Research in English and Applied Linguistics
REAL Studies 8

Research in English & Applied Linguistics

Graduate Academic Writing in Europe in Comparison

edited by
Josef Schmied

Cuvillier Verlag Göttingen
Internationaler wissenschaftlicher Fachverlag

REAL Studies 8

Previous Titles

Vol. 1: English for Central Europe - Interdisciplinary Saxon-Czech Perspectives
Vol. 2: Stress Patterns of Nigerian English
Vol. 3: Complexity and Coherence: Approaches to Linguistic Research and Language Teaching
Vol. 4: English Projects in Teaching and Research in Central Europe
Vol. 5: Academic Writing in Europe: Empirical Perspectives
Vol. 6: Exploring Academic Writing in Cameroon English: A Corpus-Based Perspective
Vol. 7: English for Academic Purposes: Practical and Theoretical Approaches

Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives

---

Edited by

Josef Schmied

Bibliographical information held by the German National Library
The German National Library has listed this book in the Deutsche Nationalbibliografie (German national bibliography); detailed bibliographic information is available online at http://dnb.d-nb.de.
1st edition - Göttingen: Cuvillier, 2015

Nonnenstieg 8, 37075 Göttingen, Germany
Telephone: +49 (0)551-54724-0
Telefax: +49 (0)551-54724-21
www.cuvillier.de

1st edition, 2015
This publication is printed on acid-free paper.

ISBN 978-3-95404-959-2
eISBN 978-3-7369-4959-1

Preface

The present volume draws on the experience of teaching academic writing at Chemnitz University of Technology and the Workshops and Summer Schools developed together with partners from Macedonia, Montenegro, Albania and Kosovo that were organized with the support of the German Academic Exchange Service (DAAD) in South Eastern Europe in 2013 and 2014. The students' perspective is considered crucial, and half of the articles were contributed by MA and PhD students. The cooperation in this project was made enjoyable for everyone by the enthusiasm of the students and in fact it was a major aim of the project to integrate as many students as possible and relate the theoretical part as strongly as possible to the practical experience, so the students' perspective was considered as important as the teachers' perspective and large parts of the volume were discussed among MA and PhD students from the participating universities.

This volume introduces key principles of academic writing. It explains the data compilation and analysis, so that graduates can use their own database to decide which conventions to follow in their own writing. It gives examples from different genres (from term papers to PhD theses) and different university contexts (from Germany, Albania, Macedonia, Serbia, Poland, and even China in comparison).

The introductory contribution by Josef Schmied is a state-of-the-art summary for junior researchers in the field focusing on the concepts of metalanguage and genre, it explains the *ChemCorpus* as a reference corpus, shows exploration principles and illustrates empirical approaches through text elements and numerous examples.

The following four contributions focus on teaching practical aspects and the wider context of learning at (post-) graduate levels, i.e. in MA and PhD programmes: Helen Forbes makes a concrete and detailed proposal for vocabulary learning. Elena Ončevska Ager concentrates on the important and current topic of feedback. Irina Petrovska and Bisera Kostadinovska write on a special approach to the teaching of writing. Magdalena Rogozińska analyses the PhD programmes at Polish universities and discusses how students' writing skills could receive "adequate attention".

The following four comparative contributions are written by Chemnitz MA and PhD students on the basis of their theses and/or the *ChemCorpus*: Sven Albrecht tries to prove his hypothesis that connecting adverbials are more frequently or more elaborately used in untimed than in timed student writing in the *ChemCorpus* and comparable text collections. Dana Beyer follows the same students' usage of hedging in their BA and their MA theses. Jessica Küchler compares the culture-specific modal auxiliaries *may* and *will* in term papers, the first academic texts written by Chinese and German students, and proves that a relative small topic can provide empirical evidence for a vast discussion on general and culture-specific writing. Jacinta Edusei compares code glosses with exemplifying and reformulating functions in Albanian and German BA theses and demonstrates that a relative small, explorative case study may yield interesting results that indicate the need for a more detailed and complex study.

The following five contributions discuss language- and culture-specific writing problems again on an empirical basis of specific Serbian, Macedonian or Albanian corpora: Jasmina Djordjevic focusses on introductions and illustrates negative transfer from Serbian in 50 essays or term papers by third-year students in English, the lowest level of academic writing analysed in this volume. Zorica Trajkova compares commitment and detachment in eleven Macedonian BA theses to eleven German BA theses. Bledar Toska analyses different types of linkers in a big corpus (almost 10 Million words) of 160 Albanian PhD dissertations from different fields (Economics, Humanities, Natural and Social Sciences). Armela Panajoti adds to the well-known debate on authorial identity (i.e. in personal pronouns) in a sample of the same corpus (about 3 Million words) and highlights tentatively some differences between male and female students writing Albanian PhD theses. The final contribution by Silvana Neshkovska and Jovanka Lazarevska Stanchevska analyses a small sample (selected 50 pages from three Macedonian PhD theses) and discuss all errors in detail. All these case studies show that very different corpora and corpus samples can be used profitably for empirical analyses of very different features important in academic writing.

The individual contributions show, also in their styles, the different research and style traditions in European universities. The emphasis on practical approaches in teaching and research allows us to take theoretical backgrounds into consideration wherever needed and still make this project and this volume accessible to young researchers. Author-specific idiomaticity and some grammatical special features (like the well-known article usage by speakers of Slavonic languages) were left untouched, when the general understanding was not affected. Even the formatting of references was only harmonized when functional necessities appeared prominent (e.g. to distinguish independent publications like books from dependent publications like articles). It has to be emphasized at the very beginning of this volume that this is not intended as a model that has to be followed by graduates in various parts of Europe but that this is a documentation of the different conventions and practices of academic writing in different parts of Europe, which graduates in this field must be aware of, so that they can make their own decisions about which conventions to follow and which conventions to neglect in their own writings. Thus they may find a good compromise between the necessities of the field and their own identity as independent scholars.

December 2014 Josef Schmied

Contents

Graduate Academic Writing in Europe in Comparison: A Research-Based Approach to Metalanguage and Genre[1]

Josef Schmied (Technische Universität Chemnitz)

Abstract

This contribution introduces the key concepts of academic writing, metalanguage and genre. Metalanguage is seen to include all writer-reader interaction, esp. stance and engagement markers. The concept of genres in academic writing is discussed as a core-periphery model with the research article in the centre and the conference presentation, research monograph, handbook article and the chain from BA through MA to PhD thesis as other core genres. All concepts are explained and illustrated by examples from the *ChemCorpus*, which can serve as a (partial) reference corpus to all the other national mini-corpora in the SE European academic writing project and beyond. A research-based approach means that writers do not learn rules, but discover patterns and conventions themselves, either by testing ideas from textbooks or by exploring their own small corpora, even to test whether their linguistic variables are appropriate for their text/genre or socio-biographical variables. They can also use comparisons with similar corpora to position themselves in the spectrum between individual identity and disciplinary convention. Through this approach graduates gain skills that should be useful for their own writings at university and even for their professional life afterwards.

1. Introduction

This state-of-the-art introduction tries to show to junior researchers and practitioners in academic writing that this is an interesting and useful field: interesting because it is accessible to young researchers with only a few basic "theoretical" concepts and useful because students may profit from practical knowledge in the field for writing their theses during their studies or even for improving and editing other writers' work after their studies. Of course, no student of English or writing specialist in general would want to evaluate and discuss research findings in other disciplines, because they would not have the technical and subject-specific knowledge. But if we assume that research should be written in such a way that it can be understood not only by specialists, but by other academics as well; then graduate writers not only impress the direct supervisors, but even demonstrate that they can make a contribution in a sub-discipline or small research area to the academic discourse.

This "introduction" takes only some initial research experience for granted: term papers as first practices and (hopefully) a BA thesis as a first writing project.

---

[1] This contribution is based on the discussion with partners in the Chemnitz Academic Writing research group, and project partners in Macedonia, Montenegro, Kosovo, and Albania. Discussions with Dana Beyer were always helpful to clarify ambiguities or find even better illustrative examples. We all have to thank the German Academic Exchange Service and their committed advisors (not "administrators") for their continued support during the project including the finances for this final documentation.

It does not discuss learner English below C1 level (in the *Common European Framework of Reference for Languages)* or all problems of general writing, which may also play a part, but concentrates on the special features that make writing "academic". Our approach can be described as practical, research-based and empirical. It is "practical", since it encourages young researchers to use a similar hands-on approach and illustrates all concepts by concrete examples from the *ChemCorpus* (see below). It is "research-based", since it goes beyond practical language exercises, leading the young researcher to take a functional perspective, either by trying to prove the usefulness of conventions, guidelines, "rules", wherever possible or even by trying to find regular usages and choices in other writers academic texts. It is "empirical", since it encourages readers to test the assumptions and hypotheses they may have had themselves or they may have read in this introduction or other instructional academic texts.

Although young researchers may be painfully aware that academic writing has its discipline-specific vocabulary and idiomaticity, we take for granted that all academic writing today is expected to use an accessible metalanguage to guide the reader and to make the reading process as easy as possible. This is based on the conviction that good academic work does not have to "hide" behind complex vocabulary, discourse and style "to reflect the complexity of the subject" or "impress the reader".Well-written academic work includes the special attention to the reader all the time, even that the evaluation of the writing (in terms of a good mark, further funding or a positive citation) depends only partly on the "actual work", but also on the convincing, honest and open presentation in a conscious attempt to use metalinguistic choices for their writer–reader interaction that encourages academic discourse and the advancement of learning. Keeping a balance between individuality and convention is the central challenge in academic writing and every writer has to be aware of this in the writing process. Thus even writers who intend to "construct their academic self" towards the unconventional end of the continuum must be aware of the conventions, either by reading about them or by identifying them through the practical, empirical and research-based approach offered here. This encouragement of "rhetorical consciousness raising" (and related RCRI ="rhetorical consciousness raising instruction") has been discussed for a long time (since Swales 1990: 213).

