A second way of incorporating oral narratives into the writing classroom is by means of triangulation. This technique can be used in large classes, as it involves

2. I would like to acknowledge that this suggestion came from Pauline Sumner of the Dyslexia Support Service, Middlesex University.

students in interacting with their peers. Hyland (2008) points out that oral interaction with peers in writing workshops is useful as a means of providing both practical and emotional support for students as well as helping students to develop confidence and independence as writers. During triangulation, students provide mutual support for each other in groups of three as well as helping each other cognitively by focusing on academic content. One student 'tells the story'; the second student listens and asks questions (such as those asked by the tutor in a tutorial); the third student makes a note of the questions. Once this process is complete, the group works together from the notes, trying to formulate a clear set of research questions, which can then form the basis for information-gathering in the next stage of assignment writing. Once the student telling the story is clear about the questions, the students in the group swap roles until all three students have had an opportunity to tell their stories.

The peer-group triangulation usually needs to be preceded by some form of modelling by the tutor, and two other students in front of the whole class. During this modelling process, the tutor should play the role of the questioner, emphasising the need for analysis in the asking of the questions. This means moving beyond the real-world narrative by asking questions such as:

- Why did this happen?
- What could be the reasons for it?
- What literature might be able to cast light on this? etc.

3.3 Writing a draft of the narrative and annotating it

This strategy involves students in writing a first draft of the narrative, to which they can then add what could be called 'Why?', 'Reference needed' or 'Evidence needed' annotations. Here is an example from HA, a student doing an MA in Product Design. HA was an experienced designer with a wealth of real-world experience, who spoke Portuguese as her first language. She had already had one tutorial in which she told the story of the design of her product, which at that stage was very much a work in progress, as she was struggling to resolve certain design problems. She was then encouraged to write this up in a very loose draft, which she presented as follows:

> *Following a brainstorm (see Appendix Z) a series of concepts for the tool changing mechanism were sketched. For choosing the most appropriate concept a weighting table was created based on (...GET REFERENCE) model. This type of concept selection technique is not perfect, but I felt it was the most adequate WHY??? Because of the technical criteria that will define the choice of the concept; because it put in evidence the elements that needed research and then the information used on the weighting table.*

What is striking about HA's draft is that in the process of writing her narrative, she has already begun to make the transition into a more academic mode. Thus, when telling the story to the tutor, HA used the first person singular 'I' throughout. In this sample of her writing, we can see that she is already making use of the passive voice (e.g. *was sketched*, *was created*), a grammatical form often identified as pervasive in academic writing.

This technique of annotation can be developed in many different ways. For example, students can be encouraged to use the 'Comment function' in software such as Word to annotate their own narratives with notes suggesting the bridge from simple narrative to argumentation. Students can also work in groups, annotating each other's narratives in the same way.

3.4 Providing models for analysis and discussion

Another useful strategy is that of providing students with models to critique, preferably of authentic student writing. Students may be given one model, which is a written version of the oral narrative, and a second, which is a more academic version of this. Students can then be asked to complete some discussion tasks, which draw out the differences between the two texts. Here is a sample of such an activity. It should be noted that both texts A and B were written by SR, the education student mentioned earlier. She wrote Text A immediately after our first tutorial; when she returned for her second tutorial we discussed some of the ways in which she could make her narrative more 'academic', and she then went away and wrote Text B.

Task 1

Read Texts A and B. What differences do you notice between them? Which text is more 'academic'? Why?

Text A

As part of my teaching practice I had the opportunity to teach three Geography lessons to a year two class. Geography was timetabled for one afternoon a week compared to the literacy and numeracy lessons, which were taught every morning. I feel this in itself gave out the wrong message to children about Geography, that it was not important as literacy and numeracy. Also at my first teaching practice both Geography and Religious Education were taught by a different teacher while the class teacher had her PPA time. I felt the children were not very motivated to learn in Geography, when they did not have their own class teacher teaching them. Also the teacher who taught this subject relied heavily on the LCP Schemes of work which comprised mostly of worksheets. Children did not have any Geography

books which meant they always had to work on worksheets. Having to do worksheets in every Geography lesson made children unenthusiastic and during a lesson the children would easily lose interest in the topic studied. Looking at it from all these aspects it can be said that the children were given hidden messages that Geography is not an important subject in education.

Text B

Firstly, it is important to discuss the importance of Geography as a National Curriculum subject, and why children need to learn this subject. One can agree with Scoffham (2004) that Geography helps us to understand and make sense of the world and also helps us to think about how the world may change in the future. Geography helps to prepare children for the future by helping them develop enquiry and investigative skills. No other subject can help children gain an insight into places and environments all over the world, as well as to develop an understanding about their rights and responsibilities they have themselves and towards others. Williams (2005) emphasises that by helping children learn about the world Geography prepares them for their future life and employment. Besides all this, yet still, according to the Ofsted report (2004/2005), on the whole the provision for Geography has still not achieved a 'good' in standards compared to other subjects.

As part of my first teaching practice I had the opportunity to teach three Geography lessons to a year two class. Geography was timetabled for one afternoon a week compared to the literacy and numeracy lessons, which were taught every morning. I feel this in itself gave out the wrong message to children about Geography, that it was not important as literacy and numeracy. In support of this Pirog and Tracz (2003) suggest that having only one lesson of Geography per week can make Geography is seen as a low status subject in the eyes of primary pupils. Also at my first teaching practice both Geography and religious Education were taught by a different teacher while the class teacher had her PPA time. I felt the children were not very motivated to learn in Geography, when they did not have their own class teacher teaching them. Also the teacher who taught this subject relied heavily on the LCP Schemes of work which comprised mostly of worksheets. Children did not have any Geography books which meant they always had to work on worksheets. Having to do worksheets in every Geography lesson made children unenthusiastic and during a lesson the children would easily lose interest in the topic studied. Looking at it from all these aspects it can be said that the children were given hidden messages that Geography is not an important subject in education.

Task 2

i. Read the first paragraph of Text B again. It is making a number of claims about the importance of geography in the National Curriculum.
   - What are these claims? Can you list them?
   - Do you think the claims are backed up sufficiently? If not, how could they be improved?
   - Are the links between the different sentences in Paragraph 1 clear? If not, what changes could you make to make the first paragraph more coherent?

ii. Read the second paragraph of Text B again. In this paragraph, the writer uses the first person singular ('I'). Do you think that this is appropriate? If so, why? If not, why not? If not, what techniques could be used to avoid using 'I'?

Question (i) of Task 2 aims to encourage students to think how cogent argumentation involves the effective backing up of claims, while Question (ii) asks students to think critically about the appropriate use of the first person singular in academic writing. Thus, in this activity, students work with authentic pieces of writing to notice and discuss the changes that take place when a piece of narrative is turned into a different genre. A follow-up activity to this might include asking students to write a third version of the text (Text C) in which nearly all of the elements of narrative are occluded, and academic argumentation is fore grounded more fully. Or students could be asked to write a narrative similar to Text A before being exposed to Texts A and B. Once they have studied Texts A and B, they could then be encouraged to rewrite their Text A so it is closer to Text B.

4. Further complications

This paper has proposed that oral narratives could be harnessed as a useful tool in helping students who have to write academic essays based on reflective practice. It has been suggested that using oral narratives allows tutors to legitimate the situated, personal experience of their students as a first stage towards developing more complex academic arguments. But this proposal raises a number of problematic issues for further reflection and research.

First of all, can oral storytelling be used as a basis for developing proficiency in other academic genres, or is it only possible to use it with assignments grounded in reflective practice, as discussed here? If it has a wider use, to which other genres of academic writing could it be linked?

Secondly, in what language should the oral storytelling take place? In this paper, no separation has been made between the needs of students who speak English as a first language or who are bilingual, and those for whom English is a

foreign or second language. This reflects the all-inclusive nature of the English Language and Learning Support at Middlesex University, which teaches academic writing to all students regardless of language background, not only to students on English for Academic Purposes courses. However, in the case of some students, particularly those who speak English as a foreign language, telling an oral story may arguably be more difficult than writing one. In critiquing what he calls the SCT (Socio-Cultural Theory) of writing, Weissberg (2008: 34) makes the point that learning to write in a second language (L2) is not necessarily analogous or equivalent to learning to write in one's mother tongue (L1). The assumption that it is analogous or equivalent may also mask significant individual differences between different L2 writers. Thus, it may be that some students may find that telling a story in English only increases the difficulty of a task for them. An alternative might be to tell the story in the mother tongue, where this is practically possible, e.g. in contexts where the tutor knows the student's mother tongue very well. With some other students, it may be important to recognise that writing may still be an easier skill than speaking so that the writing of the narrative should precede any scaffolded discussion of it.

A third difficulty with the use of oral narratives for developing academic writing skills is whether this kind of activity takes writing tutors too far from our role as teachers of specialised language. Should the role of the writing specialist be to concentrate on teaching the linguistic features that go to make up different academic genres? By helping students unpack the complications of their narratives are we straying too far outside of our area of expertise, and becoming involved in substantive content issues that reflect the epistemologies of different disciplines in which we are non-expert? In other words, is the work described in this paper best undertaken by disciplinary specialists, such as education or design lecturers?

Perhaps one way through this difficulty is suggested by Turner and Hocking (2004). They identify the generic characteristics of the tutorial and the post-graduate dissertation in contemporary fine art study, and demonstrate how the spoken mode of the tutorial with an art lecturer primarily mediates reflection on practice, while the written mode generates engagement with theory. They argue that the relationship between these modes is dialectical, with both feeding off and promoting the development of the other. This synergy between the two modes can be underlined by art tutors and writing tutors working in close collaboration. Close collaboration between subject specialists and writing tutors means that subject specialists benefit from gaining the insights and analytic tools of writing specialists, while writing tutors are able to gain a more insider perspective on a particular discipline. While this kind of collaboration is often challenging and expensive to organise, it can be very beneficial for students.

5. Coda

This paper began with a personal narrative from the university chalk-face. The coda with which it ends acknowledges that the writing of this paper has been informed by the "autobiographical impulse in our theory-making" (Neilson 1998: 226, cited in Cadman 2005). As writing tutors, teachers of EAP, lecturers in academic literacies, or whatever else we call ourselves, can the stories we tell from our living and breathing classrooms be used as a bridge into theory-making? For if this is so, the theory-making might then find its way back into the classroom, and perhaps other narratives will follow...

Acknowledgements

I would like to thank my students HA and SR for allowing me to use their work in this paper.

References

Andrews, R. (ed.). 1989. *Narrative and Argument*. Milton Keynes: Open University Press.
Andrews, R. 1997. Reconceiving argument. *Educational Review* 49(3): 259–269.
Andrews, R. & Mitchell, S. 2001. *Essays in Argument.* London: Middlesex University Press.
Bazerman, C. 1997. The Life of Genre, the Life in the Classroom. In *Genre and Writing: Issues, Arguments, Alternatives,* W. Bishop and H. Ostrom (eds), 19–26. New Haven: Boynton Cook.
Becher, T. & Trowler, P. R. 2001. *Academic Tribes and Territories,* 2nd edn. Milton Keynes: Open University Press.
Cadman, K. 2005. Towards a 'pedagogy of connection' in critical research education: A REAL story. *Journal of English for Academic Purposes* 4: 353–367.
Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.
Granville, S. & and Dison, L. 2005. Thinking about thinking: Integrating self-reflection into an academic literacy course. *Journal of English for Academic Purposes* 4: 99–118.
Holmes, J. 2005. Story-telling at work: A complex discursive resource for integrating personal, professional and social identities. *Discourse Studies Quarterly* 7(6): 671–700.
Hyland, F. 2008. Scaffolding during the writing process: The role of informal peer interaction in writing workshops. In *The Oral-Literate Connection,* D. Belcher & A. Hirvela (eds), 168–190. Ann Arbor MI: University of Michigan Press.
Hyland, K. 2004. *Genre and Second Language Writing.* Ann Arbor MI: University of Michigan Press.
Labov, W. 1972. *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.
Laurillard, D. 2002. *Rethinking University Teaching*. London: Routledge Falmer.

Lock, G. and Lockhart, C. 1998. Genres in an Academic Writing Class. *Hong Kong Journal of Applied Linguistics* 3: 47–64.

Loughran, J.J. 1996. *Developing Reflective Practice: Learning and Teaching and Learning through Modelling.* London: The Falmer Press.

Middlesex University. 2009. Available at: <http://www.mdx.ac.uk/aboutus/newsevents/factsandfigures.asp> 19 July, 2009.

Nair, R.B. 2003. *Narrative Gravity: Conversation, Cognition, Culture.* London: Routledge.

Neilson, L. 1998. *Knowing her place: Research literacies and feminist occasions*. San Francisco, CA: Caddo Gap Press.

Paige-Smith, A. & Craft, A. (eds). 2008. *Developing Reflective Practice in the Early Years.* Maidenhead: Open University Press.

Piróg, D. and Tracz, M. 2003. The Status of Geography in the Polish Education System. *International Research in Geographical and Environment Education* 12(2): 164–170.

Schön, D.A. 1983. *The Reflective Practitioner.* New York NY: Basic Books.

Schön, D.A. 1987. *Educating the Reflective Practitioner.* San Francisco CA: Jossey-Bass.

Scoffham, S. (ed.). 2004. *Primary Geography Handbook.* Sheffield: Geographical Association.

Toolan, M. 2001. *Narrative: A Critical Linguistic Introduction.* 2nd edn. London: Routledge.

Tribble, C. 1996. *Writing.* Oxford: OUP.

Turner, J. & Hocking, D. 2004. Synergy in art and language: Positioning the language specialist in contemporary fine art study. Art, *Design and Communication in Higher Education* 3(3): 149–162.

Weissberg, R. 2008. Critiquing the Vygotskian approach to L2 literacy. In *The Oral-Literate Connection,* D. Belcher & A. Hirvela (eds), 26–45. Ann Arbor MI: University of Michigan Press.

Wenger, E. 1998. *Communities of Practice: Learning, Meaning and Identity.* Cambridge: CUP.

Williams, A. 2005. Geography Curriculum Policy. The Reintegration Service, Gloucestershire County Council Education. Retrieved online from www.gloucestershire.gov.uk/utilities/action/act_download.cfm?mediaid=17482. Accessed 28/03/08.

UrgentiAS, a lexical database for medical students in clinical placements

Architecture, use and evaluation

Kris Buyse, Eva Saver, An Laffut and Herlinda Vekemans*
Leuven Language Institute (LLI)

The UrgentiAS database was compiled to meet the needs of Dutch-speaking Belgian medical students attending classes of *Medical English/French/Spanish* to prepare for their clinical placements abroad. The database provides students with an easy-to-use multilingual medical lexicon, which can help them acquire the vocabulary needed for their placements in *obstetrics/gynaecology, pediatrics and internal medicine*. Its main purpose is to ensure that the *representative vocabulary is presented and acquired in context, both semantically and syntactically*. Our *corpus analysis* of *both scientific and popular medical language use* resulted in an online database (partially available on paper) consisting of terminology cards with associative contexts, to be used in a *setting of blended learning*.

Keywords: Lexical database, language of Medicine, online resources, terminology.

1. Introduction

UrgentiAS (the Dutch **Urgenti**elexicon **A**rtsen**S**tage, lexicon[1] for medical students in clinical placements) was compiled to meet the needs of *Dutch-speaking Belgian medical students* attending classes of *Medical Spanish/French/English* to prepare for their *clinical placements in obstetrics and gynaecology, paediatrics and internal medicine in hospitals abroad*. Not only do these students need sound knowledge of the professional medical terminology in the language of the host country, they also need to be able to use it in many different communicative situations. They are

* With the cooperation of Stéphane Ostyn, Geert Peeters and Serge Verlinde.

1. A *lexicon* or *glossary* is "a list of terms in one or more languages. The amount of information contained in glossaries can vary greatly, and the level of detail in any glossary will usually depend on the purpose for which it is intended" (Bowker & Pearson 2002: 137).

expected to be sufficiently skilled in dealing with written communication, i.e. consulting and reading medical literature, and writing medical reports such as medical files, anamneses, diagnoses and treatment plans. Furthermore, accurate and fluent oral communication, both with patients and (para)medical staff of the guest hospital, is indispensable in the daily routine of their clinical placements. For English and French, interesting and valuable course material is available on the market, but for Spanish the need was pressing.

In order to meet this specific need for medical vocabulary in different contexts, we aimed to develop a *representative and easy-to-use online multilingual medical lexicon*, which can help students acquire the vocabulary needed in their placements abroad, and which allows them to consult it abroad whenever necessary.

As Nielsen and Mourier (2007: 119–121) state, lexicographic resources "must be linked to the *intended user groups*, the user's *linguistic and factual competences* and their *needs* in the relevant situations of use (...). [They are] designed to fulfil one or more *functions*, contain *lexicographic data* supporting the function(s) and contain *lexicographic structures* that combine and link the data in order to fulfil the function(s)" (italics ours). In this paper we will hence describe and discuss (1) the learners, (2) their linguistic and educational context and needs, (3) our approach (regarding lexicographic data, linguistic structures and their didactic function), (4) our corpus and (5) the corpus analysis, and (6) some preliminary data on the efficacy of the use of the resulting lexicon by the learners.

2. Learners

Our main target group or "primary user group" (Nielsen & Mourier 2007: 123) consists of more than 100 undergraduate students of the fifth year of medicine at K.U.Leuven (Catholic University of Leuven) who are interested in doing their *placements of obstetrics and gynaecology, paediatrics and internal medicine* in hospitals in French, English and especially Spanish speaking countries. Before they can apply for a clinical placement abroad, students have to pass a basic language test in the target language (CEFR level B2 for French and English, two of the three foreign languages taught in Flemish secondary schools; A1-A2 for Spanish, which is not part of the Flemish curriculum). Having passed this test, they can apply for a clinical placement abroad. Medical language courses are compulsory for those who have been granted a place. Due to the students' very intensive medical training programme, the number of contact hours of the medical language courses has been limited to 40 per year for Medical French and Medical English, and to 60 hours for Medical Spanish, following a preparatory course in basic Spanish.

UrgentiAS can be useful to *other, "secondary"*[2] *target groups* as well: based on their experience in placements in Belgian hospitals, students also report a growing need for foreign language proficiency, both for French (Belgium's second language) and for English. Furthermore, it can be used for training purposes during exchange programmes in the European Union, or for work purposes in the French-speaking part of Belgium as well as in other countries around the world. Finally, non-native medical students, PhD researchers in the biomedical field and doctors at our university can also benefit from the UrgentiAS lexicon.

At present the English and French modules are only available for students and staff with a K.U.Leuven computer login. The Spanish module has been published (electronically and in print[3]) and has thus become available to others elsewhere, for example to those working in the field of translating and interpreting for medical and social purposes.

3. Linguistic and educational context

The *basic needs* of our main target group are related to their educational context: our undergraduate medical students have a very demanding medical curriculum, as a result of which there is little time left for the acquisition of foreign languages. In other words, their *factual competence* contrasts sharply with their *linguistic competence* (Nielsen & Mourier 2007: 123). Pressed for time, students tend to resort to *simplistic word-by-word memorization*. The results of the basic language test that these students have to take clearly show that this is not a very effective way of acquiring vocabulary, as in this way their memory is not supplied with the necessary *contextualisation, useful collocations*[4] *and terminological variation*. Attention to *lexical, grammatical and orthographic precision remains limited*. Teaching materials available on the market for Medical Spanish, especially Spanish-Dutch

2. See Nielsen and Mourier (2007: 123).

3. Kris Buyse and Eva Saver (2008). UrgentiAS. Urgentielexicon Artsenstage Spaans. Leuven: Alta Uitgeverij (www.altamedia.be). ISBN 978 90 8579 024 2.

4. "An arrangement or juxtaposition of words or other elements, especially those that commonly co-occur, as *rancid butter, bosom buddy*, or *dead serious*" (http://www.thefreedictionary.com/collocation, accessed 27–09–2009). As Bowker and Pearson (2002: 18) point out, "people who are learning an LSP in a foreign language (...) may experience some interference from their native language. For example, a native French speaker who is learning the LSP of optical scanning technology in English may know that the term for 'tête' is *head* and the term for 'numériseur' is *scanner*, but when it comes to putting a sentence together, s/he may rely on intuition to come up with a structure that uses English words but is based on French syntax (e.g. 'tête de numériseur' translated as head of the scanner instead of scan head."

bilingual material, are inadequate to provide students with advanced, representative and contextual material. For Medical French and Medical English, the choice of materials is wider, but again rather limited in a bilingual form. In general, coursebooks focusing on language in context are few and far between on the market, as a result of which language lecturers often have to compile their own teaching materials (Balboni 2000: 105). The function of our lexicon is hence both *knowledge- and communication-oriented* (see Nielsen & Mourier 2007: 122): it provides our students with representative information on specialized vocabulary and correct use in authentic contexts. It is designed to be used in a setting of *blended learning*[5], especially in a didactic context facilitating frequent classroom practice of communicative skills such as doctor-patient simulations, presenting medical information and writing case reports. Therefore the lexicon focuses *on both reception and production* in the relevant situations of use (ibid. p. 124).

A lexicon or glossary is to be preferred to a (traditional) dictionary because some intrinsic disadvantages of dictionaries are their "inherent incompleteness", their size and the lack of contextual usage information:

> Most users would prefer to have a dictionary that will fit in their ruck-sack, which means that the lexicographers who create the dictionaries have to choose which information to include and which to leave out. Unfortunately, their choices do not always correspond with the needs of LSP[6] users. For example, acronyms and other abbreviated forms (...) are a common feature in LSP's, but these are frequently omitted from dictionaries. (...) Another of the most common criticisms of dictionaries is that they do not provide enough in the way of contextual or usage information. LSP learners have to pay attention to how terms are used, which means that in addition to information about the meaning of a term, they also need information about how to use that term in a sentence (e.g. what other words 'go with' the term in question) (Bowker & Pearson 2002: 15–16).

On the other hand, it is true that for a long time many LSP lecturers had relatively little idea of corpus linguistics, terminology and phraseology[7], one of the

5. "A learning solution created through a mixture of face-to-face and online learning delivered through a mix of media" (http://www.e-learningcentre.co.uk, accessed on 27–09–2009).

6. LSP: Language for Specific Purposes (viz. Medicine, in this case). Bowker and Pearson (2002: 25): "Perhaps the easiest way to describe LSP is to put in opposition to LGP, which refers to *language for general purposes*. LGP is the language that we use every day to talk about ordinary things in a variety of common situations. In contrast, LSP is the language that is used to discuss specialized fields of knowledge."

7. *Corpus linguistics* "is the study of language as expressed in samples (corpora) or "real world" text. This method represents a digestive approach to deriving a set of abstract rules by which a natural language is governed or else relates to another language. Originally done by hand, corpora are largely derived by an automated process, which is corrected."

consequences of which was that there tended to be a rather limited focus on representative contexts in actual class practice (Cabré & Gómez de Enterría 2006: 69).

4. Lexicographic data, linguistic structures and their didactic function

UrgentiAS is innovative in several respects. Not only is it multilingual (with a *global, unified concept for all the languages it covers*, alongside *language-specific tools*), it also combines the use of a *new, specialized LSP corpus*[8] *and ICT materials* on the one hand with the newest didactic insights on the other, such as *contextual vocabulary learning, deep level processing, Lexical Approach, blended learning, Content and Language Integrated Learning (CLIL), Task Based Language Teaching (TBLT)*[9]. The main output is an online database of *terminology cards with associative context* (see Appendix 1). A card contains not only the slots deemed necessary by recent research on terminology (translation, definition, subject field, grammatical information, phraseological information, related terms, notes about usage, synonyms, antonyms; see Nielsen & Mourier (2007: 126–133), Bowker & Pearson (2002: 142)), but also rather more innovative features such as words of the same class or family, paronyms[10], pre- and suffixes, geographical variants, valency[11], acronyms and pronunciation. Furthermore, "a traditional bilingual LSP dictionary

Terminology "is the study of terms and their use – of words and compound words that are used in specific contexts".
Phraseology "studies collocations of words (phraseologisms, phraseological units, idioms), where the meaning of the whole collocation is different from the simple sum of literal meanings of the words".
(http://www.thefreedictionary.com, accessed 12–10–2010)

8. A *corpus* is a "large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria. There are four important characteristics to note here: 'authentic', 'electronic', 'large' and 'specific criteria'. These characteristics are what make corpora different from other types of text collections" (Bowker & Pearson 2002: 9).
A *special purpose corpus* "is one that focuses on a particular aspect of a language. It could be restricted to the LSP of a particular subject field, to a specific text type, to a particular language variety or to the language used by members of a certain demographic group" (ibidem, p. 12).

9. See the following paragraphs for further explanations of those concepts.

10. "A word linked to another by similarity of form." (http://grammar.about.com, accessed on 27–09–2009).

11. "In linguistics, **verb valency** or **valence** refers to number of arguments controlled by a verbal predicate. It is related, though not identical, to verb transitivity, which counts only object arguments of the verbal predicate. Verb valency, on the other hand, includes all arguments, including the subject of the verb." (http://dictionary.babylon.com/Valency%20(linguistics, accessed 27–09–2009).

focuses exclusively on terms, but in any LSP text the terms account for less than fifty percent of the words; this kind of dictionary does not assist the user with all the words surrounding the terms" (Nielsen & Mourier 2007: 124; see also Cabré & Gómez de Enterría 2006: 65–66). This insight led us to pay special attention to different types of collocations, as a result of which the concept of "collocation" (see the Lexical Approach of Lewis 2000) has become one of the key features of our database. LSP dictionaries also tend to pay less attention to categories other than the noun (e.g. verbs, adjectives and adverbials), as nouns usually best represent field-specific concepts (ibidem, p. 66). In UrgentiAS we have aimed for a more balanced treatment of all categories. Finally, we have tried to include a great variety of authentic examples extracted from our corpora.

This is the complete field list:

- translation(s) (App. 1)
- domain (medical discipline) (App. 1)
- words of the same class or family (e.g. *amamantar > la mama, mamar, mamografía,* etc.; see App. 1)
- sample sentences in which the terms are used (App.1)
- (near-)synonyms, antonyms (App. 1 and 2)
- paronyms and homonyms (e.g. *el cura / la cura*)
- collocations: verbal, nominal, adjectival, adverbial or prepositional syntagms (App. 1)
- variants from different registers (scientific versus popular) (App. 1)
- geographical variants (App. 3)
- abbreviations and acronyms: the "acronymania", very typical of the professional medical jargon and frequently used in medical reports, has also been taken into account because abbreviations and acronyms are far from transparent to non-native speakers
- explanations, when contributing to a better understanding of the term (e.g. *SVD = spontaneous vaginal delivery*)
- information on morphology and syntax: transitivity, irregular conjugation, article, gender: feminine form if different from the masculine one (Spanish or French), valency (e.g. *introducir X en Y*)
- prefixes and suffixes, typical of the medical jargon (e.g. *-itis, neuro-*), with hyperlinks to other semantically related terms (e.g. *-itis > inflammation*)
- pronunciation in the English module, as experience has shown that speakers of Dutch often experience difficulties with and make mistakes against pronunciation (e.g. *hypertrophy* or *anaesthesia,* see App. 3)

The combination of these didactic and lexicographic principles ensures that students attending classes and learning vocabulary by means of the lexicon have

'significant encounters' with the target vocabulary (Buyse & Thiry 2008: 110). Although in our experience medical students generally acquire vocabulary more quickly than the average student does (the average student is able to acquire about five to twelve words per class for productive use, and twice as many for receptive use), learners' vocabulary acquisition potential should not be overestimated (see a.o. Cervero & Castro 2000 and Schmitt & McCarthy 1997). Systematic vocabulary teaching is certainly necessary, incidental learning (e.g. by reading, from the media, from native speakers) does not suffice: intentional learning is necessary as well, since "un mot vu n'est pas un mot su", or as it is said in English "vocabulary will not take care of itself". Cognitive anchoring (*deep level processing*, see Schmitt 2000) plays a very important role in vocabulary integration: in order to store new vocabulary in the memory, learners need to be exposed to it or to be confronted with it four to seven times on average (the exact number is subject to discussion). Learners need to encounter the new vocabulary several times in a relatively short time span, in various ways (e.g. in other contexts, in different class components, types of learning material, transferring it from active to passive use and vice versa) and always in significant and different contexts (see a.o. Schmitt 2000, Hulstijn & Laufer 2001, Lewis 2000). Learners remember new words more easily when they have pronounced, read and written them than when they were just repeated ten times (ibidem). In brief, vocabulary is not acquired through accumulation but via intermediary stages, through progressive, non-linear structuring.