This is also necessary since the conventions (as understood by the reader, evaluator, supervisor, etc.) and expectations seem to have become stricter over the last few decades. This can best be seen by the concept of genre, which is used today quite differently from 25 years ago and is very useful as it differentiates according to functional necessities in a modern globalised and digitalised academic world. Today, teachers as well as students can compare academic writing world-wide and see cultural (or language-specific) differences – and finally decide where to "position" themselves in the academic research space (cf. Hyland 2012). If young researchers aim at demonstrating familiarity, expertise and intelligence, they can do that on the basis of the conventions irrespective of how much their readers are aware of them.

2. Understanding key principles

Since academic writing specialists today cannot be specialists in all the possible subjects where English is used as an international lingua franca, they have to collaborate with the subject specialist and concentrate on the formal contexts and conventions that are important in all disciplines. Our focus in the following discussion will not be on specific conventions, but rather on the choices that young graduates have to be aware of and make according to the chosen academic identity. This applies in particular to non-native speakers who have not yet developed the necessary feel for the language that they use together with most of their colleagues as an international lingua franca (Mauranen 2013). But this raises the question of how we can develop such *Sprachgefühl* more consciously starting from the experience in finding conventions and pattern through empirical learning from authentic materials. This brings our approach in line with current philosophies of learning: Social constructivism suggests that knowledge and social reality are created through daily interactions between people and particularly through their discourse (cf. Hyland 2012).

2.1. Focus on metalanguage

The focus in this contribution will be on metadiscourse or metalanguage, i.e. the language that includes various conventions in author–reader interaction. The basic guideline or the basic principle for all our decisions should be reader-appropriate metalanguage, which depends on various factors, in particular the experience that readers have in the academic levels from BA to postdoctoral theses. The reader-specific perspective is particularly important in cases of hedging and boosting, which are two sides of the same coin of stance (cf. Trajkova 2015). The pragmatic meaning of both types of interaction is clear: In the case of hedging (*the data suggest*), the writer protects himself from the possible attack of other subject specialists by emphasizing the tentative position: writers invite constructive criticism and cooperation; whereas by boosting (*I am convinced*), academic writers indicate that they would be willing to “fight it out”, if other specialists were to attack them in those points.

Figure 1 (from Hyland 2005: 177) summarizes the whole spectrum of metadiscourse with two types of interaction, stance and engagement: Hedges and boosters can be seen as two ends of the stance continuum. Attitude markers can be expressed in verbs (*praise*), adverbs (*appropriately*) and adjectives (*best, excellent*) and allow the authors to align or distance themselves from previous authors. Self-mention through personal pronouns (*I/me/my* vs. *we/us/our*) is a feature of “soft” sciences (Hyland 2009: 76); the choice between plural and singular forms for the author is a disputed feature of discipline conventions.

Engagement can be seen in reader pronouns: *you/your* are used to bridge the communication gap between writer and reader. Directives are mainly expressed by imperatives (*take the following example*) and modals (*must, will*). Personal asides

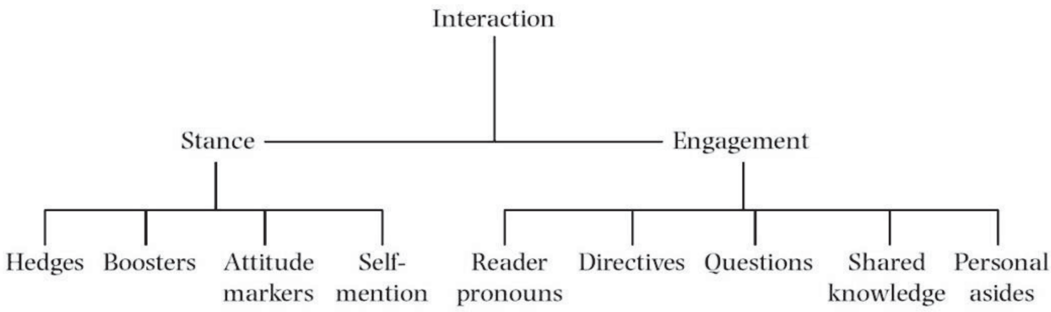


Figure 1: Types of interaction (Hyland 2005: 177)

are rare in most genres (*if you wish*). Appeals to shared knowledge create a community of discourse (*of course*, *as we know from* …). Rhetorical questions expect no answer but ask the readers to think and judge for themselves (*what can we do?*).

In this article, we cannot discuss discipline-specific content. We can only discuss metalanguage conventions, i.e. the discourses that have developed in specific sub-disciplines and departments and are applied to the genres that the students have to write first within their own departments and later on beyond (Schmied 2012 on an empirical study from a South African university). Theoretically, we assume that students move from a more individual approach to a department-specific approach and later to a subdiscipline-specific approach, so that their work can be compared world-wide, so that they can compete with other researchers in publications in appropriate journals.

2.2. Focus on genre

The concept of genre has made a surprising career in applied linguistics over the past 25 years. Since Swales' 1990 definition, it has replaced earlier concepts (like styles, register or text-type) and has been widely adopted and extended. Swales (2004: 4) sees his own studies in a "continuing phenomenon […] an increasing generification of administrative and academic life". He himself has added hierarchies (ibid: 12) and genre chains, sets and networks to the system (ibid: 18-25). Bhatia (2004: 23) proposed a widely accepted detailed concept:

(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 structured and conventionalised constructs, with constraints on allowable contributions not only in terms of the intentions one would like to give expression to and the shape 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 greater 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 conventionalised 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 communicative purposes’.

(5) Genres are reflections of disciplinary and organizational cultures, and in that sense, they focus on social actions embedded within disciplinary, professional and other institutional practices.

(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.

An alternative or complementary perspective is rare: Hyland (2009) distinguishes between research discourses (with the focus on research articles and conference presentations), instructional discourses (with a focus on university lectures, seminars, undergraduate textbooks), student discourses (with a distinction between undergraduate and postgraduate genres) and popular discourses (with a focus on science journalism). In a more complex set-up, this list of genres could be viewed in a central-peripheral frame and presented as in Figure 3 below.

2.3. Argumentative structures in genres

In research discourses, the trend towards a uniform argumentative structure seems most obvious: IMRAD/IMRD is an acronym for Introduction, Methods, Results, And Discussion. Originally, research articles in natural sciences were typically structured in this basic order. The IMRAD format has been adopted by a steadily increasing number of academic journals, since the first half of the 20th century. This structure has come to dominate academic writing in the sciences, although it is not suitable for many genres of academic writing, like case reports and meta-analyses, which usually have non-IMRAD structures. Recently, the structure has been expanded into AIMRAD to include a structured summary or abstract (and even, I may add, a list of 4-6 keywords for the index or for specialised searches). For applied linguistics, it may be appropriate to adapt the acronym to AIMAC, which concentrates on the analysis, including the interpretation and possibly even the summary and a complex conclusion (see Table 1 below).

For the beginner’s analysis of such sections, it may be useful to add a substructure wherever possible and some typical lexemes or phrases that can serve as indicators in academic text analyses.

The CARS (=Creating A Research Space) model (Swales 1990) has been an important breakthrough in the field and has stimulated a number of further studies. Swales’ CARS model for introductions is based on his study of articles across a range of disciplines. He identified the following moves as common among most articles (Figure 2):

Move 1: Establishing a territory
- Step 1 Claiming importance and/or
- Step 2 Making topic generalizations and/or
- Step 3 Reviewing items of previous research

Move 2: Establishing a niche
- Step 1a Counter-claiming or
- Step 1b Indicating a gap or
- Step 1c Question-raising or
- Step 1d Continuing a tradition

Move 3: Occupying the niche
Step 1a Outlining purposes or
Step 1b Announcing present research
Step 2 Announcing principle findings
Step 3 Indicating article structure

Figure 2: CARS model (Swales 1990: 141)

A similar structure has been found in book reviews (Motta-Roth 1998: 35) and review articles (Diani 2004: 109). But it may be difficult to present a detailed series of moves and steps as a model for a complete thesis (and maybe it is more appropriate for MA than BA levels). Even Swales & Feak (2012) only mention moves explicitly in the introduction and the methodology sections.

Table 1 attempts to combine the IMRAD and the CARS concepts and adds a few key lexemes or phrases as examples that may indicate the structure or sub-structure to the reader or that may be used consciously by the writer to meet the expectations of the examiners. As in the CARS model, not all steps are necessary and some are alternatives. What is not indicated in this table is the important network system that connects, for instance, the relevance in the issue section and the contextualisation or outlook in the conclusion, the tests indicated in the methodology section and the significance in the analysis, etc. Some of these networks seem elementary, but it is surprising how often thesis writers fail to check whether all research questions raised in the methodology section are answered explicitly or adequately in the summary or whether all hypotheses are rejected or confirmed.

structure	**substructure (tactical alternative)**	**key lexemes/phrases (as indicators)**
A = abstract	keywords in context	*focus, discuss, approach*
I = issue	new	*not enough research yet*
	relevant	*important, academic discourse, practical application*
	focussed	*concentrate, emphasise, purpose*
M = methodology	previous research, i.e. lit. review incl. evaluation	*concept developed, review, refer to, proceed to, claim*
	hypotheses possible?	*research question*
	data base	*corpus, data collection*
	tests/procedure	*calculate*
A = analysis	examples as evidence	*illustrate, show, prove*
	statistical tables as summaries	*table, figure, diagram, graph, bar*
	significance to generalise	*significant, chi*^{2}
C = conclusion	summary	*in conclusion, finally/at last, we have shown, discussed above*
	interpretation	*this proves that*
	contextualisation	*in a wider perspective, apply*
	limitations	*more data, beyond the scope*
	outlook	*further research is necessary, predict, dissemination/application of results*

Table 1: Revised IMRAD structure: AIMAC

The lexemes and phrases are just examples of what experienced and quick readers (who just scan the text in their routine when they have many other texts to read) expect to find in such sections and hopefully tick off as “familiar” in their mental genre schemata and, as an examiner, even in the printed thesis submitted. Such coherent academic lexemes and phrases help to create a global framework in top-down planning, which is a major headache for young academic writers (cf. p. 17 below). This is useful because for some writers at least, getting this overarching span from the first to the last page is the real difference between writing a short term paper (10 pages/5,000 words) and a BA thesis. Of course, this global AIMAC metalanguage form is just a scaffold that has to be filled with concrete academic content, but it is useful because it helps the writer to concentrate on the local details and thus to create a positive “familiar good standard” impression on the reader/examiner, who notices that the writer is a member of his/her academic discourse community.