The entries for the four languages (Dutch, English, French and Spanish) are linked for concepts present in all monolingual corpora. However, UrgentiAS also reflects the *specificity, and to a lesser extent the culture, of each language*. For instance, as English medical terms are more difficult to pronounce compared with the same terms in Spanish or French, audio archives with pronunciation were added to the English component (App. 3). On the other hand, the Spanish corpus reflects the geographical variety in the Hispanic world and the opposition between the more "scientific" tendencies in Romance languages versus the "vulgarizing" Germanic languages (e.g. "otite / otitis" versus "keelontsteking / inflammation of the ear").

In try-out sessions UrgentiAS proved to be a valuable learning tool and reference work suitable for use in the classroom (where the medical terms can be further contextualised) as well as for self-study. A certain amount of well-organised self-study saves *classroom time for communicative activities* in a setting of task-based language learning and content and language integrated learning[12], using a part of

12. *Task-based language learning* (TBLL), also known as *task-based language teaching* (TBLT) or *task-based instruction* (TBI) "focuses on the use of authentic language and on asking students to do meaningful tasks using the target language. Such tasks can include visiting a doctor, conducting

the lexicon as starting point: activities such as interaction with native speakers on predefined subjects, exploration of cultural aspects and medical contents of the lexicon, the use of audio and video material, as well as student presentations on certain parts of the lexicon, doctor-patient simulations on certain diseases defined in the lexicon, reading and producing different specialized text types. The latter is all too often neglected in LSP (Sanz 2001: 50): "It is doubtful that the phraseology has been presented in the classroom and/or in a lexicon, and it is almost certain that there was no room for either text types or communicative strategies that students need to express themselves adequately and efficiently in the context concerned" (Cabré & Gómez de Enterría 2006: 60; translation ours). In the classroom context, a global simulation of all situations in which the professional will perform is thus required, which, incidentally, also has a motivating effect on the learner (ibidem, p. 61 & 77). A special classroom activity with the lexicon concerns the periodical evaluations of the parts of the lexicon the students have to learn, testing in authentic contexts both active and passive knowledge, with or without the book open.

5. Corpus

To compile our corpus and the resulting database we started from the lexical input necessary for the clinical practice that students need to learn for their clinical placement, and which therefore needs to be taught in the Medical Spanish/ French/ English courses.

To obtain language material representative of the target domain, we compiled an *open, synchronic and multilingual corpus* with material from *different monolingual corpora,* in order to build a *multilingual database*[13]. In line with Bowker and Pearson (2002: 49–61), we compiled a *multi-author* corpus of *different types* of

an interview, or calling customer service for help. Assessment is primarily based on task outcome (in other words the appropriate completion of tasks) rather than on accuracy of language forms. This makes TBLL especially popular for developing target language fluency and student confidence" (http://en.wikipedia.org/wiki/Task-based_language_learning, accessed on 13–10–2010). *Content and Language Integrated Learning* (CLIL) "has become the umbrella term describing both learning another (content) subject such as physics or geography through the medium of a foreign language and learning a foreign language by studying a content-based subject" (http://www.teachingenglish.org.uk/think/articles/content-language-integrated-learning, accessed on 13–10–2010).

13. A *monolingual corpus* is "one that contains texts in a single language, while multilingual corpora contain texts in two or more languages; (...) a synchronic corpus presents a snapshot of language use during a limited time frame, whereas a diachronic corpus can be used to study how a language has evolved over a long period of time; (...) an open corpus, also known as a monitor corpus, is one that is constantly being expanded. This type of corpus is commonly used in lexicography because dictionary makers need to be able to find out about new words or changes

relatively recent *full texts,* such as textbooks, informative literature prepared for patients, instructional texts, protocols, newspaper articles and other popularized discussions, both in print and electronic.

Specialized texts differ from general texts in a number of respects, such as structure, style, terminology, theme and function (Cabré & Gómez de Enterría 2006: 32–36). In general, specialized texts centre on one specific theme, couched in objectifying language, with a vocabulary typical of the discipline and with unambiguous(ly used) terms. The degree to which such texts conform to this description depends on the author (expert, semi-expert of journalism) and the target audience, viz. experts, semi-experts (such as students) or a wide and more general audience; and the level of abstraction increases with the number of experts involved. On the other hand, recent students indicate that this assumption of "unambiguity" has to be qualified (Cabré & Gómez de Enterría 2006: 53 versus Gómez de Enterría 2001: 9): so-called specialized terms are in fact used in various domains, as well as in various situations within one specific domain; and they derive their transparency and unambiguous meaning only from the actual context or discipline in which they are used. The term *vector*, for instance, is used in both mathematics and physics, and the term *tone* is not only used in musicology, but also in, for example, acoustics. As a matter of fact, prior to an investigation of domain-specific terminology, the text types current in the domain have to be examined, as text type is highly dependent on discipline (see Parodi 2009). Because of the unique nature of specialized texts in general and the text types within a domain in particular, it is imperative that in a didactic context text types are studied. Following Cabré and Gómez de Enterría (2006: 62–63), we believe that students have to be offered all text types relevant to the "Special Purpose" (in this case medical language as used in the context of a clinical placement), ranging from highly specialized (characterized by a high degree of information density, precision, conciseness and objectivity, and hence a limited lexicon and syntax[14]) to more popularising and didactic, in which case these characteristics are less clearly present. Contrary to Gómez de Enterría (2001: 9), however, we found that loan words and loan translations from English, which are supposed to be typical of Spanish LSP (because of the slower technological progress in Hispanic countries) are *less* frequent in medical language than in general language use, a consequence of the Latin roots of most words. Another interesting finding that runs counter to

in meaning. In contrast, a closed or finite corpus is one that does not get augmented once it has been compiled". (Bowker & Pearson 2002: 12–13).

14. "Typical of the syntax of LSP texts is the frequent use of nominalisation, highly specific adjectives, verbs in the present tense and first person plural, impersonal constructions and the verb *to be*" (Gómez de Enterría 2001: 10; Gómez de Enterría 2009; our translation).

Gómez de Enterría (2009) is that medical texts in Spanish language areas are *not* shorter than general texts (see below).

This is why we examined the following different text types:

1. For French we contacted CENTAL (Centre de traitement automatique du langage, "Centre for computerised language processing" at UCL, the Belgian Catholic university of Louvain-la-Neuve) and obtained permission to use a corpus of 300,000 anonymised medical records from the Saint-Luc hospital. We also included the French translation of *Donde no hay doctor* (Mbow 1992), as well as several manuals (among others Fassier & Talavera-Goy 2008).
2. For Spanish we collected language material from clinical cases from medical professionals throughout the Hispanic world, from the Internet (e.g. www.fisterra.com and www.edadis.com) and from manuals (a.o. Florián Reyes & Ruiz Martínez 2005 and *Donde no hay doctor* (Mbow 1992)). Our Spanish corpus currently totals 300,000 words.
3. For English we worked in a similar manner, collecting language material both from books (among others Glendinning & Ron Howard 2007) and online sources (App. 4).
4. We also collected online materials from RSS feeds (for the methodology, see Fairon 2006 and Fairon, Macé & Naets 2008). Unfortunately, the data contained too much "noise" and "silence" for our purposes[15].

Thanks to an agreement with the K.U.Leuven University Hospital, we obtained anonymised medical records in Dutch, similar to the ones from UCL. In spite of this similarity, the medical records from the Dutch-speaking Belgian hospital are different in form from those of the French-speaking Belgian hospital: the records in Dutch are shorter and come in the form of an outline with less attention to context, whereas the records in French come in long, much more contextualised French (and Spanish) letters. Students in clinical placements abroad need to be aware of these and other *discourse differences*.

Since most of our data were collected from clinical cases, anamneses and diagnoses written by doctors and addressed to their colleagues, they only contain *specialized medical language*.

Because medical students in clinical placements also need to communicate with patients, *medical language used by and to patients,* as well as *cultural aspects of medical treatment* also needed to be included in our database. A fine example of this less scientific, popularised medical language for Spanish is found in the work of David Werner (the series *Donde no hay doctor,* see Werner 1973/1984), based

15. "Noise refers to unwanted items that are erroneously retrieved (i.e. patterns that *are not* terms), while silence refers to cases where patterns that *are* terms do not get retrieved" (Bowker & Pearson 2002: 169).

on field work in popular contexts, frequently in developing countries. The texts in this series have been written in simple language, easy to understand for people who have not been medically trained. These texts also contain Latin-American (frequently Mexican) *alternatives*, a very useful addition to the Spanish component for content and language integrated learning. They describe alternative treatments and beliefs in Latin-America: e.g. "Mexican villagers have the following home cures for poisonous snakebite: 1. to use 'guaco' leaves; 2. to bite the snake; 3. to apply tobacco; 4. to apply the skin of a poisonous lizard; 5. to smear the snake's bile on the bite" (English version of Werner 1984: 3). For the English component, information from the free online Merck Manual, as well as texts on doctor-patient forums (see App. 4), proved invaluable.

Following Bowker and Pearson (2002), we believe that the corpus compiled in the two-year project is sufficiently large to meet the needs of our project:

> There are no hard and fast rules that can be followed to determine the ideal size of a corpus. Instead, you will have to make this decision based on factors as the needs of your project, the availability of data and the amount of time you have. It is very important, however, not to assume that bigger is always better. You may find that you can get more useful information from a corpus that is small, but well designed than from one that is larger but is not customized to meet your needs. (...) In our experience, well-designed corpora that are anywhere from about ten thousand to several hundreds of thousands of words in size have proved to be exceptionally useful in LSP studies. (Bowker and Pearson 2002: 45–48)

The corpus is also designed as an open one, since "an open corpus that can be updated on a regular basis is likely to be more appropriate for many of your LSP needs (...), given the dynamic nature of LSP and the importance of staying abreast of current developments in the subject field" (Bowker & Pearson 2002: 12–13& 48; see also Torruella & Llisterri 1999: 50).

6. Corpus analysis

To analyse the corpus, with particular attention to the detection and extraction of the most frequent terms and collocations, we followed the basic steps of corpus extraction (see Bowker & Pearson 2002: 31–32). As our goal was to identify the most frequent terms, no time was invested in adding markup[16] to the corpus. After

16. Markup is a time consuming activity that can be used to (i) encode information about the appearance and structural components of documents, (ii) facilitate the exchange of documents between different computing systems (using SGML, HTML of XML) and (iii) make explicit the linguistic features of a text, such as the parts of speech ("POS tagging"). For more details, see Bowker and Pearson (2002: 75–91).

compiling a word list[17], we looked at these potentially interesting terms in context. To this end, the Leuven Language Institute's tools for corpus analysis (CATS, Corpus Tools Analysis) were used, as well as two concordancing programmes[18], viz. Unitex – especially for lemmatisation and searches with regular expressions and grammatical categories – and WordSmith – for collocations and examples. In order to avoid problems such as homonymy[19], the concordances were carefully examined, and we set a low minimum frequency value with a view to discovering important terms that do not occur frequently, as "it may well be the terms that are less commonly used that pose difficulties for LSP users because these are the words that may not be included in dictionaries or other published resources" (ibidem, p. 166). As we did not have a corpus of *language for general purposes* ("GPC", see before) for all languages at our disposal, we worked on the basis of not only Keywords (i.e. "words which occur with an unusually high frequency in a text or corpus when that text or corpus is compared with another [general purpose] corpus" (ibidem, p. 114–115), but also frequency lists. The advantage of this procedure is that words which *also* occur frequently in a GPC (the so-called *subtechnical terms*, see Vangehuchten 2005) are not overlooked. This is important because LSP users have to be able to understand and use not only the terms that occur *only* in a specific domain, but also the general words which are frequently used in it (also see Blanpain et al. 2008: 83). Furthermore, general words can acquire a different meaning in LSP than in LGP.

In the next move, the terms were sorted on the basis of frequency and alphabetical order (allowing us to group family words and prefixes, see above for their importance for medical purposes) and reverse alphabetical order (allowing us to retrieve important suffixes)[20]. Finally, frequent collocations were identified through frequent word clusters and the rearrangement of concordances by the first and second words to the left and to the right[21].

---

17. This is "a list that presents all the words in the corpus and indicates how often each appears. This list can be organized in such a way that the most frequently occurring words are presented at the top of the list" (Bowker & Pearson 2002: 31).

18. "A concordancer allows the user to see all the occurrences of a particular word in its immediate contexts. This information is typically displayed (...) in a KWIC display, [where] all the occurrences of the search pattern are lined up in the centre of the screen with a certain amount of context showing on either side. As with word lists, it is possible to sort concordances so that it becomes easier to identify patterns" (ibidem, p. 13).

19. "The state or quality of a given word's having the same spelling and the same sound or pronunciation as another word, but with a different meaning, as *race* ('tribe') and *race* ('running contest')" (http://www.thefreedictionary.com, accessed on 27–09–2009).

20. For more details on alphabetical and reverse word lists, see Bowker and Pearson (2002: 116–120).

21. For more details on clusters and concordances, see Bowker and Pearson (2002: 120–131).

	Words	Collocations
English	2934	1318
French	2324	1745
Spanish	2454	1506

Figure 1. Number of words and collocations

In order to structure the data, i.e. to arrange, label and present the lexical terms, we developed a flexible database combining MySQL and PHP scripts (originating from the online application PhpMyAdmin 2.10.0*)*. This application allows for continuous content management, updating and expansion.

The database, we feel, does not contain too many terms, making it a learning tool rather than an extensive and prohibitive traditional dictionary. However, the database is sufficiently extensive to offer our learners an answer to language problems they encounter frequently in their clinical placements abroad (see Figure 1).

The end-user interface allows the user to *search in different ways*: by card (through an alphabetic index), by (part of a) key word (regular search function), by medical specialisation or by chapter of the Medical French / English/Spanish course (App. 1). Searches can be performed in Dutch or in the target foreign language and can combine all the above search criteria.

Bearing in mind several criteria such as clarity, accessibility and ecological concerns, we also provided a compact print version as a user-friendly learning tool.

Finally, for the scientific / medical side of our database project, we relied on support from specialists in Belgium and abroad[22] because, as Bowker and Pearson (2002: 17) and Gómez de Enterría (2009) point out, expert advice is necessary if the lexicographers are not specialists in the subject field.

7. Evaluation

Towards the end of the project, the database was evaluated by the students of Medical English and Spanish. Students were encouraged to add comments and suggestions. The aim of the evaluation was to assess whether the use of the lexicon had altered students' perceptions of vocabulary acquisition, e.g. by taking more note of

22. One of the co-supervisors of the project is an M.D. affiliated with the University of Leuven, Faculty of Medicine. In addition, the Spanish section of the database has been double-checked by a Belgian doctor who, besides teaching at the University of Quito, has been running a health center in Ecuador for more than a decade, and thus has extensive experience with the Latin-American medical reality.

the context of a word, of the different aspects of one term, of the existence of related words, etc. As the French component of the database had not been sufficiently integrated in the course yet, the evaluation for Medical French was postponed to a later date.

From this evaluation we can conclude that UrgentiAS is a very powerful tool to learn how to use the new vocabulary. The fact that the vocabulary is embedded in a network of complementary information was appreciated: on average 95% of the students of Medical Spanish, who had been working with the database for 25 weeks at the time they participated in the evaluation, evaluated the tool positively. The students of Medical English had not worked with the lexicon at the time of the evaluation. They were given access to the database for half an hour to do a vocabulary exercise, but even this short span of time proved sufficient for them to respond positively (8 x 90%, 2 x 80%). For 4 questions in the evaluation, the time provided to work with the database had been too short and the responses to these items scored somewhat lower on average (70% for "It is easier for me to remember new vocabulary with UrgentiAS" and "I've become better at avoiding typical errors (pronunciation, spelling, conjugation, combinations...); 60% for "I now pay more attention to different aspects of a new word (article, gender, conjugation, preposition...) instead of just learning the word itself" and "I now pay more attention to the typical aspects of a language (e.g. typical pronunciation problems, terminological variation, typical abbreviations...)".

Moreover, comparison of test results for Medical Spanish suggests a significant improvement in language acquisition and skills: in the academic year 2005–2006, when students did not use the new lexicon, the results of test 2 (at the end of May) were hardly better than those of test 1 (at the beginning of January); whereas the results from the academic years 2006–2007 (with the experimental version of the lexicon) and those of 2007–2008 (with a first version of UrgentiAS) were not only 5% to 11% higher on average, the average difference in results between tests 1 and 2 improved by approximately 12% and 23% respectively. Furthermore, the number of morphological mistakes (i.e. against word form), mistakes against morphosyntaxis (i.e. the combination of words in the sentence) and mistakes against spelling had halved between the two test occasions. This indicates that students have become more aware of how to use new vocabulary in context: they used the words more correctly (spelling and morphology) and in combination with other words (grammar). Compared to the results of test 1 in 2005–2006, the results of this year (2008–2009) have improved by more than 25%.

Although the way in which UrgentiAS was used both in and outside class was the only difference with the classes of the years before, we need to be careful in comparing test results. The evaluation and comparison need to be continued before any clear conclusions can be drawn, but they provide important tools for us,

as the obligation to use the same learning materials and methods for all students in one academic year makes the use of other techniques to measure efficacy (such as control groups) impossible.

8. Conclusion

In conclusion, we are able to offer our students an *easy-to-use multilingual medical lexicon* firmly embedded in *semantic and syntactic context*. As it can be used both in the classroom and at home, it ensures that *students have enough 'significant encounters' with the target vocabulary* so that they can learn and *use these lexical structures effectively and accurately*.

In surveys among our target user group, based on try-out sessions, the tool was warmly welcomed. Most students were strongly convinced that the UrgentiAS lexicon can help them in their medical vocabulary learning process, an assumption that is backed up by the comparison of recent test results.

References

Balboni, P.E. 2000. *Le microlingue scientifico-professionali*. Torino: UTET Libreria Srl.

Blanpain, K., Heyvaert, L. & Laffut, A. 2008. Collex-Biz: A corpus-based lexical syllabus for Business English. ITL, *International Journal of Applied Linguistics* 155: 77–94.

Bowker, L. & Pearson, J. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

Buyse, K. & Thiry, B. 2008. ElektraRed®, NetVocList® et LinguaProLio®: Trois recettes de vitamines pour dynamiser la langue. *Le langage et l'homme* 43(2): 109–128.

Cabré, M.T. & Gómez de Enterría, J. 2006. *La enseñanza de los lenguajes de especialidad. La simulación global.* Madrid: Gredos.

Cervero, M.J. & Pichardo Castro, F. 2000. *Aprender a enseñar vocabulario*. Madrid: Edelsa/ Tandem.

Fairon, C. 2006. Corporator: A tool for creating RSS-based specialized corpora. In *Proceedings of the Workshop Web as corpus.* EACL 2006, Trento.

Fairon, C., Macé K. & Naets H. 2008. GlossaNet 2: A linguistic search engine for RSS-based corpora. In *Proceedings of LREC 2008*. Workshop WAC4, Marrakesh.

Fassier, T. & Talavera-Goy, S. 2008. *Le français des médecins: 40 vidéos pour communiquer à l'hôpital.* Grenoble: Presses universitaires de Grenoble.

Florián Reyes, L. & Ruiz Martínez, A.M. 2005. *Ciencias de la salud. Manual de español profesional. Nivel intermedio-avanzado. Libro del Alumno + Cuaderno de Ejercicios.* Alcalá: Universidad de Alcalá.

Glendinning, E.H. & Howard, R. 2007. *Professional English in Use: Medicine.* Cambridge: CUP.

Gómez de Enterría, J. (ed.). 2001. *La enseñanza/aprendizaje del español con fines específicos*. *Las lenguas de especialidad. Su aplicación a la enseñanza del español como lengua extranjera.* Madrid: Edinumen.

Gómez de Enterría, J. 2009. El lugar que ocupan las Lenguas de Especialidad para la Enseñanza del Español Lengua Extranjera. Keynote lecture on the XX International Conference of ASELE on "El español en contextos específicos: Enseñanza e investigación". Comillas, September 23–27, 2009.

Hulstijn, J.H. & Laufer, B. 2001. Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning* 51(3): 539–558.

Lewis, M. 2000. *Teaching Collocation: Further Developments in the Lexical Approach, LTP Teacher Training*. London: Language Teaching Publications.

Mbow, M.L. (ed.). 1992. *Là où il n'y a pas de docteur.* Dakar: Enda tiers-monde.

Nielsen, S. & Mourier, L. 2007. Design of a function-based internet accounting dictionary. In *Dictionary Visions, Research and Practice: Selected Papers from the 12th International Symposium on Lexicography, Copenhagen 2004* [Terminology and Lexicography Research and Practice 10], H. Gottlieb & J.E. Mogensen (eds), 119–139. Amsterdam: John Benjamins.

Parodi, G. 2009. Corpus, discurso y géneros: Español en contextos académicos y profesionales. Keynote lecture on the XX International Conference of ASELE on "El español en contextos específicos: Enseñanza e investigación". Comillas, September 23–27, 2009.

Roberts, R.P. 2007. Dictionaries and culture. In *Dictionary Visions, Research and Practice: Selected papers from the 12th International Symposium on Lexicography, Copenhagen 2004* [Terminology and Lexicography Research and Practice 10], H. Gottlieb & J.E. Mogensen (eds), 277–299. Amsterdam: John Benjamins.

Sanz, A. 2001. La enseñanza de las lenguas especiales a estudiantes extranjeros: el español de la medicina. In *La enseñanza/aprendizaje del español con fines específicos. Las lenguas de especialidad. Su aplicación a la enseñanza del español como lengua extranjera,* J. Gómez de Enterría (ed.), 41–55. Madrid: Edinumen.

Schmitt, N. & McCarthy, M. (eds). 1997. *Vocabulary, Description, Acquisition and Pedagogy*. Cambridge: Cambridge Language Teaching Library.

Schmitt, N. 2000. *Vocabulary in Language Teaching*. Cambridge: CUP.

Torruella, J. & Llisterri, J. 1999. Diseño de corpus textuales y orales. In *Filología e informática. Nuevas tecnologías en los estudios filológicos*, J.M. Blecua, G. Clavería, C. Sánchez & J. Torruella (eds), 45–77. Barcelona: Editorial Milenio.

Vangehuchten, L. 2005. El léxico del discurso económico empresarial: identificacion, selección y enseñanza en Español como lengua extranjera con fines especificos. Iberoamericana: Madrid.

Werner, D. 1973/1984. *Donde no hay doctor: Una guía para los campesinos que viven lejos de los Centros Médicos*. California: Fundación Hesperian. <http://www.healthwrights.org/spanish/Libros/doctor.htm>

Protocols. URLs:
<http://medlineplus.gov/spanish/>
<http://www.medicinainformacion.com/index.htm>
<http://www.fisterra.com/cursos/listaCasos.asp>
<http://escuela.med.puc.cl/paginas/Departamentos/CirDigestiva/ArchCasosEsp.htm>
<http://www.edadis.com>

Appendixes

Keys to the symbols in the appendixes:

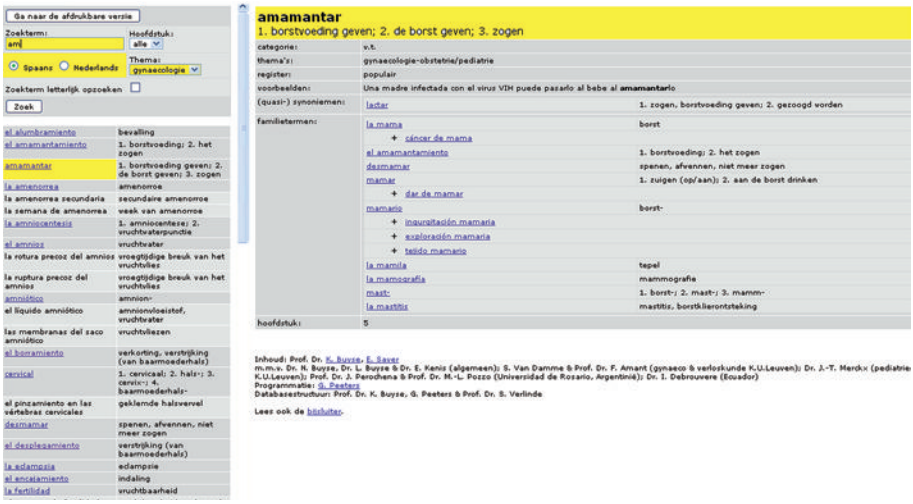
Electronic version:

(Dutch term)	***(English translation)***
Categorie	(grammatical) category
Thema's	themes, domains
Collocaties	collocations
Register	register
Verklaring	explanation
Gebruiksinfo	usage
Nota	note
(quasi-)antoniem	(near)-antonym
(quasi-)synoniem	(near)-synonym
Paroniem(en)	paronym
Voorbeelden	examples
Familietermen	family words
Hoofdstuk	chapter
Uitspraak	pronunciation

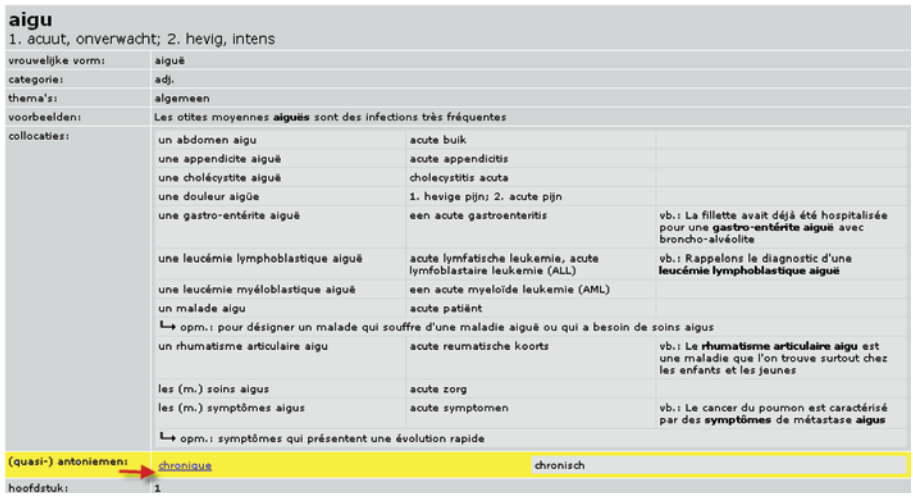
Print version:

≈	synonym
><	antonym
≠	paronym
fam.	family words
A, G...	abbreviation of the domain(s): general medicine, gynaecology...
Wetenschappelijk	scientific
Populair	popular
1, 3, 5, ...	number of the chapter in the coursebook

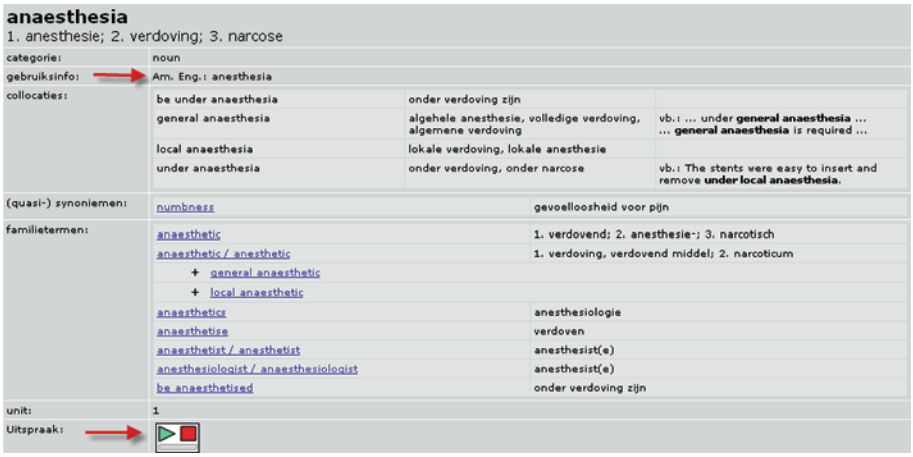
Appendix 1: electronic interface (Spanish), with search panel, word list and the terminological card of the verb "amamantar" (translation, category, domains, register, examples, (near-) synonyms, family words and chapter



Appendix 2: antonyms (French)



Appendix 3: geographical variants and pronunciation (English)



Appendix 4: corpora used for the English component of the UrgentiAS database

1. **Gynaecology/obstetrics**
   - Integrated Management of Pregnancy and Childbirth, Pregnancy, Childbirth, Postpartum and Newborn Care: A guide for essential practice.
     (23.000 words)
2. **Pediatrics**
   - Pedriatic surgery handbook
     (28.500 words)
3. **Internal medicine**
   - Yale university school of medicine heart book
     (203.000 words)
4. **General**
   - Where there is no doctor
     (127.000 words)
   - Course in Medical English (ILT course + Glendinning, Eric H. & Ron Howard (2007). Professional English in Use: Medicine. Cambridge: Cambridge University Press.
     (32.500 words)
   - The Liverpool Hospital Trauma Handbook
     (9.500 words)
   - Society of Teachers of Family Medicine
     (74.000 words)

5. **Merck Manual**
    1. Merck Cardiovascular Disorders (108.500 words)
    2. Merck Critical Care Medicine (160.000 words)
    3. Merck Ear, Nose, Throat and Dental Care (53.600 words)
    4. Endocrine and Metabolic Disorders (82.000 words)
    5. Gastrointestinal Disorders (71.500 words)
    6. Gynecology and obstetrics (87.000 words
    7. Hematology and oncology (75.800 words)
    8. Pediatrics (170.100 words)
    9. Injuries and Poisoning (71.800 words)
    10. Genitourinary disorders (76.800 words)
6. **Clinical cases**
    - Clinical cases_1: http://www.indiana.edu/~m555/cases/cases.html (6.100 words)
    - Clinical cases_2: http://medinfo.ufl.edu/year2/mmid/bms5300/cases/index.html (34.900 words)
      Clinical cases_3: several websites (11.400 words)
7. **Forums**
    - Netdoctor_discussions_children: 5.500 words
    - Netdoctor_askthedoctor_children
    - Heartsurgery_doctors: 31.700 words

Using natural language patterns for the development of ontologies

Elena Montiel-Ponsoda and Guadalupe Aguado de Cea
Polytechnic University of Madrid

The combination of certain linguistic units that recurrently appear in text genres has attracted the attention of many researchers in several domains, as they can provide valuable information about different types of relations. In this paper, the focus is on some of these combinatory units, referred to as Lexico-Syntactic Patterns (LSPs) that provide information about conceptual structures present in ontologies, also called Ontology Design Patterns (ODPs). The final end is to create a repository of LSPs associated to the ontological structures they convey to help novice users in the development of ontologies. In this paper we present the different strategies we have followed to identify LSPs, as well as an excerpt of the repository of LSPs-ODPs that is currently being built.

Keywords: Conceptual relations, Lexico-Syntactic Patterns, Ontological Structures, Ontology Design Patterns

1. Introduction

Recurrent combinations of certain linguistic units have long attracted researchers' attention both from the theoretical and practical point of view. Lexicographers, second language teaching professionals, among others, have analysed these combinations taking into account the morpho-syntactic, semantic and pragmatic features that these collocated units present. According to Aguado (2007: 182), depending on the combination of features adopted in collocational studies, the approaches comply mainly with syntactic criteria, lexico-syntactic criteria, semantic and pragmatic criteria, though other criteria such as statistical and conceptual (Hearst 1992; Meyer & Mackintosh 1996; Feliu & Cabré 2002) have also been applied. The purpose of these studies can range from lexicon building, dictionary making, language teaching to new areas in the natural language processing field, such as information extraction or ontology development, to mention just a few.

One of these areas is Terminology, in which conceptual relations play a decisive role, since they illustrate "the network of concepts underlying the terms of a domain" (Meyer 2001: 280). Concepts and conceptual relations together with their linguistic realizations, i.e. the terminological units, are the basic research objects in terminology. This author (2001: 280) also reports that there are two types of conceptual characteristics: (1) attributes (e.g. colour, height, weight) that hold for a certain concept without involving other concepts, and (2) relations, which link concepts and help to describe domain knowledge (e.g. hyponymy, meronymy, causality). In the context of Computational Terminology, the interest is addressed not only to the conceptual relations, i.e. semantic relations, but also to understanding, describing and formalizing their linguistic properties beyond their discovery capability (Auger & Barrière 2008: 8).

With the aim of describing a domain of knowledge we need to find out, first, the relevant concepts of the domain and their description, and second, how they are related to each other. In this paper we will deal with conceptual relations in that they are centred on how world objects are related to each other and on the lexical realizations that convey a certain relation between concepts. We will focus on taxonomy and meronymy relations as they are of the utmost importance in developing ontologies. We have followed a corpus-based onomasiological approach to pattern discovery, trying to identify all those patterns that provide the above-mentioned relations.

2. Identifying conceptual relations by means of linguistic patterns

Conceptual relations are defined by Feliu (2004: 27) as elements "that link two or more specialized knowledge units in a particular subject field", and they are formally represented as:

R (a b, n)

where *R* represents the relation, *a* and *b* are knowledge units, and *n* foresees the case when a relation links more than the two elements *a* and *b*. In her work, she analyzes verb-oriented conceptual relations, i.e., those relations in which verbs are the ones that convey a specific meaning. The objective of her research was to detect those "linguistic patterns" that expressed conceptual relations and to apply them in terminology extraction. A catalogue of linguistic patterns conveying the relations of similarity, inclusion, sequential processes, causality, instrument, meronymy, and association for the Catalan language is included in Feliu and Cabré (2002).

With a similar objective, Meyer (2001: 290) identified "knowledge patterns" with the aim of extracting terminology in a semi-automatic way. In Marshman

et al. (2002) knowledge patterns are defined as "words, word combinations or paralinguistic features of texts which frequently indicate conceptual relations", and are divided into:

a. Lexical knowledge patterns, which consist of words or groups of words.
b. Grammatical knowledge patterns, which involve combinations of grammatical categories.
c. Paralinguistic knowledge patterns, which are neither lexical nor grammatical, but include elements of text such as punctuation or parentheses.

Marshman et al. (2002) focused on the first group, the identification of "lexical knowledge patterns" used in French for conveying three types of conceptual relations: hyperonymy (*est un* / *type de* / *forme de* [is a/type of/form of]), meronymy (*consiste en* / *partie de* / *comporte* [consist of/part of/includes]), and function (*utilise pour* / *permet* / *function* [is used for/allows/function]).

Recently, Sierra et al. (2008) also identified "definitional contexts" in Spanish for the extraction of semantic relations with the aim of integrating them in a system for the automatic extraction of definitional knowledge. In those contexts, verbal patterns are the ones that convey the semantic relations between terms and their definitions. The authors classify the identified verbal patterns into four groups of "definitional verbal patterns": a. Analytical (such as *referir* [to refer to]); b. Functional (such as *servir* [to serve]); c. Meronymy-extensional (such as *componer* [to compound]); and c. Synomymy (such as *equivaler* [to be equivalent to]).

In Computational Linguistics, Hearst (1992) also identified some linguistic patterns with the goal of extracting information that would help in building up large lexicons for natural language processing. Hearst mainly focused on the automatic acquisition of hyponymy relations from texts by means of what she called "lexico-syntactic patterns". Hearst's patterns are said to be "easily recognizable, occur frequently and across text genre boundaries, and indisputably indicate the lexical relation of interest". The set of patterns identified by this researcher had the following characteristics: (1) they were directly extracted from texts, and (2) they had as main elements prepositional phrases, paralinguistic signs or conjunctions (but not verbs). Examples of Hearst's patterns are shown below:

NP[1] *such as {NP*$_1$*, NP*$_2$*... (and | or) NP*$_n$
NP[1] *{,NP}* {,} or other NP*
NP {,NP} {,} or other NP*
NP {,} including {NP,} { or | and } NP*
NP {,} especially {NP,} { or | and } NP*

1. NP: Noun Phrase

Since then, there have been many authors that have applied Hearst's patterns for the automatic discovery of lexical items, mainly in the Natural Language Processing field. In the next section, we will deal with the use of lexico-syntactic patterns in Knowledge Engineering.

3. Lexico-syntactic patterns in Knowledge Engineering

Ontologies are one of the central research subjects in Artificial Intelligence and Knowledge Engineering, since they allow to represent knowledge for machines and to add semantics to the information on the Web. Moreover, ontologies represent a domain of knowledge by defining the concepts of that domain and the relations among them. So far, work on ontology development could be identified with terminology work. However, ontologies go some steps further, in the sense that definitions of concepts and relations among them are formalized, which means that they are made understandable also to machines. And last but not least, the knowledge represented in an ontology captures the consensual knowledge of a community of domain experts. This has been summarized by Studer et al. (1998: 161) in one of the most cited definitions of ontologies that states that an ontology "(...) is a formal, explicit specification of a shared conceptualization" (based on Gruber (1993)).

Broadly speaking, ontologies consist of five main components: concepts, attributes, relations, instances, and axioms. Concepts identify types or classes of objects. Attributes refer to features or characteristics that define objects. Relations represent dependencies between concepts, or how concepts relate to each other. Instances are specific, real objects that belong to a certain class of objects. Finally, axioms can be defined as rules used to check constraints and to automatically infer knowledge in the ontology. Consider an ontology about cartoon animals, where *cartoon mouse* would be a type or subclass of *cartoon animal*, i.e., a concept in the ontology; *gender, size*, and *colour* of *mouse* would be the attributes; *cartoon mouse* could be related to *cartoon cheese* by means of the relation *eats*; a certain mouse called *Mickey* could be an instance of the concept *cartoon mouse*; and an axiom would establish that the relation *eats* from *cartoon mouse* could only go to *cartoon cheese*, and not to any other type of *cartoon food*. A common representation of an ontology has been included in Figure 1 with this highly simplified extract of a cartoon animal ontology.

As in Terminology, ontology development requires the discovery of the concepts of a specific domain, their properties, how they are related to each other, and the instances that belong to the identified concepts. Since this is a time and resource consuming activity, much effort has gone into the automatic acquisition of the different ontology elements from texts. For this purpose, linguistic patterns

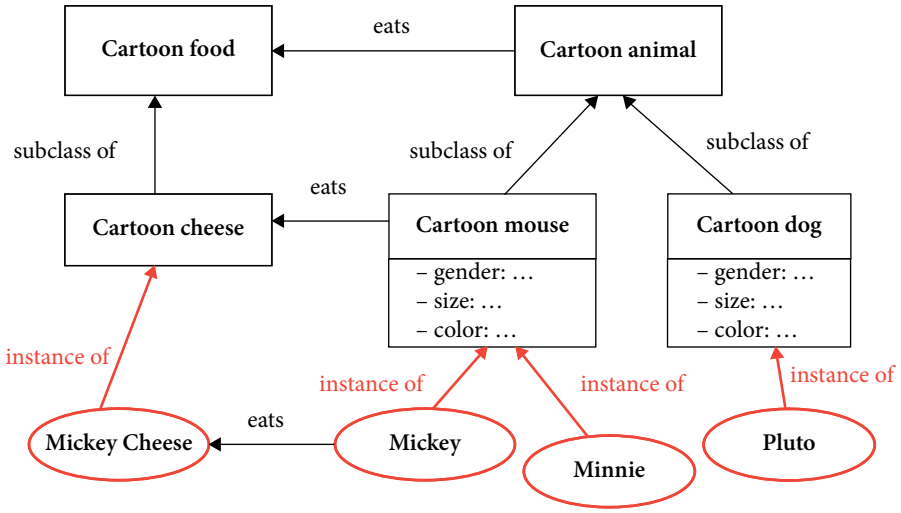


Figure 1. Example of the main ontology components

have been applied to extract ontology elements in order to speed up ontology development. Some researchers, Berland and Charniak (1999), Snow et al. (2004), or Cimiano and Wenderoth (2007), among others, have extended the original set of Hearst's patterns with additional ones that express hyponymy relations, or new ones expressing relations such as meronymy, agency, cause, etc. Some patterns were similar to Hearst's ones, that is, not verb-centred. Others had verbs as main elements.