There are more theoretical approaches to the argumentative analysis in discourse, but their application to academic genres has hardly begun, although the choice between more empirical and more theoretical, more critical and more epistemological argumentations promises to reveal interesting differences between national academic traditions in Europe and beyond (e.g. Smirnova 2013). Siepman et al. (2011) use the concept of rhetorical moves and give some intriguing examples without developing it into a comprehensive organisation concept.

2.4. Metalanguage features in academic writing

The list of prominent discourse features discussed in textbooks of academic writing is extensive. Swales & Feak (2012) discuss features related to their CARS and genre model (e.g. abstracts, reviews, commentaries, critiques, data commentary, introduction/methodology/discussion/result sections; we can see examples in 5.1. below), but also textual features like claims and hedging/boosting, linking words/phrases and connectors (this is why we will discuss them in 5.2 below). Siepman et al.’s (2011) Guide focuses on construction/paragraph organisation, focus/topic/message/climax sentence, but also on well-known functional concepts like elaboration, exemplification, given/new information or information packaging. Both handbooks emphasise less the qualitative (corpuslinguistic) perspectives on interaction markers (see p. 4 above and 5.3. below) or lexical bundles. But lexical bundles (*it can/could be argued that*) have been analysed in detail elsewhere (e.g. Biber/Barbieri 2007 or Dontcheva-Navratilova 2012), especially to compare native vs. non-native usage. Chen/Baker (2010), for instance, claim that non-native writers prefer VP-based bundles, whereas native writers prefer NP-based bundles. They compare the 1990 FLOB-J (the academic part of the Freiburg-Lancaster-Oslo/Bergen corpus) with BAWE (British Academic Written English) and a (Chinese) non-native corpus. The most frequent bundles are surprisingly different for the two British corpora and many bundles are not really academic (cf. the criticism in Simpson-Vlach/Ellis 2011 or Salazar 2014: 149): Although *on the other hand* is an expected non-native usage (functioning like *by contrast* and not as a counterpart to *on the one hand*), it also occurs almost half as often in FLOB-J (34 vs. 19, but 0 in BAWE). Similarly, *at the same time* is used very often by non-native and less than half as often by the native writers (24, 5/10 times) and as well as 16 vs. 10 times in BAWE (0 in FLOB-J). Real academic phrases like *in the context of* occur only 5 times in the Chinese corpus and 19 times in FLOB-J, but not at all in BAWE – but does that mean anything? Only *as a result of* is used equally well in all three databases (12, 17 and 9 times). For the analysis of larger bundles, we need very large corpora if we do not want to restrict ourselves to the most frequent constructions only like *can be used/explained/seen/found*, etc. – all clearly hedges. At least, such studies remind us that we need to discuss good academic English on the basis of a good general idiomatic English – and maybe our *ChemCorpus* family is not big enough for this.

Finally, it has to be mentioned that even punctuation is an important element in academic writing and author–reader interaction, because good pronunciation structures the text for the reader and thus makes it easier to process. This is reflected in different guides as Greenbaum's (1996) *Oxford English Grammar* and Swales & Feak's *Handbook*, which includes an interesting punctuation "decision" figure (2012: 39), in particular to choose between periods, commas and semicolons. Siepman et al.'s *Guide* (2011) devotes a whole chapter 4 (p. 192-224) to punctuation. Schmied (2013: 32ff) shows how punctuation has be included in good writing training, because it has functional values (i.e. to make life easier for the reader) that go far beyond the narrow-minded application of old style rules.

3. Exploring genres in academic writing today

Academic writing skills were explicitly included in many new curricula at European universities (Schmied 2011) and many take a genre approach. Students at Master's level have written term papers and their Bachelor theses. Thus they are aware of the concept of genre, but they may not see it as directly related to their own work. They have to develop a feel for the usefulness of the concept of genre and the metalanguage genre conventions of their discourse community to move successfully from student to research discourse during their MA studies. Our survey includes two approaches that complement each other:

Table 2 shows a matrix approach to genres in academic writing, trying to complement the most central variables of research/teaching, spoken/written and specialised/learner (and a few others) by concrete figures on writer's experience and length, and trying to combine them as mega-types in special contexts. The comments column serve as a basis for further discussion. The approach does not produce a neat mosaic and not all possible genres have been included, like a roundtable (talk) with statement, comments and question/answer session during conferences.

Figure 3 (below) offers a gradient approach from more central to peripheral genres and even a few examples of academic related genres from journalism, politics, marketing and technology. But even some peripheral genres (like press releases or obituaries) may be debatable. Since it is impossible to put all genres in a two-dimensional space, the dimensions instruction-, research- and job-orientation have been added as further dimensions.

CONTEXT	GENRE	type	medium	audience/ readership	expertise in years	length	purpose	***comments***
article	**research** (scholarly) **article**	r	w	specialist	5	5,000+	drive research	in double-blind peer-reviewed journal with impact factor
book	research book	r	w	specialist	6	10.000	drive research	also monograph, trend towards IMRAD
	textbook	r	w	novice	8	6-10,000	introduce novices	also course book, intoduction?
	handbook	r	w	specialist	10+	10,000+	state-of-the-art	honour+reliable?
	book reviews	r	w	specialist	4	1.000	evaluate/critique	also section in book/thesis
	state-of-the-art review	r	w	specialist	10+	2.000	start project?	research field survey
	article collection (ed)	r	w	specialist	10	150+p	state-of-the-art?	multiple authors,1 topic/area;conference?
project doc.	project proposals	r	w	specialist?		5.000	accepted for qualif./funding	a promise with milestones to check progress
	BA/MA project proposal	t	w	learner	3	1.000	qualification	trend to IMRAD, research questions/hypotheses
	PhD project application	r	w	learner	5	5.000	qualification/funding	trend to IMRAD, research questions/hypotheses
	(thesis/PhD) defense	r	s	learner	3-5y	20+20-40m	qualification	trend to IMRAD, research questions/hypotheses; open discussion
	project/progress report	r	w	specialist		6-10,000	justify expense	multiple authors,1 topic/area
conference	conference presentations	r	s	specialist	5	15+5m	drive research?	protect area? ask advice?
	key-note (lecture)	r	s	specialist	10+	45-60+15-30m	state-of the-art?	famous scholar,personality
	plenary (lecture)	r	s	specialist	10+	30-60m+15	research overview	broad topic?
	progress report	r	s	specialist	5	10m	demonstrate research? ask advice?	
	(conference) abstract, proposal	r	s	specialist	5	300-600w	acceptance	
	poster introduction	r	s	specialist	5	3m	view poster	very consice+attractive?
	conference poster	r	w	specialist	5	1 A0	introduce research,popularise?	Rtrend to IMRAD; fig/tables+ref.s
	conference proceedings (ed)	r	w	specialist	10+	150+p	document research	multiple authors,1 topic/area
	conference report	r?	w	public	3	1.000	drive research?	popular? dissemination
university	lecture	t	s	learner	10+	45-90m	disseminate knowledge	competes with textbook? problem attention span
teaching	student/seminar presentation	t	s	learner	1	10-20m	qualification	problem discussion? self-protective?
	Wiki	t/r?	w	learner	1		collaborate in Knowledge creation	groupwork/multiple authors
	classroom discussion	t		learner	1		collaborate in knowledge creation	problem interaction
	field notes	t	w	self?	1		collect information	data collection methodologies?
	BA thesis	t/r?	w	specialist	3	40+p	qualification	
	MA thesis	r	w	specialist	5	60+p	qualification	
	PhD thesis	r	w	specialist	8	200+p	qualification	conventions,not a research book
	habilitation/postdoctoral thesis	r	w	specialist	10+	200+p	qualification	tend to small; replaced by articles in p-r j.?
subsidiary?	(article) abstract	r	w	specialist	5	1-300w	read? full article	part (free; to decide worth paying?)
	handout	t/r?	w	l/sp	1	1-2p	support,take-home	large diagrams,figures,statistics; examples;references
"valorisation"	university journal,newsletter	r	w	public	5	1-2p	demonstrate "value"	untrained?
	popular blog	r	w	public?	1	1+1+1	time-line of development (projects,carrer)	technical platform for old genres?
	popular science book	t/r?	w	public	8	80-200?	create interest in research?	

Table 2: Matrix of genres and wider contexts in academic writing

For a classification of current academic genres, we have to start with the research article, which is the prototypical type of research texts and well researched since Swales' (1990) analysis. It is written by the specialist for the specialist and serves the purpose of driving research further in the field; it usually requires considerable experience in academic writing, which may start from MA level, but most journals would probably only accept articles written by experienced researchers that have been in the business for considerable time. However, the names of researchers do not play a role in the decision and evaluation process because the standard procedure in renowned journals with a high impact factor is a double-blind peer-review process, so that neither the reviewer knows the author nor the other way round. This does not mean that writer and reviewer are "strangers". A good editor knows which reviewer would be interested in doing the job. To make sure that the review is done quickly and reliably, editors usually choose reviewers that belong to the same discourse (sub-)community, so that the review is not only a burden that the reviewers are usually not paid for, but also interesting and useful for the reviewers' own academic work. Additionally, writing the review should make them feel honoured and well-informed about current developments in their academic field. Of course, reviewers do not only expect to find general academic words and phrases, but also discipline-specific terminology. In case the reviewer has a somewhat different specialisation, a "standard", "familiar" general academic metalanguage may help even more.

Traditionally, the best-known academic genre family are books: The standard qualification at book length is usually the research book, which is also written by the specialist for the specialist and usually after a PhD/Master's thesis. It is considerably longer than the research article, and this is the monograph that also drives research, although it is usually more difficult to publish, and the publication takes more time than smaller research texts, like the scholarly article. It tends to follow, in particular in the natural sciences, an IMRAD structure (see above). In contrast to the research book, a text book is a book for students, a novice readership. Though it is often shorter than a research book, it also requires special expertise: Its purpose is to introduce novices to the area and to make them enthusiastic, so that they go on with their reading. A similar concept is pursued by a course book or even an introductory book, which are also geared towards lower levels at university teaching and sometimes even follow the structure of a university course of 12 to 14 weeks. They all have to be sufficiently different from a lecture series or a conference "paper" selection, which are rarely published. A handbook, by contrast, is written by even more experienced specialists. It is usually quite thick and detailed with many state-of-the-art articles, so that the specialist perspective at a certain point in time is shared with other academics; again it is an honour to be asked to contribute to a handbook, however that implies that editors and publishers expect a reliable contribution, which is the reason why students interested in concept formation may take a conventiently short contribution in handbooks as a reliable starting point for further research. Printed texts that derive from spoken genres like key-note lectures and conference collections do not serve this purpose so clearly.