In any case, no research has been oriented to the identification and use of linguistic patterns with the aim of helping users in ontology development, which is the aim of this approach. Basically, our hypothesis is that certain linguistic patterns can correspond to ontological relations which can be directly translated into formal representations in the ontology. These formal representations of ontological relations, which are expressed in the ontology by means of first-order logic predicates, will be referred to in this research as *ontological structures*. These structures are not sometimes limited to simple binary relations but they become complex structures with restrictions (axioms) and additional properties, as will be explained with more detail in the following sections.

In this research we have also adopted the name of Lexico-Syntactic Patterns (LSPs henceforth) as proposed by Hearst (1992), but we have redefined them, according to our purposes, as "linguistic schemas or constructs derived from recurrent expressions in natural language that consist of linguistic and paralinguistic elements that follow a certain syntactic order, and that permit extracting some conclusions about the meaning they express" (Aguado de Cea et al. 2008). Our LSPs are mainly verb-centred linguistic patterns in line with Feliu and Cabré

(2002), Marshman et al. (2002), and Sierra et al. (2008).The main objective of this research is to create a repository of LSPs associated to the ontological structures they express.

As a starting point for our research we consider those *ontological structures* that correspond to consensual modelling solutions, and that have received the name of Ontology Design Patterns (ODPs) in Knowledge Engineering. ODPs have been defined as "archetypal solutions to design problems" (Gangemi: 2005). In other words, ODPs are specific ontological structures that have shown to be suitable for solving a particular design problem, and that are described and presented to users in repositories or manuals, so that users can reuse them in their ontology development. Design patterns in general are deemed especially suitable for the development of resources for three main reasons: (1) they speed up the development process, (2) they encourage the reuse of best practices, and (3) they allow less experienced users to produce a better design (Prechelt: 1997).

Because of these reasons, we decided to focus on those ontological structures regarded as ODPs (although some of them coincide with simple ontology relations) and look for the linguistic patterns or LSPs that realize them in natural language. As a result we obtained a set of LSPs associated to ODPs and stored them in a repository that we have called LSPs-ODPs repository. This repository will be the grounding of a system that will permit identifying (semi-) automatically when a sentence introduced by the user corresponds to an LSP, and in its turn, to an ontological structure or ODP. An overview of the system for automatically recognizing ontological structures from natural language sentences is outlined in Figure 2. Both the repository of ODPs our research is based on, as well as the repository of LSPs-ODPs presented in this chapter, were developed within the European Project NeOn[2]. Some ODPs are already available online at www.ontologydesignpatternsn.org.

According to Suárez-Figueroa et al. (2007), ODPs can be divided into three main groups: Logical ODPs, Content ODPs and Architectural ODPs.

a. *Logical ODPs* include domain independent patterns, i.e., those patterns that can be used across domains. The most representative pattern in this group is the *Subclass of relation* pattern, which allows the construction of taxonomies or hierarchies in the ontology. Other patterns in this group are the *Datatype property* pattern, and the *Object property* pattern. The names of these patterns are related to the paradigm and implementation language in which most ontologies are nowadays built: OWL DL[3]. A *Datatype property* pattern is an

2. http://www.neon-project.org

3. The most popular Web Ontology Language (OWL) is based on Description Logics (DL) (for more details see http://www.w3.org/TR/owl-guide/)

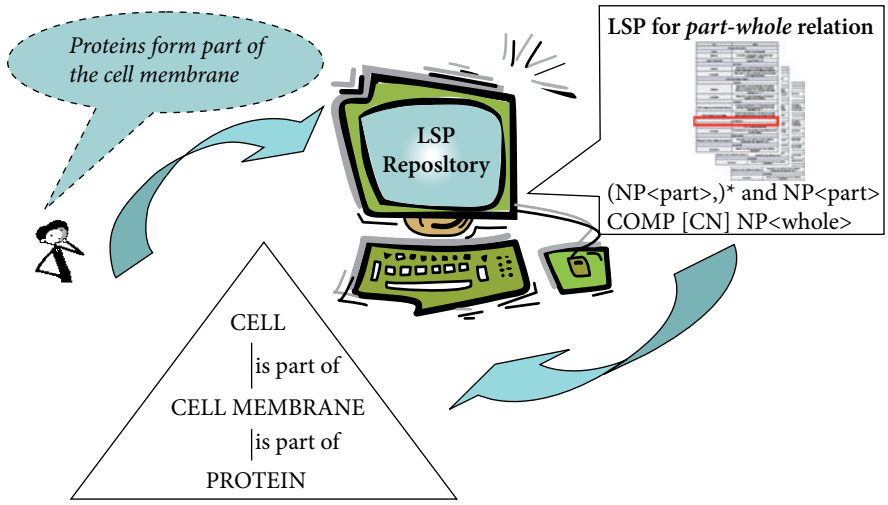


Figure 2. Overview of the system for LSPs recognition

ontological structure that permits relating a class with its attributes. For example, it would allow us to represent in the cartoon animal ontology, that *cartoon mice have gender*. *Object property* patterns permit to establish simple *ad hoc* relations between classes. For instance, the relation expressed by the verb to *eat* in *cartoon mice eat cartoon cheese*.

b. *Content ODPs* include domain specific patterns, i.e., patterns that allow representing specific relations that may only happen in certain domains. For example, the *Participation pattern*. We could make use of the Participation pattern in our ontology if we would need to represent that *cartoon animals participate in cartoon films during a certain period of time*. As we can see, these are more complex constructions that usually contain combinations of Logical ODPs. Within the *Content ODPs* group we find the so-called *Part-Whole relation* pattern. This pattern is responsible for the representation of meronymic relations, i.e., the ones that hold between an object and its parts. One could argue that this pattern should belong to the Logical ODPs. However, researchers on ODPs have decided to put it under this group, since the resulting structure is sometimes quite complex.

c. *Architectural ODPs* are also domain independent patterns, but they characterize the overall structure of the ontology. For the time being, these patterns are out of the scope of this research.

3.1 Difficulties of a natural language approach

Although being conscious of the fuzzy boundaries of natural language statements for conveying formal logical principles, as pointed out in Marshman (2008) quoting Lakoff (1975: 221), we claim the usefulness of the proposed natural language approach for identifying ontological structures for two main reasons.

The first reason is related to the use of precise and concise language when the user communicates with the system. The user is expected to realize in natural language those domain aspects he or she needs to store in the knowledge base with the aim of creating a sound resource that will provide him or her with powerful reasoning means. This would support the idea that expressions of uncertainty as those identified by Marschman (2008) would not come up when interacting with the system in the ontology development process. In her contribution, Marschman refers to quantifiers, hedging, modal verbs or negation as textual elements that may distort the message or render it useless when identifying information about conceptual relations. Let us go back to our toy ontology of cartoon animals. In that case, it would have no sense to introduce a sentence like: *It seems that cartoon mice eat cartoon cheese*, in the system, because "seems" indicates epistemic modality, which is not the aim of ontology development, but rather to model a specific domain. So we should reformulate the input ignoring this kind of expression. However, novice users in ontology development may be tempted to use some of these textual elements when expressing what they want to model in the ontology. Therefore, some recommendations or guidelines as to the linguistic expressions used will be offered to the user before proceeding with ontology development.

The second reason for adopting a natural language based approach is that we contend that fuzzy input statements will not arise. On the one hand, as mentioned above, the guidelines will provide clear information to the user before using the system, and, on the other hand, all the previous steps the user has to go through before performing the ontology conceptualization *per se* will prevent him or her from using uncertainty expressions. As detailed in most ontology development methodologies[4], users are supposed to have followed some steps in the conceptualization phase, before we expect them to be expressing in natural language what they want to model in the ontology. Those steps are mainly related to:

a. The specification of the requirements the ontology has to meet. For example, what uses the ontology is going to have, who the target users will be, etc.
b. The elicitation of the knowledge to be included in the ontology. For example, in the illustrated cartoon animal ontology, the information to be included would be *cartoon animal*, *cartoon food*, *cartoon mouse*, *Mickey*, *Minnie*, etc.

4. For a review of ontology development methodologies see Gómez-Pérez et al., (2003)

These are not trivial tasks, and domain experts normally need the assistance of ontology engineers to carry them out. However, once those previous activities have been accomplished, the expression of the specific knowledge that has to be included in the ontology will be more reliable. The argumentation presented here needs more evidence, and research is currently being conducted.

4. Strategies in the discovery of LSPs

At this stage, the methodology applied for extracting natural language expressions equivalent to ontological structures or ODPs, and transformed afterwards in LSPs, conformed to the following strategies:

a. To select available verb-centred patterns in literature and adapt them to our notation schema, following symbols and abbreviations based on a well known notation form in Computer Science, the Backus-Naur Form[5] (see Table 6 in the Appendix). For instance, one of Cimiano and Wenderoth's patterns (2007) that expresses the *subclass of* relation, "NP_{QT} is a kind of NP_F", would be transformed into "[(NP<subclass>,)* and] NP<subclass> be [CN] NP<superclass>", according to our notation (see LSP 1 in Table 1).
b. To identify ontology related concepts, and search the Web for common verbal constructs linking them according to an onomasiological approach. As input resources we used the WordNet[6] lexicon and the AGROVOC[7] thesaurus, which contain semantically related lexical items or terms. For example, we took the terms *protein* and *cell* from WordNet, conscious of the *part-whole* relation that holds between them, and introduced both terms in a search engine. We obtained sentences like: "A typical human cell contains millions of these proteins (...)", corresponding to our LSP 2 in Table 4 for the part-whole relation: "NP<whole> COMP [PARA] [(NP<part>,)* and] NP<part>".
c. To search in domain descriptive and encyclopaedic documents for verbal constructs that link concepts according to the ontological relations in ontology constructs. For this goal, we used descriptive documents from the Web, specially Wikipedia[8], and documentation used in the development of ontologies in some European projects such as Esperonto[9] and NeOn.

5. http://cui.unige.ch/db-research/Enseignement/analyseinfo/AboutBNF.html
6. http://wordnet.princeton.edu/
7. http://www.fao.org/aims/ag_intro.htm
8. http://es.wikipedia.org
9. http://www.esperonto.net/semanticportal/jsp/frames.jsp

As a result of applying these strategies we obtained a set of sentences in which concepts were related by means of different verbal constructs. We could state that some sentences corresponding to the same ontological structures followed a similar schema, despite concepts coming from different domains of knowledge. For illustrating this process, let us consider two sentences in English from different knowledge areas expressing the *Subclass of relation* ODP:

1. Animals are divided into two major categories: vertebrates and invertebrates.
2. Medications are generally classified into two groups: over-the-counter (OTC) medication and prescription only medicines (POM).

The verbs *divide into* and *classify into* indicate a *subclass of* relation, in which the so-called superclass is at the left-hand side of the verb, and the subclasses are at the right-hand side of the verb. In fact, there is a group of sentences constructed in a similar way from which we could draw an LSP embracing all of them:

NP<superclass> CATV [CD] [CN] [PARA] (NP<subclass>,)*and NP <subclass>

In principle, we can say that this LSP has a direct correspondence to the ODP representing the *subclass of relation*. Should the user introduce in the system a sentence in natural language which matches this LSP, the information contained in the sentence would be transformed into an ontological structure modelling that specific knowledge. However, this correspondence is not always so flawless, and difficulties are mainly imposed by ontology modelling needs on the one hand, and language ambiguities on the other, as will be explained in more detail in the next section.

5.1 The challenge of matching LSPs and ODPs

When specifying the subclass relation that defines hierarchies in ontologies, the user can access the *subclass of relation* ODP. Then, it is very important to determine if the identified subclasses that belong to the same superclass can share instances, i.e., if subclasses are disjoint (as pointed out in Gómez-Pérez et al. 2003: 134). In fact, *disjoint classes* is a Logical ODP that allows to represent a set of disjoint classes. Going back to our cartoon animal ontology, *cartoon mouse* and *cartoon dog* could be said to be disjoint classes, since no instance of the *cartoon mouse* class can be at the same time an instance of *cartoon dog*. In addition to disjointness, there is a further characteristic of the *subclass of* relation that has to be made explicit when developing ontologies, namely exhaustiveness (*ibidem*). This is related to the possibility of identifying if the set of spotted subclasses are all the possible ones to be covered by the superclass. There is as well a Logical ODP for representing *exhaustive classes*. In our example, *cartoon mouse* and *cartoon dog*

could be said not to exhaustively reflect the subclasses of *cartoon animal*, since we would be excluding too many candidates.

However, both specifications of the *subclass of relation* are needed for assuring correctness in subsequent reasoning possibilities offered by the ontology. In fact, this is highly recommended in ontology modelling, although not so obvious for untrained users in ontology engineering who may simply not be aware of the fact that the omission of this kind of information can lead to inconsistencies. For example, should we forget to define *cartoon mouse* and *cartoon dog* as disjoint classes, this could lead to an inconsistency statement of the sort: *Mickey is a cartoon mouse and a cartoon dog*. Therefore, in our purpose of assisting novice users in ontology development with a natural language based approach, we would need to identify those linguistic elements that implicitly convey that meaning. For instance, from a sentence like *Cartoon animals are either cartoon mice or cartoon dogs*, though being false, we could derive an LSP that would correspond to the *subclass of relation* ODP, plus the *disjoint classes* ODP and the *exhaustive classes* ODP, as explained in the next section.

Regarding language ambiguities, these can also pose some problems to the direct correspondence between LSPs and ODPs. If we take some verbs such as *divide* or *include*, they can realize a *subclass of relation* as in sentence (1) "Animals are divided into two major categories: vertebrates and invertebrates", or they can indicate a *part-whole relation* in certain contexts, as in "The cerebrum is divided into two major parts: the right cerebral hemisphere and the left cerebral hemisphere". For those reasons, we may encounter LSPs that can correspond to different ontological structures, and that will require disambiguation. See Table 3 for more examples.

6. Lexico-syntactic patterns for the *subclass* of and the *part-whole relations*

In this section, our aim is to show an excerpt of the LSPs-ODPs repository we are currently developing. We try to collect some of the most usual ways a language has for expressing the ontological structures that are considered consensual modelling solutions. This repository is currently in a more advanced stage for English, but we mean to develop an equivalent repository for the Spanish and the German languages. The motivation behind a multilingual repository is that we believe that users should be allowed to develop ontologies in their own language. Although for the time being the set of discovered LSPs is not exhaustive, it aims at being representative of the most typical ways in which a language can express the identified structures.

In Table 1 we present the set of LSPs for the *subclass of relation* ODP, identified up to now for the English language (EN) as the *LSP Identifier* indicates. For each of the identified LSPs, we have added an example in natural language for the sake of clarity.

Table 1. LSP for the *subclass of relation*

LSP Identifier: *LSP – Subclass of – EN*
Formalization
1. [(NP<subclass>,)* and] NP<subclass> be [CN] NP<superclass>
2. [(NP<subclass>,)* and] NP<subclass> (classify as) | (group in|into|as) | (fall into) | (belong to) [CN] NP<superclass>
3. There are CD | QUAN [CN] NP<superclass> PARA [(NP<subclass>,)* and] NP<subclass>
4. [A(n) | QUAN] CN of NP<superclass> be | include [(NP<subclass>,)* and] NP<subclass>
Examples
1. *Odometry, speedometry and GPS are types of sensors.*
2. *Thyroid medicines belong to the general group of hormone medicines.*
4. *There are several kinds of memory: fast, expensive, short term memory, and long-term memory.*
5. *Some examples of peripherals are keyboards, mice, monitors, printers, scanners, disk and tape drives, microphones, speakers, joysticks, plotters and cameras.*

Let us take Example 1 from Table 1, "Odometry, speedometry and GPS are types of sensors". Without further information, we are not able to determine whether the list of sensors compiled in the example is exhaustive or not. In this sense, some lexical markers could indicate that kind of information, such as the use of the adverb *only*, or its synonyms *solely* and *exclusively*. In other words, these kinds of adverbs reflect exhaustiveness as can be seen in the following example: "Odometry, speedometry and GPS are *the only* types of sensors". Moreover, exhaustiveness may also be assured when the number of subclasses is explicitly mentioned as in the following example "Sensors are divided into *two* groups: contact and non-contact sensors".

If these linguistic elements, adverbs or adverbial expressions are not explicit, our system would need to rely on user interaction to identify the ontological structures corresponding to certain natural language realizations and specify if the subclasses are all the possible ones corresponding to the main class. For that purpose, we are developing a further strategy, in which whenever a subclass of relation is identified by the system, it systematically asks the user if the list of elements or subclasses in the sentence is complete or exhaustive.

Regarding disjointness, it is our claim that whenever types, classes, categories, etc., are enumerated and introduced by verbal markers, such as *classify into* or *divide into*, they refer to sets that cannot share instances. This assumption has been validated in an initial experiment presented in (Aguado de Cea and Montiel-Ponsoda, 2009), in which the British National Corpus[10] was employed to search for

10. http://www.natcorp.ox.ac.uk/

sentences including the verbal forms: *classify into, divide into, split into*, and *separate into*. Nearly 90% of all the analyzed sentences showed that classes were pairwise disjoint, according to real knowledge of the world. This statement was confirmed in those cases in which some adjectives appeared modifying the so-called Class Name (CN), such as "...classified into two *distinct groups*..." or the superclass, as in "...divided into two *separate scientific communities*...". Apart from the enumeration criteria and the use of some adjectives that modify the terms designing classes or class roles, the conjunction *either...or*, or *neither...nor* has also been identified as straightforwardly pointing to disjointness (Völker et al. 2007). However, for the remaining cases in which disjointness may not seem so obvious or cannot be assured, user interaction is also foreseen.

Because of these particular characteristics of the *subclass of relation*, we decided to create a specific section for those LSPs that could be directly identified with the *subclass of relation* ODP, the *disjoint classes* ODP and the *exhaustive classes* ODP (cf. Table 2). Sentences matching these LSPs could be said to refer to subclasses contained in a superclass not sharing instances, and being covered by the superclass.

Leaving apart the specificities of certain ontological relations that need to be made explicit in the ontology, a further challenge of any natural language-based approach is polysemy. In the present approach, polysemy has to do with the fact that the same linguistic structure can be identified with different ontological structures. Differences in the conveyed meaning are clear for humans, but not for machines. In Table 3 we have included ambiguous LSPs corresponding either to the *subclass of relation* ODP or the *part-whole relation* ODP.

Strategies involving user interaction are also being researched to disambiguate between *subclasses* and *parts* conveyed by the same linguistic patterns, in this case, the polysemous verbs in Table 3. User interaction would consist in launching a question to the user whenever one of the LSPs in Table 3 is matched. However,

Table 2. LSPs for the ontological structures: *Subclass of, disjoint* and *exhaustive classes*

LSP Identifier: *LSP- Subclass of – Disjoint classes – Exhaustive classes - EN*
Formalization
1. NP<superclass> be [either] NP<subclass> or | and NP<subclass>
2. NP<superclass> be CATV | divide | split | separate | group in|into CD CN [PARA] [(NP<subclass >,)* and] NP<subclass>
Examples
1. *Animals are either vertebrates or invertebrates.*
2. *Sensors are divided into two groups: contact and non-contact sensors. Membrane proteins are classified into two major categories, integral proteins and peripheral proteins.*

Table 3. LSPs for the *subclass of* and *part-whole* ontological structures

LSP Identifier : *LSP – Subclass of / Part whole – EN*
Formalization
1. NP<class> include | comprise | consist of [(NP<class >,)* and] NP<class>
2. NP<class> be divide | split | separate in|into [CD] [CN] [(NP<class >,)* and] NP<class>
Examples
1. a. *Common mass storage devices include disk drives and tape drives.* (Subclass of)
b. *Reproductive structures in female insects include ovaries, bursa copulatrix and uterus.* (Part-whole)
2. a. *Marine mammals are divided into three orders: Carnivora, Sirenia and Cetacea*. (Subclass of).
b. *The cerebrum is divided into two major parts: the right cerebral hemisphere and left cerebral hemisphere.* (Part-whole)

other strategies are also being considered to automate the process and avoid constant interaction with the user. A possible strategy would consist in accessing ontology libraries and looking for the two components of the relation and how they have been related in other ontologies. Similarly, this could be done by accessing other kinds of resources containing some semantics, such as lexicons or thesauri. Let us take sentence 2 (b) "The cerebrum is divided into two major parts: the right cerebral hemisphere and left cerebral hemisphere". By looking at "cerebrum" and "hemisphere" in a lexicon such as WordNet, we find that the relation holding between them is a part-whole relation. In default of data in such resources we would have to resort to user interaction.

Finally, we present those LSPs identified in English for expressing meronymy (see Table 4). As the reader can observe from the listed examples, one of the most usual linguistic patterns to express parts of an object is missing, namely, *to have*. The reason for this is that *to have* can refer as well to attributes of classes, represented by the *Datatype property* ODP, or the relation of possession, which would be represented by the *Object property* ODP, as an *ad hoc* relation (see Table 5). Let us take the example "Birds have feathers" in Table 5. This could be modelled in the ontology with the *part-whole relation* ODP, or with the *Object property* ODP. The final decision will depend on how the user wants to represent this knowledge in the ontology (as parts of birds or as a relation of possession), and on further ontology modelling decisions. In this case, interaction with the user is also needed to solve the ambiguity.

Table 4. LSPs for *part-whole relation* ontological structure

LSP Identifier: ***LSP- Part whole – EN***
Formalization
1. [(NP<part>,)* and] NP<part> COMP NP<whole>
2. NP<whole> COMP [PARA] [(NP<part>,)* and] NP<part>
3. [The] part of a NP<whole> be [PARA] [(NP<part>,)* and] NP<part>
4. [(NP<part>,)* and] NP<part> be part of a NP<whole>
Examples
1. *Proteins form part of the cell membrane.*
2. *Water is made up of hydrogen and oxygen.*
3. *The parts of a tree are the root, trunk(s), branches, twigs and leaves.*
4. *A device is any part of a computer.*

Table 5. LSPs for the ontological structures: *Datatype property, object property* or *part-whole relation*

LSP Identifier : ***LSP – Datatype property/Object Property/Part whole – EN***
Formalization
1. NP<class> have NP<class>
Examples
1. *Birds have feathers.* *Water areas have names in natural language.*

7. Potential applications of the LSPs repository to ESP/EAP

In the research work presented here, we have focused on the application of the LSPs repository to the development of ontologies by inexperienced users. However, linguistic patterns, as a combination of lexis, syntax and semantics, have proven to be extremely useful resources in a wide range of domains ranging from Terminology and Translation to the Teaching of Specialized Languages.

As already mentioned in the introduction, these lexico-syntactic patterns can help terminologists to recognise the conceptual relations when modelling the conceptual map of a specialized domain. Moreover, since terminological acquisition is both a linguistic and cognitive process, these patterns can help to speed up the identification of the terms and terminological units that define a domain, mainly by using some software applications such as WordSmithTools©. In the same sense, lexicographers could also apply definitional patterns for the extraction and

enrichment of definitions in dictionaries in a semi-automatic way. In fact, the set of patterns that have been identified in this research represents a subset of the so-called *definitional verbal patterns* (Sierra et al. 2008), since they help to detect hyperonymy-hyponymy as well as meronymy relations.

For translators (and interpreters) that are not domain experts, it is essential to quickly acquire the background knowledge of certain subject fields if they aim to succeed in the process of transferring knowledge and communicating effectively. These definitional and classification patterns, that are used across domains, can trigger the translator's ability to capture the cognitive schemas of a domain alongside the phraseology used for this purpose in bilingual texts.

Finally, in the Languages for Specific Purposes field, classification is one of the five rhetorical functions (Trimble 1985) found in academic writing for the transmission of knowledge in technical language. These rhetorical functions are linguistically realized by means of certain patterns that are present in all academic writing courses. Thus, on the one hand, the LSP teacher should make students aware of the structuring function of such patterns that allow specialists to organize and structure the knowledge of a certain subject field. Furthermore, non-native students of a technical discipline need to master these patterns if they want to reproduce them when writing their reports, dissertations or assignment tasks as they are normally used to convey knowledge, move from the *known* to the *unknown*, or from the more general to the more specific. On the other hand, these patterns can undoubtedly help learners to gain an overview of the domain, even if their objective is not to become experts in the subject field.

8. Conclusions

Linguistic patterns in Terminology and Knowledge Engineering have proven to be highly beneficial for extracting valuable information, thus speeding up terminology and ontology work. Our proposal of identifying Lexico-Syntactic Patterns (LSPs) that correspond to ontological structures or Ontology Design Patterns (ODPs) can help users in the development of ontologies by using a system that permits detecting automatically the ontological relation expressed in the sentence introduced by the user. The core of this research is the repository of LSPs matched to ODPs that will enable to automatically recognize ontological structures as they match ODPs. We have also analyzed some of the problems that LSPs present regarding exhaustiveness, disjointness and language polysemy, and have explained some of the solutions that are currently being investigated. At present, the LSP repository is being extended to cover a subset of Logical ODPs and Content ODPs in English. We also plan to enhance the repository with LSPs for Spanish and German.

Acknowledgements

The research described in this paper is supported by the European project *Lifecycle support for networked ontologies (NeOn)* (FP6–027595), and the Spanish project *GeoBuddies* (TSI2007–65677C02). In addition, it is partially co-funded by an I + D grant from the *Universidad Politécnica de Madrid*. We would also like to thank Mari Carmen Suárez-Figueroa for her help with ontological aspects.

References

Aguado de Cea, G. 2007. A multiperspective approach to specialized phraseology: Internet as a reference corpus for phraseology. In *The Texture of Internet: Netlinguistics in Progress*, S. Posteguillo, M.J. Esteve & M.L. Gea-Valor (eds), 182–207. Newcastle: Cambridge Scholars.

Aguado de Cea, G., Gómez-Pérez, A., Montiel-Ponsoda, E. & Suárez-Figueroa, M.C. 2008. Natural language-based approach for helping in the reuse of ontology design patterns. In *Knowledge Engineering: Practice and Patterns*, *Proceedings of the 16th International Conference on Knowledge Engineering (EKAW)*, 32–47. Berlin: Springer.

Aguado de Cea, G. & Montiel-Ponsoda, E. 2009. When phraseology and ontologies meet. In *Proceedings of the XI Simposio Internacional de Comunicación Social*, 80–84.

Auger, A. & Barrière, C. 2008. Pattern-based approaches to semantic relation extraction. *Terminology* 14(1): 1–19.

Berland, M. & Charniak, E. 1999. Finding parts in very large corpora. In *Proceedings of the 37th Annual Meeting of the ACL*: 57–64.

Cimiano, P. & Wenderoth, J. 2007. Automatic acquisition of ranked qualia structures from the web. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*: 888–895.

Feliu, J. & Cabré, M.T. 2002. Conceptual relations in specialized texts: New typology and an extraction system proposal. *TKE 2002*. Nancy, 45–49.

Feliu, J. 2004. Relacions conceptuals i terminologia: Anàlisi i proposta de detecció semiautomàtica. PhD dissertation, Universitat Pompeu Fabra.

Gangemi, A. 2005. Ontology design patterns for semantic web content. In *Proceedings of the 4th ISWC*, M. Musen et al. (eds), 262–276. Heidelberg: Springer.