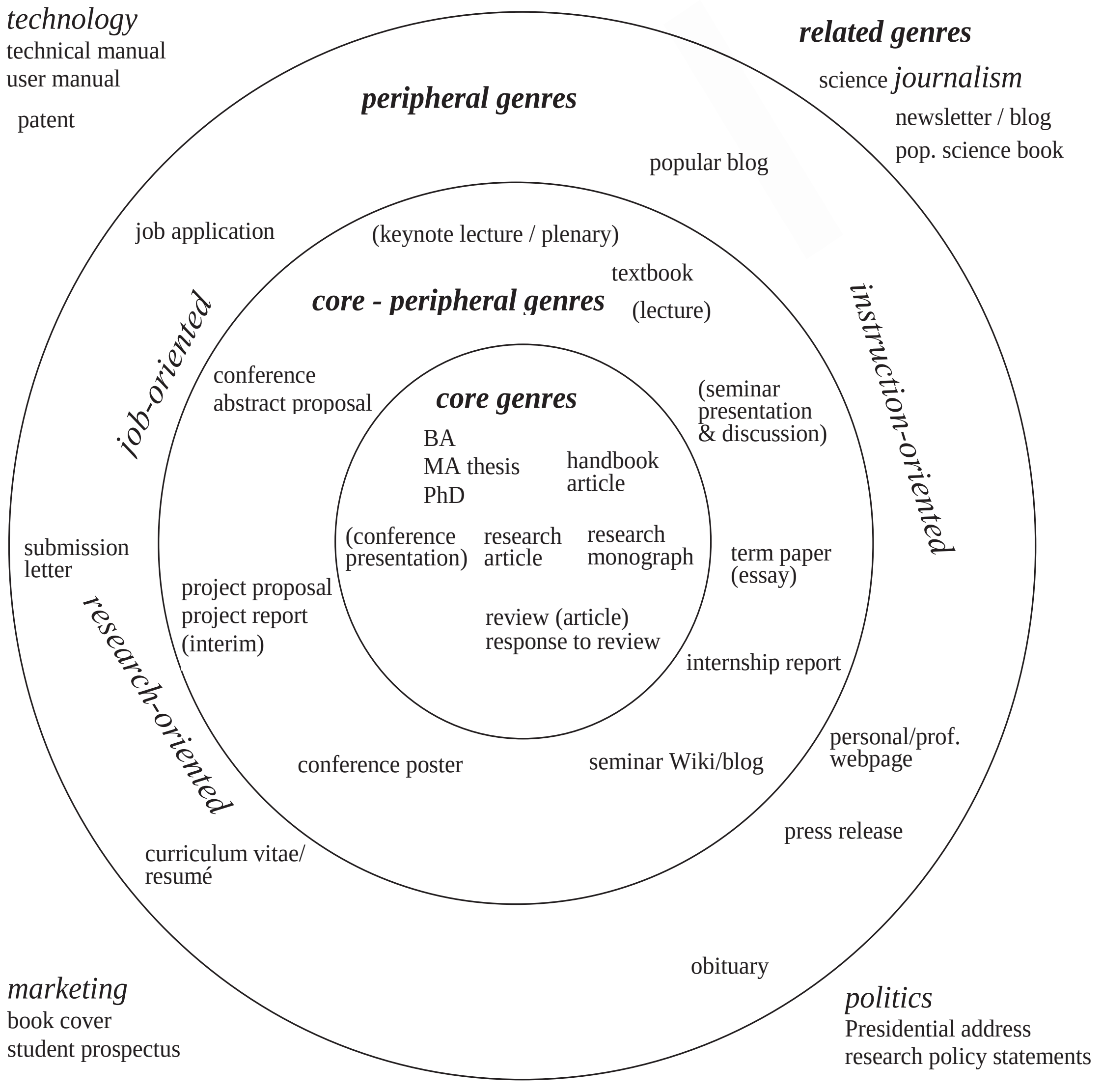


Figure 3: A prototype approach to academic writing

A further genre family are project proposals. They are usually teaching-oriented and may have a length of around 5,000 words, in which the researchers promise to reach certain milestones, so that in the end a project can be completed profitably; these milestones allow not only the researchers themselves, but also the funding agency or supervisors to check whether the progress is adequate to achieve a specific goal in a certain time. At university level, the first academic writing project is usually a BA thesis, which may still be considered a teaching genre, although it is supposed to introduce the students to independent research. Such a project is usually written at the end of a three or five-year period of learning at university and it often has a length of 15-20,000 words. The purpose is, of course, quite elementary, to be awarded a certain degree in the university career. A PhD project application by contrast is really research-oriented and it usually takes some five years of experience to write a proposal that may be around 5,000 words, so that a funding agency is tempted to spend money on the applicant or a professor is willing to supervise the project. An oral genre that is related to a project application can be a thesis or PhD defence which is usually done after three or five years at university level; the standard format in many universities is a 20 minute presentation and a 20 to 30-minute discussion based on this presentation or the PhD thesis as such. Finally in this category of projects, the project or progress report has to be mentioned, that is usually about 6-10,000 words long. It is often written by multiple authors because of the funding needs and research experience.

An additional family of genres are conference genres: The prototypical conference presentation is spoken, rarely spoken to be written, and it requires a few years of experience although it is only 15 to 20 minutes long plus a 5 to 10-minute discussion afterwards. A more advanced conference contribution is a keynote lecture which requires usually some longer research experience (of more than 10 years) and again, it is a contribution of 45 to 60 minutes first and a discussion of 15 to 30 minutes afterwards. Usually the presenter is a senior scholar, famous for his/her books or personality. Related to the keynote lecture is the plenary lecture, which is often quite similar, sometimes a little shorter, but it also requires some similar experience. The conference abstract or conference proposal has to be written and submitted to the academic committee organizing the conference and 300 to 600 words are usually enough to decide whether the contribution can be accepted or not for the conference. New academic genres at conferences are poster introductions and conference posters. Poster introductions are usually very short and the related conference poster is usually only one A0 sheet, often including several tables and figures. Due to the fact that it is supposed to relate a usually young researcher to more experienced researchers or peers, it also contains an address and email that can be used for suggestions and general academic discussion afterwards. A conference is often made memorable by the conference proceedings, which are an edited collection of articles that are based on a conference presentation. They are usually not too long, especially when they are submitted for electronic publication on a CD (shortly after or even before the conference). Such a volume often includes multiple authors, but should concentrate on merely one topic and area. A conference report is then a more popular research genre which basically serves to disseminate the information on the topic of the conference. It can often be written by relatively young scholars and it is also rarely longer than two pages.

Special cases are genres that relate the academic work to the general public. This type of genre family I have called *valorisation* (in Table 2), because the French term shows more clearly that the purpose is to disseminate information on academic activities and convince a general public that academics pursue useful projects. Traditionally, such genres are considered peripheral, although they may include advanced technology. Thus popular blogs can easily include pictures or sounds and popular science books can also be developed further in this direction.

The central steps of academic qualification genres lead from the first academic attempt in a BA thesis (pursued after three years of study and often little more than 40 pages) to an MA thesis (after five years of study and over 60 pages), but the expectations differ. The contrast to a PhD thesis is striking in quality and in quantity, since that can only be pursued after around eight years of study altogether and may take up at least 200 pages today. Quite similar is the next career step. If it exists at all, a postdoctoral thesis or *Habilitation* at traditional universities requires more than 10 years of experience, although the length does not have to be more than the PhD thesis. However, all these theses serve the writers' further personal qualification and are integrated into a relatively rigid and well-documented sequence of procedures.

Finally, it is worth mentioning some genre-related issues: ambiguous and supplementary genres. Thus abstracts serve different purposes: conference abstracts can be only promising proposals, the project just needs to be finished before the conference; article abstracts have to be finished and the complete version of the article has to be available immediately after the abstract is read. Oral contributions are traditionally accompanied by handouts, which serve not only as a guideline for the reader in the free speech that is expected in academic speaking today; they may also include figures and tables that do not fit well into a presentation and that the listeners may want to take home for further study. A handout is usually a genre that focusses on research, though it can also be used for teaching purposes at university level.

4. Compiling corpora to analyse conventions

Although students may be well aware of the big reference corpora from the British National Corpus to the Brigham Young corpora (Davies 204-), it is also important that students get used to analysing much smaller and more specific corpora or extract subcorpora that are more comparable with their own materials for analysis. It is extremely important for young researchers to recognize the conventions of their own subdisciplines and compare them to the much broader options that departments and individual researchers are able to choose from. A corpus approach is appropriate not only for non-native speakers of English, who are not sure whether they can rely on their intuitions, but also for novices to base intuitive judgments on real-data evidence.

Finally, many students are inclined to use "Web as corpus" techniques to find out about lexical patterns (e.g. look for Google responses to queries of *different from* vs. *different than,* cf. Schmied 2015).

4.1. Identifying textual, social and linguistic variables

When we compile a new corpus, we usually start with the biggest context variables: these include text variables like written vs. spoken, genre, etc. and social variables like native vs. non-native, gender, age/educational level, etc. The importance of these variables depends on whether we can set up simple logical research hypotheses (e.g. female writers are more "careful", "tentative" and thus prefer more or different features).

For the linguistic features, it would be most useful if we could refer to standard textbooks (or grammars and dictionaries) or model articles to find a comprehensive list of features, preferably in different functional (sub-) categories. Unfortunately, many young researchers notice that usually there is no uniform opinion on these categorisations in their secondary literature or references and the delimitation of their linguistic concepts and their "ingredients" is their first major task.

An issue closely related to this is the question of how big a research corpus has to be. This of course depends on the frequency of the linguistic variable analysed and the number of textual and social variables included in the analysis.