Gómer-Perez, A., Fernández-López, M. Corcho, O. 2003. *Ontological Engineering*. New York: Springer.

Gruber, T.R. 1993. Toward principles for the design of ontologies used for knowledge Sharing. In *International Workshop on Formal Ontology*, N. Guarino & R. Poli, (eds), Padova, Italia.

Hearst, M.A. 1992. Automatic acquisition of hyponyms from large text corpora. In *14th International Conference on Computational Linguistics*, 539–545.

Lakoff, G. 1975. Hedges. A Study in meaning criteria and the Logic of Fuzzy concepts in D. Hockney, W. Harper and B. Freed (eds). *Contemporary Research in Philosophical Logic and linguistic Semantics*. 221-271. Dordrecht: Reidel.

Marshman, E., Morgan T. & Meyer, I. 2002. French patterns for expressing concept relations. *Terminology* 8(1): 1–29.

Marshman, E. 2008. Expressions of uncertainty in candidate knowledge-rich contexts. A comparison in English and French specialized texts. *Terminology* 14, 124–151.

Meyer, I. & Mackintosh, K. 1996. Refining the terminographer's concept-analysis methods: How can phraseology help? *Terminology* 3(1): 1–26.

Meyer, I. 2001. Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework. In *Recent Advances in Computational Terminology* [Natural Language Processing 2], C. Bourigault, C. Jacquemin & M.-C. L'Homme (eds), 279–303. Amsterdam: John Benjamins.

Prechelt, L. 1997. An experiment on the usefulness of design patterns: Detailed description and evaluation. Technical Report. University of Karlsruhe.

Scott, M., 2004. *Wordsmith Tools* version 4. Oxford: OUP.

Sierra, G., Alarcón, R., Aguilar, C. & Bach, C. 2008. Definitional verbal patterns for semantic relation extraction. *Terminology* 14(1): 74–98.

Snow, R., Jurafsky, D. & Ng, A. Y. 2004. Learning syntactic patterns for automatic hypernym discovery. In *Advances in Neural Information Processing Systems* 17, L.K. Saul, Y. Weiss & L.Bottou (eds), 1297–1304.

Studer, R., Benjamins, R. & Fensel, D. 1998. Knowledge engineering: Principles and methods. *Data & Knowledge Engineering* 25(1–2): 161–198.

Suárez-Figueroa, M.C., Brockmans, S., Gangemi, A., Gómez-Pérez, A., Lehmann, J., Lewen, H., Presutti, V. & Sabou, M. 2007. NeOn D5.1.1. *NeOn Modelling Components.* NeOn Project.

Trimble, L. 1985. *English for Science and Technology*. Cambridge: CUP

Völker, J., Vrandecic, D., Sure, Y. & Hotho, A. 2007. Learning Disjointness. In *Proceedings of the ESWC2007*, E. Franconi, M. Kifer & W. May (eds), 175–189. Berlin: Springer.

Appendix

Table 6. Symbols and Abbreviations used in the description of LSPs

Symbols & abbreviations	Description
AP<…>	Adjectival Phrase. It is defined as a phrase whose head is an adjective accompanied optionally by adverbs or other complements as prepositional phrases. AP is followed by the semantic role played by the concept it represents in the conceptual relation in question in <…>, such as e.g., *property*.
CATV	Verbs of Classification. Set of verbs of classification plus the preposition that normally follows them. Some of the most representative verbs in this group are: *classify in/into, categorize in/into, subclassify in/into, subcategorize in/into.*
CD	Cardinal Number.
CN	Class Name. Generic names for classes usually accompanied by preposition, such as *class, group, set, type, kind, category, species, sort, example.*
COMP	Verbs of Composition. Set of verbs meaning that something is made up of different parts. Some of the most representative ones are: *contain, hold, consist of, compose (of), make up (of), form (of/by), form part of, constitute (of/by)*.
NP<…>	Noun Phrase. It is defined as a phrase whose head is a noun or a pronoun, optionally accompanied by a set of modifiers, and that functions as the subject or object of a verb. NP is followed by the semantic role played by the concept it represents in the conceptual relation in question in <…>, e.g., *class, subclass*.
PARA	Paralinguistic symbols like colon, or more complex structures as *as follows*, etc.
PREP	Prepositions
QUAN	Quantifiers such as *all, some, most, many, several, every,* etc.
REPRO	Relative pronouns such as that, which, whose.
( )	Parentheses group two or more elements.
*	Asterisk indicates repetition.
[ ]	Elements in brackets are meant to be optional, which means that they can be present either at that stage of the sentence or not, and by default of appearance, the pattern remains unmodified.
¬	Elements preceded by this symbol should not appear in the pattern.

Notes on contributors

Douglas Biber is Regents' Professor of English (Applied Linguistics) at Northern Arizona University. His research efforts have focused on corpus linguistics, English grammar, and register variation (in English and cross-linguistic; synchronic and diachronic). He has written numerous books and monographs, including academic books published with Cambridge University Press (1988, 1995, 1998, 2009), John Benjamins (2006, 2007), the co-authored Longman Grammar of Spoken and Written English (1999), and three grammar textbooks published by Longman.

Bethany Gray is a PhD candidate in Applied Linguistics in the Department of English at Northern Arizona University. Her research investigates register variation through corpus linguistics, and her recent work has focused on documenting synchronic variation across written academic registers and disciplines. Her collaborative efforts have examined the differing grammatical complexities of spoken and written English, and have tracked historical change in scientific research writing through diachronic comparisons with other historical registers. This research has particularly focused on the development of complex noun phrases in academic prose. She is also interested in the lexico-grammatical marking of stance, cohesion in writing, and recurrent discontinuous lexical sequences. These research efforts have resulted in published book chapters and research articles in journals such as Journal of English for Academic Purposes, English for Specific Purposes, and TESOL Quarterly.

Carmen Pérez-Llantada is Professor of English at the University of Zaragoza (Spain). She is EAP instructor at both degree and master levels. Her research interests are rhetorical and genre analyses of ESP discourses with a particular focus on the spread of English as a global lingua franca and the linguistic, intercultural and sociolinguistic implications. At present, she is involved in the research project entitled "Generic integrity of academic and professional communication: An analysis of genres and its correlation with the discourse practices and cultural identities of different professional communities", funded by the Spanish Ministry of Science and Education (FFI2009–09792).

Sara Gesuato is associate professor of English at the University of Padua, Italy, where she teaches English language and linguistics. Her research interests include

academic discourse, text analysis, lexical semantics and tense and aspect. Recently, she has examined the structure and wording of oral and written expressive speech acts, investigated the phraseology and content of academic genres, studied the temporal and aspectual meanings of catenative motion verbs, and carried out corpus-based analyses of near-synonyms and of compounds with singular vs plural nominal premodification. She is currently working on pedagogical applications of speech act analysis and the phraseologies of anticipatory-it constructions.

Mercedes Jaime-Sisó (PhD) is a Senior Lecturer at the University of Zaragoza, where she has been teaching ESP at the Faculty of Veterinary Sciences since 1988. At present she also teaches academic scientific writing in two master courses. As member of InterLAE (Interpersonality in Written Academic Language) research group, her current research focuses on the evolution of scientific RAs structure and titles in online publications, and the integration of ITC and language learning. She has recently contributed to the development of a web-based tool <http://www.inglescomercial.net> to assist Spanish professionals in their communication in international trade. She is presently preparing the publication of an English-Spanish dictionary specialized in Veterinary terminology.

Pascual Pérez-Paredes completed his PhD in Applied Linguistics in 1999 after a research stay at The Univesity of Texas at Austin. His main interests are quantitative research of register variation and the compilation and use of language corpora.. He has been project coordinator of a MINERVA initiative funded by the European Commission: SACODEYL <http://www.um.es/sacodeyl> and Corpora for Content & Language Integrated Learning <http://www.um.es/backbone> in Spain. He is also involved in corpus-based international projects such as LINDSEI (UCL) and ICCI (TUFS). Some of his most recent publications include *Developing annotation solutions for online Data Driven Learning*, ReCALL Journal, and *Tracking learners' actual uses of corpora: guided vs non-guided corpus consultation,* CALL Journal. During the 2009–2010 academic year, he completed a research stay at the English Department in Northern Arizona University. For more information visit <http://www.mendeley.com/profiles/pascual-perez-paredes/>

Purificación Sánchez Hernández is a Senior Lecturer at the University of Murcia. She received her PhD in English Philology by University of Murcia. She has been a visiting professor in Calgary University. She has published mostly on the didactics of English language, and English for Specific Purposes in international journal and teacher's magazines. Her main areas of writing and research include teaching English as a foreign language, using literature in language teaching and English for Specific Purposes. She is currently working on pedagogical applications of studies with corpora and has been involved in several International Research Projects.

Pilar Aguado joined the English Department at the University of Murcia in 1990. She took her PhD in 1997 (Shakespeare's Stage Directions: F1 and Editorial Intervention in the 18th Century) and got her permanent position as a Senior lecturer in 2004. Her main teaching and research interests are Teaching English as a Foreign Language, ICTs, Materials Design and ESP. She has been a Language Advisor for CAGE Panel, Cambridge University Press, has been involved in several research projects on Learner Corpora and Orality, and collaborates as a referee in some international journals in the field of English Studies.

Carmen Sancho is Senior Lecturer in the Department of Applied Linguistics at the Polytechnic University of Madrid, where she teaches EAP, Professional Communication and in-service training seminars for engineering teachers willing to undertake English-medium instruction within the EU framework. Her primary research focus lies in two major areas: the interdisciplinary analysis of written academic and professional discourses and innovation in the teaching of academic literacies. Recently her interests have extended to pedagogical projects dealing with e-portfolios and fostering creativity and autonomy among technical students.

John Flowerdew is a Professor in the Department of English, City University of Hong Kong. His research interests include discourse analysis and English for Specific Purposes. Among other books, he has published Academic Listening: Research Perspectives: CUP; Research Perspectives on English for Academic Purposes: CUP (with Matthew Peacock); Academic Discourse: Longman; Second Language Listening: Theory and Practice: CUP (with Lindsay Miller); and Lexical Cohesion and Corpus Linguistics: John Benjamins (with Michaela Mahlberg). His latest books are Advances in Discourse Studies: Routledge (with Vijay Bhatia and Rodney Jones) and Critical Discourse Analysis in Historiography (Palgrave Macmillan).

Shaeda Isani is full professor at Grenoble University (France). She teaches in the Applied Languages Department where she is Director of Studies for Masters Programmes. Her research in English for Legal Purposes focuses on two fields of enquiry, visual semiotics of law-related professions and specialized narrative in popular fiction as a vector of specialized/professional cultures. Widely published in these areas, she has been editor of several special editions. She directs the ESP branch of the research group ILCEA/Gremuts and is also a Board member of the French national research group in ESP studies, the GERAS, where she in charge of the Law SIG. <shaeda.isani@u-grenoble3.fr>.

Gillian Diane Lazar is a Senior Lecturer (Academic Writing and Language) in the Learner Development Unit (formerly English Language and Learning Support) at Middlesex University, U.K., where she also teaches on the MA TESOL and Applied Linguistics and supervises dissertations. She has published in international journals,

learner dictionaries and teachers' magazines. She is the author of *Literature and Language Teaching: A guide for teachers and trainers* (1993), *A Window on Literature* (1999) and *Meanings and Metaphors* (2003), all published by Cambridge University Press. Her research interests include using literature in language teaching, figurative language, academic literacies, materials development and online learning.

Kris Buyse is associate professor at the faculty of Arts and the Leuven Language Institute of the K.U.Leuven (ILT), as well as at the department of Applied Language Studies of the Lessius University College; he teaches Spanish language, linguistics and didactics at the Faculties of Arts, Economics and Medicine. He has published on the corpus based investigation of translation, lexicography, contrastive linguistics, LSP, CALL and Spanish as a foreign language. He also directed the UrgentiAS project.

An Laffut holds a PhD in English Linguistics. She is an EAP/ESP lecturer at K.U.Leuven (Leuven Language Institute, ILT), where she collaborated on English part of the UrgentiAS project. She is also involved in the Teacher Training Department (EFL section).

Eva Saver holds a masters degree in Romance Languages from K.U.Leuven. She collaborated on the Spanish and French part of the UrgentiAS project and she is a lecturer at Leuven Language Institute (K.U.Leuven, ILT), where she teaches medical Spanish. She is also a Spanish teacher in adult education at CLT Language Centre (Leuven).

Herlinda Vekemans is a lecturer at Leuven Language Institute (K.U.Leuven, ILT). She has been teaching medical English in the Faculty of Medicine for over 15 years and co-coordinated the UrgentiAS project. She also teaches legal English in the Law Faculty and academic writing and presentation skills to doctoral students in Biomedical Sciences. As part of the ILT team providing tailor-made language services to K.U.Leuven personnel and research staff, she translates and revises publications in the biomedical field.

Elena Montiel-Ponsoda is senior researcher at the Ontology Engineering Group of the Universidad Politécnica de Madrid. She received her M.A. in Conference Interpreting and Translation (2000) by Universidad de Alicante, her B.A. in Technical Interpreting (2003) by Hochschule Magdeburg-Stendal, Germany, and her PhD in Applied Linguistics (January 2011) by Universidad Politécnica de Madrid. Her research activities are at the intersection of Translation, Terminology, Natural Language Processing, and Ontologies. She has participated in several national and international projects in the fields of Semantic Web and Information Technologies.

Guadalupe Aguado is associate professor at the Facultad de Informática. She received her PhD in English Philology and her MSc in Translation by Universidad Complutense Her current research activities include, among others: Terminology and Translation in the field of Information Technology and Natural Language Processing, as well as the relationship between the ontological and the terminological fields in its application to the Semantic Web. She is a member of the Ontology Engineering Group at the Universidad Politécnica de Madrid and she regularly participates in different national and European projects on these topics.

Index