For a sound statistical analysis not only an absolute number of cases is important, but also their relative distribution in the cells for each variable combination. Of course, such variables can be taken out and the results collapsed later-on during the analysis, but initially at least it would be most useful to have enough cases to exclude variables empirically as unimportant or insignificant at the end and not impressionistically at the beginning.

4.2. Comparing corpora in Academic Writing

The *ChemCorpus* can be compared with other available current academic corpora of student or academic writing, less to the *ICLE* Corpora, which have mostly been compiled in the 1990s of argumentative student essays which are a different genre and not really part of academic writing today, more with the *BAWE* corpus, in particular if we take the English or the linguistics parts and students that have German as a first language only. It can be compared quite well with the *Brno Corpus*, which was developed in a parallel project over the last five years and includes a clear division between linguistics, literature and methodology specialisations. These non-native academic corpora can also be compared to the academic or academic humanities parts of the *British National Corpus* (*BNC*; from the early 1990s) or the *Corpus of Contemporary American English* (*COCA*; or even the *TIME magazine*), bearing in mind that those are of course much more heterogeneous as far as the production date and the socio-geographical data of writers are concerned (Davies 2004-). For explorative purposes, the academic corpora mentioned above can of course be used when the frequency of linguistic phenomena analysed is normalized, per 1 million words, for instance.

4.3. Compiling stratified and *ad-hoc* or test corpora

Corpus compilation methodology and corpus design are discussed in all standard textbooks (e.g. Biber/Conrad/Reppen 1998: 246-250). Usually, "representivity" is considered an unrealistic goal and "diversity" is aimed at through stratified sampling. This means when many texts by female students are compiled already we look specifically for texts by male students until we have half or a representative or a satisfactory number. What satisfactory means in this context depends on the importance of the variable in our research questions or hypotheses. Only a well-designed corpus will bring us reliable data for an analysis of language variation. An old issue in academic writing corpora is the size of the texts: generally, complete texts are preferable, because the argumentative, lexical or cohesion patterns may be quite different in the sections Introduction, Methods, Analysis, Results or Conclusion discussed in 1.3 above. The size of the corpus necessarily depends on the frequency of the variables investigated: pronunciation variables need only small, idiomatic or collocation variation large samples.

Sometimes it seems reasonable to compile a quick and dirty ad-hoc or test corpus first, just to see whether the (limited, insufficient, and skewed) data promise results. This is comparable to the tradition of disposable corpora from translation studies (Gavioli 2006).

4.4. The *ChemCorpus* as a reference corpus

The following matrix (Table 3) shows that the *ChemCorpus* consists of data compiled from two types of degree programs, the traditional German Magister as well as the modern BA/MA from the Bologna System. The earliest pieces of academic writing in the Bologna System are the so-called term papers, which are specific exercises where students can demonstrate their critical thinking in linguistics, literary criticism or cultural studies and get used to the formal conventions, i.e. MLA versus APA style. The structure in these papers is usually not very strict, yet all and in many cases relatively narrative, based on the available secondary literature with (hopefully) some conclusions of their own. Only little afterwards some students have to write their first report, which are partly related to their ERASMUS studies abroad, partly to other projects where this form is appropriate. Essentially, it is the review of the plans they had before the projects were started and compared to the results which may have digressed more or less from the original plan. The first piece of academic writing students have to submit to qualify for further studies is usually a BA thesis which is in many ways a requirement if students want to change universities to continue in a Master's programme. The BA thesis and the MA thesis are usually quite similar and differ in length as well as in academic rigor. MA theses tend much more towards a conventional structure (like IMRAD style), but deviations are acceptable and depend very much on the subject chosen.

The length and number of texts per section in Table 3 are just rough estimates to make the corpus coherent and logical in proportion. The average is also variable, although the requirements are standardizing and the figures given are standard in many parts of the world, but the actual original texts submitted (and also stored in their original form) have more words, since in the corpus-processing stage the number of words is reduced. This is necessary to prepare the input suitable for analysis using the standard tools (like *AntConc* in the following examples). Thus bibliographies/reference lists, appendices and even figures and tables in the text are taken out to ensure that only words in context were really included in the quantitative analyses. The figures can, of course, be adapted to local departmental standards, but all this has to be documented in the corpus manual.

genre	specialisation	number of texts	average length	total words
timed Mag paper	language/linguistics	70	4,200	0.3 Mill.
Magister thesis	language/linguistics	25	25,000	0.6 Mill.
	culture/literature	11	30,000	0.4 Mill.
Total Magister		106		1.3 Mill.
term paper BA	language/linguistics	100	4,200	0.5 Mill.
	culture/literature	100	4,700	0.5 Mill.
project report	(cultural)	120	4,000	0.5 Mill.
BA thesis	language/linguistics	80	12,000	1 Mill.
	culture/literature	80	16,000	1 Mill.
term paper MA	language/linguistics	80	5,700	0.5 Mill.
	culture/literature	80	6,600	0.5 Mill.
MA thesis	language/linguistics	40	25,000	1 Mill.
	culture/literature	40	25,000	1 Mill.
Total B/MA		720		6.5 Mill.

Table 3: *ChemCorpus* set-up by genre and specialisation (trad. Magister and B/MA)

The *ChemCorpus* can be called a monitor corpus, because it is so big that only sections of it can be used for comparative purposes. In the following comparisons (this volume), for instance, Albrecht compares adverbials in timed and untimed student writing (i.e. 4 hour test papers and the traditional theses) from the traditional Magister part of the *ChemCorpus* with the comparative native speaker corpora *BAWE* and *MICUSP*, Beyer compares hedging in the BA and MA theses by the same students included in the *ChemCorpus*, Küchler compares the term papers in the *ChemCorpus* with a compatible corpus of Chinese student writings and Edusei compares the German BA theses with Albanian theses in English. In similar ways, research questions can be derived from the setup, either as culture-specific features or as first language induced interference phenomena.

4.5. Non-native reference corpora?

The difference between native and non-native writing has been analysed in many research projects. The corpora discussed in this section have been used to compare concessives and contrastives (Wagner 2011), hedges (Beyer 2009) and sentence adverbials (Vogel 2013). In many cases, the result is that non-native writers use more of these academic discourse features than native writers. The logical conclusion is that non-native writers have been influenced by the explicit teaching at higher levels of university studies.

This raises the issue whether native writers always have an advantage compared to non-native writers or whether non-natives texts can still serve as models or have even features that are different from native writing and can still be seen as models for successful academic writers. Explicitness has often been quoted as a feature of academic writing as well as translators or advanced language users, but this has also been argued about, especially the loss of explicitness in nominalisation and passivation (Biber & Gray 2010) – and this may have contributed to the trend away from nominalisation and passivation over the last few years in many writing cultures. In more general terms, there is no reason to believe that non-native writers have to reduce their use of academic discourse features that they may have acquired through university training to adapt to the level of native writers – if we accept the concept of *native academic writers* at all (cf. 5.1 below).

5. Exploring metalanguage in the *ChemCorpus*

The research-based, empirical approach propagated in this article uses corpus-linguistic methodologies, although these have not been used systematically in the development of the field of academic writing. Many researchers have created ad-hoc corpora for their specific research projects, but a generally acknowledged corpus model has not been widely accessible until recently. Swales and his associates have used corpora as a basis for their studies, but in their textbooks only the latest editions (e.g. Swales & Feak 2012) have made the direct link when the *Michigan Corpus of Upper-level Student Papers* (*MICUSP*, a collection of 830 papers from 16 disciplines) became freely available online in 2009 (http://elicorpora.info/). Many researchers have used *Google Scholar* instead for key-word-in-context analyses of lexical and idiomatic usage.

If we look for linguistic elements that can be used analysed, we find a very wide range, which will be exemplified in the following examples from the *ChemCorpus* (see 4.4 above). In this section, we only take a qualitative view at our data, partly because the quantitative view can be seen in the following papers that also use (sections of) the corpus.

5.1. Argumentative structure

A clear connection of ideas has always been considered important to help readers follow an academic text, not only specialised readers but also general academic readers, not only native speakers but also non-native speakers – but there are no native speakers and certainly no native writers of academic English – everybody has to learn (esp. about global argumentation structures) through guides and guidebooks or through independent explorative learning.

Demonstrating a good flow of ideas has always been difficult for teachers, and evaluating it has been extremely difficult for young graduate students. IMRAD/AIMAC is always a safe macro-structure for a thesis. As I have argued more generally above (p. 7), the proposed IMRAD/AIMAC organisation and corresponding metalanguage form is a good scaffold for students that has to be filled with concrete academic content. It is useful because a table of contents like in Figure 4 (with Introduction, Literature Review, Methodology, Results and Discussion) suggests to the reader/examiner that this may be a novice writer, but in principle this student is a member of "our" academic discourse community.

Table of Contents

1 Introduction **7**

2 Literature Review **10**

2.1 Canada's settlement history 10
2.2 English in Canada 12
2.2.1 Standard Canadian English 12
2.2.2 Newfoundland English 13
2.3 Selected Phonetic Features of English in Canada 14
2.3.1 Canadian Raising 15
2.3.1.1 Mainland Canada *15*
2.3.1.2 Newfoundland *18*
2.3.1.3 Quantitative, phonetic definition *19*

3 Methodology **22**

3.1 Participants 22
3.2 Data Collection 24
3.3 Data Analysis 25
3.3.1 Selected Tokens 25
3.3.1.1 Canadian Raising *25*
3.3.1.2 The Canadian Shift *25*
3.3.2 Measurements 26

4 Results **28**

4.1 Sample 28
4.2 Canadian Raising 29
4.3 The Canadian Shift 32

5 Discussion **39**

5.1 Interpretation 40
5.2 Limitations 42
5.3 Conclusion and Outlook 44

References **46**

Appendix **49**

Figure 4: Table of contents of a *ChemCorpus* BA thesis indicating an IMRAD structure

Contents

	Abstract		1
1.	Introduction		2
2.	The Term Migration and Polish Migration Motivations		4
	2.1.	Types of Migration	5
	2.2.	Ravenstein's Migration Theory	8
	2.3.	Lee's Migration Theory: Push and Pull Factors	10
3.	Polish Migration Behavior and Motivations		13
	3.1.	Gender-Relation	15
	3.2.	Age-Relation	16
	3.3.	Length of Stay	18
	3.4.	Labour Market	21
	3.5.	Social Networks	24
	3.5.1.	Polish Diaspora	28
4.	Polish Migration in Historical Perspective		30
	4.1.	Polish Migration to the UK before 2004	30
	4.2.	Polish Migration to the UK after 2004	32
5.	Conclusion		36
	Bibliography		40
	Eigenständigkeitserklärung		43

Figure 5: Table of contents of a *ChemCorpus* BA thesis indicating a non-IMRAD structure

Writers of theses in the BA – MA – PhD genre chain must become aware of the advantages and disadvantages of following a relatively rigid organisation structure indicated through conventionalised metalanguage: it is only a scaffold that needs to be filled with content, it may not always be possible, it may be boring if there are not enough specific terms in the subordinate section headings, but it shows a clear structure, which indicates clear and global academic thinking. Sasaki (2000) claims that experts tend to be global planners in essay writing, laying out the big picture before focusing on the details; experts' plans are more elaborate whereas novices' plans are more localized.

The choice depends on the preferences of individual writers and their instructions. The writer of the thesis whose contents we see in Figure 5 gives a definition of the key term "migration", distinguishing different types and discussing two theories – this also shows clear academic thinking, although it uses fewer general metalanguage terms and does not use expected general standard phrases:

2. Theoretical background
2.1. Definition of the key concept migration
2.2. Types of migration
2.3. Theories explaining migration
2.3.1. Ravenstein's Migration theory
2.3.2. Lee's Migration Theory: Push and Pull Factors
3. Case study: Polish migration to the UK
3.1. Behaviour
3.2. Motivation etc.

The "historical perspective" is certainly an option in many theses in humanities and social sciences (like the example in Figure 5). Whether the subheadings should be more explicit like

4.1. Polish Migration to the UK before the accession of Poland to the EU in 2004, etc.

may be a matter of taste, the fewer metalanguage expressions, the more variation is possible and every writer has to choose in the editing phases whether he/she would like to be more specific and thus make the text more interesting, but also more work for the reader. The compromise between the global picture and the attractive detail is the challenge.

When we look at academic texts written by younger scholars, the organisation structure is not so clear yet, but indications of IMRAD structures can be found in linguistic papers (Figure 4) more than in literature papers; sometimes they can be found in higher levels like chapter headings and not in lower levels (as in Figure 6). This is also a good argumentative structure.

Table of Contents

1. Introduction 3

2. Theoretical Approaches

2.1. Key concepts 4

2.2. Interpretation of dreams – important terminology 5

3. Text Analysis

3.1. Mohsin Hamid's *The Reluctant Fundamentalist* 6

3.2. Dreams in *The Reluctant Fundamentalist* 7

3.3. Dreams versus reality 9

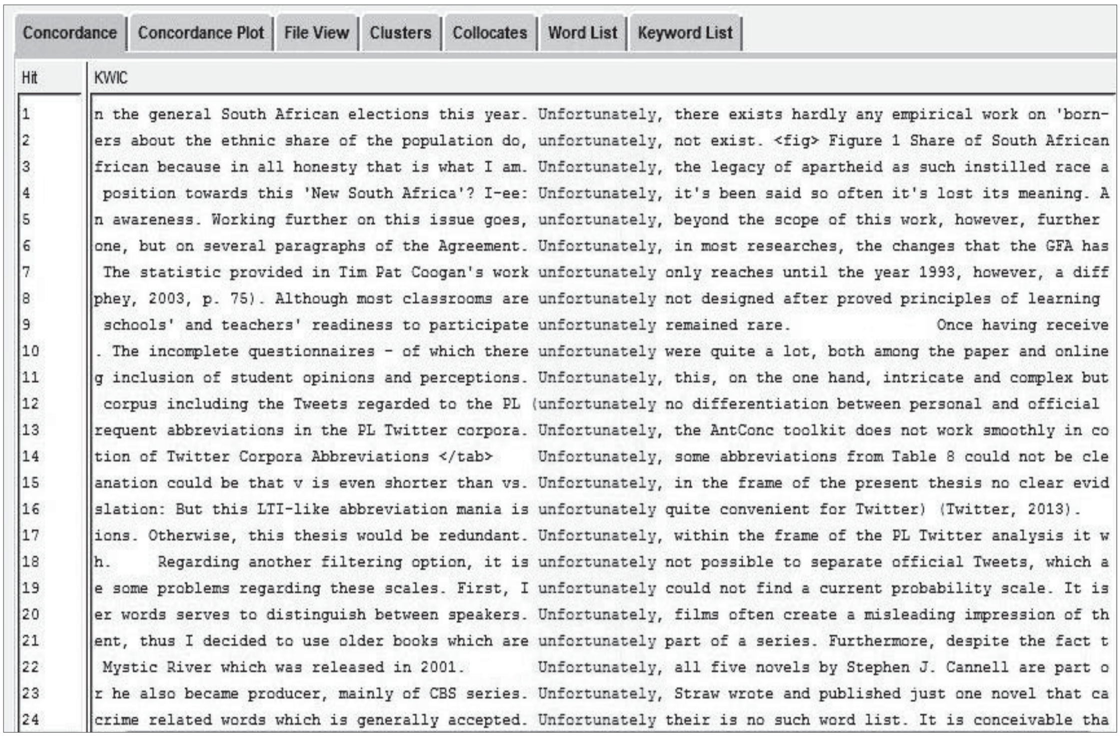
4. Conclusion 12

Figure 6: Table of contents of a literature BA term paper from the *ChemCorpus* indicating an IMRAD structure in the chapters and a non-IMRAD structure in the lower levels

5.2. Linking

Linking devices have been considered central in text analysis and text production since Halliday & Hasan's seminal book *Cohesion in English* from 1976 and they have been central to all student instruction and have become a topic in many student analyses (cf. Albrecht 2015). Although linguists do not always agree on the complete list of linking devices, the most central indicators are *and* with additive, *then* with sequential, *but* with adversative, *because* with causal function. Research hypotheses could not only include the number of linkers, but also the types and even more specific or peripheral choices.

I will only illustrate the usage of an unusual, peripheral linking adverb. Few studies of cohesive devices include the lexeme *unfortunately*, which is not as rare or "subjective" as many young researchers in academic writing assume. The following *AntConc* screenshot (Figure 7) demonstrates the usage in the *ChemCorpus*, where it is almost always a sentence adverbial set off by a comma, often collocating with negation (*no/not*) and/or *existence* (e.g. when identifying the research gap that needs to be filled). Here the use of *unfortunately* appears not unscientific, but quite acceptable, when it is used in the meaning of "contrary to expectation or popular opinion". Generally, this is a good example, how junior writers can find out whether and how certain expressions are really used in the texts of their academic community, beyond general style guides and other traditional more prescriptive sources of information.



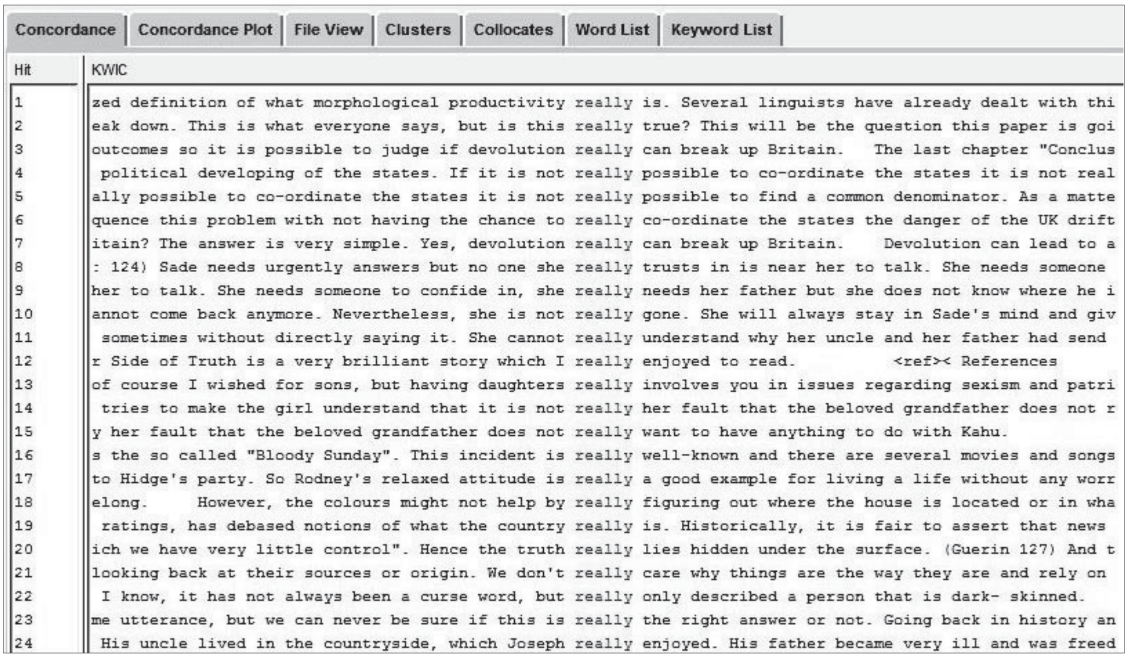
Concordance | Concordance Plot | File View | Clusters | Collocates | Word List | Keyword List

Hit	KWIC
1	n the general South African elections this year. Unfortunately, there exists hardly any empirical work on 'born-
2	ers about the ethnic share of the population do, unfortunately, not exist. <fig> Figure 1 Share of South African
3	frican because in all honesty that is what I am. Unfortunately, the legacy of apartheid as such instilled race a
4	position towards this 'New South Africa'? I-ee: Unfortunately, it's been said so often it's lost its meaning. A
5	n awareness. Working further on this issue goes, unfortunately, beyond the scope of this work, however, further
6	one, but on several paragraphs of the Agreement. Unfortunately, in most researches, the changes that the GFA has
7	The statistic provided in Tim Pat Coogan's work unfortunately only reaches until the year 1993, however, a diff
8	phey, 2003, p. 75). Although most classrooms are unfortunately not designed after proved principles of learning
9	schools' and teachers' readiness to participate unfortunately remained rare. Once having receive
10	. The incomplete questionnaires - of which there unfortunately were quite a lot, both among the paper and online
11	g inclusion of student opinions and perceptions. Unfortunately, this, on the one hand, intricate and complex but
12	corpus including the Tweets regarded to the PL (unfortunately no differentiation between personal and official
13	requent abbreviations in the PL Twitter corpora. Unfortunately, the AntConc toolkit does not work smoothly in co
14	tion of Twitter Corpora Abbreviations </tab> Unfortunately, some abbreviations from Table 8 could not be cle
15	anation could be that v is even shorter than vs. Unfortunately, in the frame of the present thesis no clear evid
16	slation: But this LTI-like abbreviation mania is unfortunately quite convenient for Twitter) (Twitter, 2013).
17	ions. Otherwise, this thesis would be redundant. Unfortunately, within the frame of the PL Twitter analysis it w
18	h. Regarding another filtering option, it is unfortunately not possible to separate official Tweets, which a
19	e some problems regarding these scales. First, I unfortunately could not find a current probability scale. It is
20	er words serves to distinguish between speakers. Unfortunately, films often create a misleading impression of th
21	ent, thus I decided to use older books which are unfortunately part of a series. Furthermore, despite the fact t
22	Mystic River which was released in 2001. Unfortunately, all five novels by Stephen J. Cannell are part o
23	r he also became producer, mainly of CBS series. Unfortunately, Straw wrote and published just one novel that ca
24	crime related words which is generally accepted. Unfortunately their is no such word list. It is conceivable tha

Figure 7: AntConc concordance (KWIC) of *unfortunately* in BA term papers the *ChemCorpus*

5.3. Stance and engagement markers

A very wide aspect of writer – reader interaction is the use and choice of stance and engagement markers (cf. Figure 1, p. 4 above). Since they have been studied



Concordance | Concordance Plot | File View | Clusters | Collocates | Word List | Keyword List

Hit	KWIC
1	zed definition of what morphological productivity really is. Several linguists have already dealt with thi
2	eak down. This is what everyone says, but is this really true? This will be the question this paper is goi
3	outcomes so it is possible to judge if devolution really can break up Britain. The last chapter "Conclus
4	political developing of the states. If it is not really possible to co-ordinate the states it is not real
5	ally possible to co-ordinate the states it is not really possible to find a common denominator. As a matte
6	quence this problem with not having the chance to really co-ordinate the states the danger of the UK drift
7	itain? The answer is very simple. Yes, devolution really can break up Britain. Devolution can lead to a
8	: 124) Sade needs urgently answers but no one she really trusts in is near her to talk. She needs someone
9	her to talk. She needs someone to confide in, she really needs her father but she does not know where he i
10	annot come back anymore. Nevertheless, she is not really gone. She will always stay in Sade's mind and giv
11	sometimes without directly saying it. She cannot really understand why her uncle and her father had send
12	r Side of Truth is a very brilliant story which I really enjoyed to read. <ref>< References
13	of course I wished for sons, but having daughters really involves you in issues regarding sexism and patri
14	tries to make the girl understand that it is not really her fault that the beloved grandfather does not r
15	y her fault that the beloved grandfather does not really want to have anything to do with Kahu.
16	s the so called "Bloody Sunday". This incident is really well-known and there are several movies and songs
17	to Hidge's party. So Rodney's relaxed attitude is really a good example for living a life without any worr
18	elong. However, the colours might not help by really figuring out where the house is located or in wha
19	ratings, has debased notions of what the country really is. Historically, it is fair to assert that news
20	ich we have very little control". Hence the truth really lies hidden under the surface. (Guerin 127) And t
21	looking back at their sources or origin. We don't really care why things are the way they are and rely on
22	I know, it has not always been a curse word, but really only described a person that is dark- skinned.
23	me utterance, but we can never be sure if this is really the right answer or not. Going back in history an
24	His uncle lived in the countryside, which Joseph really enjoyed. His father became very ill and was freed

Figure 8: AntConc concordance (KWIC) of *really* in BA term papers in the *ChemCorpus*

extensively qualitatively and quantitatively, I only want to illustrate some borderline cases (again) where students often hesitate, because they seem too emotional or subjective. The use of *really* has been strongly discouraged in academic writing (and its meanings questioned), but Figure 8 illustrates that it is still used by student writers (also in negative collocations) and this usage sounds lively, but not too informal.

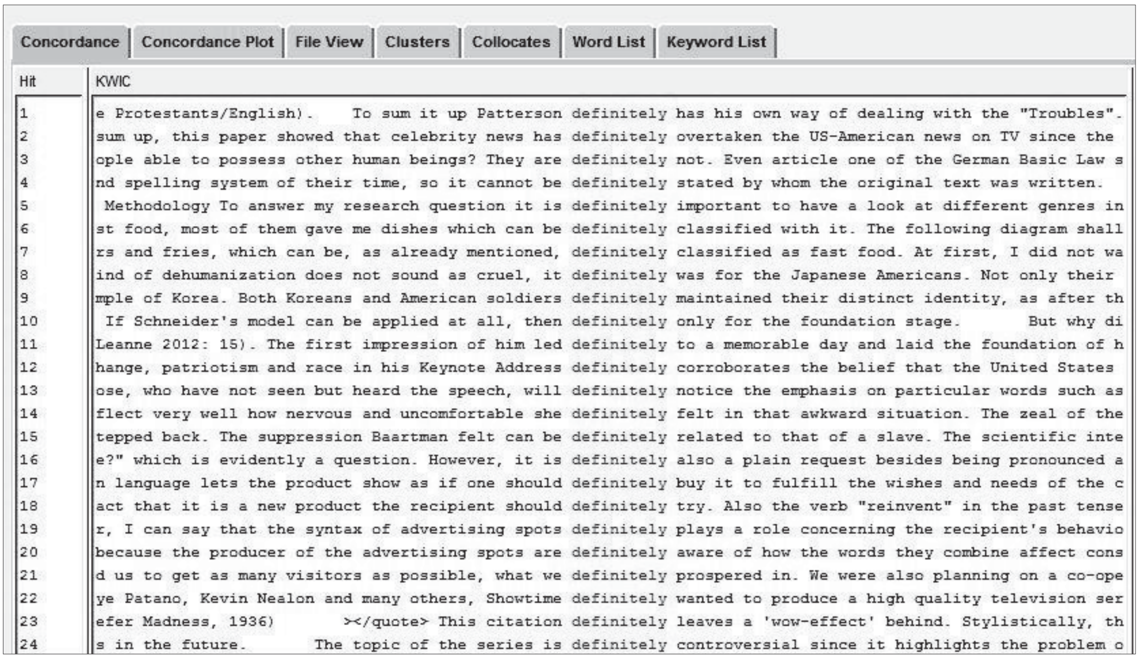


Figure 9: AntConc concordance (KWIC) of *definitely* in BA term papers in the *ChemCorpus*

Finally, the use of *definitely* demonstrates a strong stance by the author who premodifies verbs to boost his/her argumentation (Figure 9).

In our discussion, we have tried to demonstrate the usefulness of tools that help the young academic writer: Modern text-processing systems even suggest punctuation (with explanations and examples in their spell/style/grammar checkers). Modern corpus-linguistic tools allow them to check whether expressions are within the conventions of their discipline. Of course, these tools can not only be used to check whether an expression is used at all, but also by whom and in which contexts. Thus a good corpus basis (with gender, educational and language background and writing context as variables) makes many more socio- and textlinguistic analyses possible that could not be elaborated in this contribution but can be seen in many other contributions (e.g. Beyer 2015 or Edusei 2015).

6. Conclusion

This contribution has attempted to demonstrate to graduate students (and their teachers) that a few key concepts like metalanguage (features) and genre (conventions) are enough to make academic writing relatively easily accessible to graduate students. If they follow a practical, research-based and empirical (corpuslinguistic) approach, they can construct their own model collections using the *ChemCorpus* as a reference corpus that offers a (partial) comparative data-base that can be used to find models for their own writing, solutions to formal and concrete queries and help to find their own stance in practical writing challenges.

The explorative examples used in our *ChemCorpus* analyses illustrate how interactive resources can be used by writers to manage the information flow to persuade their readers to adopt their preferred interpretations. In terms of functional grammar, writers seek to display an interpersonal tenor consistent with the disciplinary identity they wish to project. This can be negotiated with the gate-keepers of international sciene only to a acertain extent (Pérez-Llantada 2012: 151f).

Metalanguage comprises all expressions that organize the content and convey the author's beliefs and attitudes towards it. Researchers do not simply discuss facts or ideas, they also wrap up their content in metadiscourse, i.e. seek to claim solidarity with their readers, evaluate previous research and their own analyses, acknowledge alternative views, etc. As Hyland (2012: 206) wrote:

> Raising student's awareness of the language options available to them in negotiating an identity they feel comfortable with is also important in EAP classes. Once again, teachers can use corpus evidence to help students move beyond the conservative prescriptions of textbooks and style guides and into the preferred patterns of expression of their disciplines. An orientation to instruction based on access to choice through genre teaching and consciousness-raising can help students understand how writing conventions are enabling rather than deterministic. It can reveal the ways that typical patterns provide broad parameters of choice through which they can craft a distinctive self.

References

Albrecht, S. (2015). Connecting adverbials in student writing: Usage in the ChemCorpus and comparable corpora. This volume, 67-80.

Bhatia, V. K. (2004). *Worlds of written discourse: A genre-based view*. London: Continuum.

Beyer, D. (2009). Gender and hedging in academic writing: A comparison of gender-preferential usage of hedges in ESL learners' Magister theses [unpublished BA thesis]. TU Chemnitz, Germany.

Beyer, D. (2015)."It is probably the reason why…". Hedging in BA and MA theses by German ESL students. This volume, 81-98.

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

Biber, D., S. Conrad & R. Reppen (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D., Babieri, F. (2006). Lexical bundles in university spoken and written registers. *English for Specific Purposes 26,* 263-286.

Biber, D., Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes 9,* 2-20.

Davies, M. (2004-). Brigham Young University corpora availbale at http://corpus.byu.edu/

Diani, G. (2004). A genre-based approach to analysing academic review articles. In M. Bondi, L. Gavioli & M. Silver (eds.). *Academic discourse, genre and small corpora*, Rome: Officina edizioni, 105-126.

Dontcheva-Navratilova, O. (2012). Lexical bundles in academic texts by non-native speakers. *Brno Studies in English 38 (2),* 37-58.

Edusei, J. (2015). Code glosses in student writing: A comparative study of Albanian and German BA theses. This volume, 119-130.

Gavioli, L. (2006). *Exploring corpora for ESP learning*. Amsterdam: John Benjamins.

Greenbaum, S. (1996). *The Oxford English grammar*. Oxford: Oxford University Press.

Halliday, M. A. K., Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hyland, K. (2004). *Genre and second language writi*ng. Ann Arbor: University of Michigan Press.

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, *7 (2),* 173-191.

Hyland, K. (2012). *Disciplinary identities: individuality and community in academic writing*. Cambridge: Cambridge University Press.

Hyland, K. & G. Diani (2009). (eds.). *Academic evaluation: Review genres in university settings*. Basingstoke: Palgrave Macmillan.

Pérez-Llantada, C. (2012). *Scientific Discourse and the Rhetoric of Globalization*. London: Continuum.

Mauranen, A. (2013). *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge: Cambridge University Press.

Motta-Roth, D. (1998). Discourse analysis and academic book reviews: a study of text and disciplinary cultures. In I. Fortanet (ed.), *Genre Studies in English for Academic Purposes*, Universitat Jaume: Castelló de la Plana, 29–58.

Salazar, D. (2014). *Lexical Bundles in Native and Non-native Scientific Writing: Applying a corpus-based study to language teaching.* Amsterdam: John Benjamins.

Simpson-Vlach, R. & N. C. Ellis (2010), An academic formulas list: new methods in phraseology research, Applied Linguistics, 31 (4), 487-512.

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing, 9 (3),* 259-291.

Schmied, J. (2011). Academic writing in Europe: A survey of approaches and problems. In J. Schmied (ed. 2011), *Academic writing in Europe: Empirical perspectives*, Göttingen: Cuvillier, 1-22.

Schmied, J. (2012). Academic knowledge presentation in MA theses: From corpus compilation to case studies of disciplinary conventions. *Brno Studies in English 38*, 149-164.

Schmied, J. (2013). English for Academic Purposes: Contrastive Perspectives in the Curriculum. In C. Haase & J. Schmied (eds.). *English for Academic Purposes: Practical and Theoretical Approaches*. Göttingen: Cuvillier, 19-40.

Schmied, J. (2015). Data from the internet for language and culture studies. A corpuslinguistic appraisal. In B. Lewandowska-Tomaszczyk, M. Kopytowska, J. Osborne & J. Schmied (eds.). *Languages and cultures in contact and contrast.* Chambéry: Éditions de l'Université de Savoie.

Siepmann D., J.D. Gallagher, M. Hannay & J. L. Mackenzie (2011). *Writing in English: A guide for advanced learners*. Tübingen: A. FranckeVerlag.

Smirnova, A. (2013). Argumentative patterns of specialized academic discourse produced in English by Russian scholars. In J. Schmied & C. Haase (eds.). *English for academic purposes: Practical and theoretical approaches,* Göttingen: Cullivier, 113-122.

Swales, J. (1990). *Genre analysis - English in academic and research settings.* Cambridge: Cambridge University Press.

Swales, J. (2004). *Research genres: Exploration and applications*. Cambridge: Cambridge University Press.

Swales, J. M. & Ch. Feak (2012). *Academic writing for graduate students* (3rd ed.). University of Michigan: University of Michigan Press.

Trajkova, Z. (2015). Expressing commitment and detachment in non-native English BA graduate papers. This volume, 143-162.

Vogel, R. (2013). Sentence adverbials in academic texts: Preferences of native vs. non-native writers. In J. Schmied & C. Haase (eds.). *English for academic purposes: Practical and theoretical approaches*, Göttingen: Cullivier, 81-94.

Wagner, S. (2011). Concessives and contrastives in student writing. In J. Schmied (ed.). *Academic writing in Europe: empirical perspectives. Research in English and applied linguistics*. Göttingen: Cuvillier Verlag, 23-49.

English for English Studies: Designing Academic Vocabulary Instruction for English Studies Majors

Helen Forbes (Technische Universität Chemnitz)

Abstract

Theoretically rooted in established research of faculty perceptions of L2 academic writing, this study contributes to recent explorations in differentiated learning outcomes for vocabulary instruction in *English for Academic Purposes* (EAP). Two instructional units were developed using the studies revealing faculty examiners' preference for lexical quality in academic writing. The instructional units were implemented and evaluated on the criteria of writing gains and learner engagement. Using professional raters and classroom observations, the study found both instructional units led to writing gains but learner engagement was considerably higher in the second of the two units. The implications for vocabulary teaching and further research are discussed.

1. Introduction

Owing to EAP's genesis as strand of English for Specific Purposes (ESP), the role of vocabulary tuition and development has enjoyed thorough examination from multiple perspectives: Corpus linguists have analyzed both expert and novice production in search of trends that inform the development of both materials and syllabi; second language acquisition specialists have investigated the relationship between different learning strategies and acquisition rates; and practitioners have assessed the merits and shortcomings of applying these materials and theories in classroom settings. The quest for best practice in EAP vocabulary instruction has resulted in an enormous body of research.

Whilst this knowledge remains relevant for EAP learners in exam situations and day-to-day participation in university life, it remains faithful to the traditional vocabulary learning outcomes of comprehension, memorization, recall and active use. These traditional learning outcomes assume the challenge facing the learner is a standardized test consisting of four components measuring linguistic performance in each skills area. Overseas students, postgraduates and English Studies majors, however, face a different challenge; these students must complete their entire degree program in English. Consequently, their academic achievement is predominantly dependent on performance in only one skill: academic writing.

Ma rightly highlights this disparity in recent exploratory research pursuing the links between learning strategies and lexical quality in free L2 academic writing (Ma 2013: 238). This study intends to contribute to the research exploring learner-relevant learning outcomes by exploring the viability of abandoning the notions of memorization and recall altogether. Based on the assumption that use in independent academic writing projects is the most valuable vocabulary learning outcome to English Studies majors, this study experiments with conceptualizing the research determining faculty raters' perception of L2 academic writing in an attempt to teach 'word choice' as a writing subskill. The desired learning outcome is that post-instructional learners will be able to adroitly select the lexical items that faculty appreciate for use in their independent academic writing projects without having attempted comprehension, memorization or recall.

2. Faculty perceptions of L2 writing quality and the EAP classroom

A multi-angled approach lends confidence to the claims that vocabulary use is the most powerful linguistic factor influencing faculty's perception of L2 writing quality. Surveys of faculty feedback (Roberts & Cimasko 2008), faculty perception tests (Santos 1988), textual analyses of pre-rated L2 writing (Engber 1995 and McNamara, Crossley & McCarthy 2010) and comparisons of vocabulary size and academic success (Laufer & Nation 1995) all confirm faculty's positive or negative response to cultivated or unrefined vocabulary use respectively. Indeed, these researchers are the voices petitioning for a 'greater emphasis on vocabulary improvement and selection' in the EAP classroom (Santos 1988: 69).

The problem, however, is a lack of compatibility between the research's individual findings and the EAP classroom. Lexical sophistication, lexical specificity and lexical diversity are incongruent with language teaching principles for unique reasons. Lexical sophistication poses the EAP practitioner two challenges; even if the hurdle of exposure is overcome, there is an axiomatic futility involved in allowing language whose desirability is defined by exclusivity to enter a classroom with the aim of boosting learners' academic success. Lexically specific items have a low-surrender value when learnt in isolation that dissuades the EAP practitioner from investing the class time required for meaningful knowledge. Lexical diversity compounds the divergence between the 'ideal' classroom for acquiring lexical quality and reality by both expanding and increasing the difficulty of the task of vocabulary learning.

Although the correlation between lexical diversity and holistic scores of writing quality fluctuates in different contexts (Jarvis 2002), the divergence of lexical diversity development in L1 and L2 learners is arguably the most depressing for language learners and teachers. Second language acquisition specialists tentatively conclude from data that vocabulary acquisition is supported by lexical networks which are more complete in L1 learners (Crossley & McNamara 2009 and Hunston, Francis & Manning 1997). The acquisition of lexical diversity, therefore, not only demands a considerable time commitment for L2 learners but its perpetual nature also necessitates a long-term diligence. Lloyd understates the issue somewhat in his dissection of the differing mechanisms fuelling the L1 and L2 lexicon when he writes 'this offers little in the way of immediate comfort' to learners in EFL contexts (Lloyd 2008 para. 10). Admittedly, this appears to hold true for all academically successful vocabulary as the research indicates the vocabulary most appreciated by faculty is unlikely to be encountered, has limited applications and is difficult to remember.

Given the disparity between the linguistic features of writing quality and the traditional principles of the EAP classroom, this study deliberately dispenses with the aim of expanding the lexicon in favor of training lexical selection as a writing subskill. The first instructional unit explores teaching lexical sophistication by sensitizing learners to the features of lexical quality and training learners to use the English Vocabulary Profile corpus from Cambridge ESOL. The rationale is research suggesting raters value lexical sophistication regardless of the presence of lexical errors (Llach 2007). The second instructional unit is guided by research demonstrating raters severely penalize lexical errors (Fritz & Ruegg 2013 and Vann, Meyer & Lorenz 1984) and emphasizes lexical appropriacy.

3. Research design

An action research methodology was deemed appropriate owing to the participant driven[1] and context-bound nature of the study. Furthermore, the study's aim was an improvement in the researcher's own teaching practices and it was unlikely that the researcher would find the resources to not be a participant in the study. Specifically, Nunan's seven step cycle (Hien 2009) was chosen on account of being the only model with satisfactory room for the preliminary investigation the genesis of the project necessitated. Following the preliminary investigation and subsequent formation of hypotheses, a teaching intervention and evaluation procedure for the first cycle were designed and implemented. The results of the first cycle guided the planning of a second comprehensive cycle.

The preliminary investigation was a cursory manual analysis of learner academic writing and expert academic writing (both L1 and L2). It was hypothesized that learners overestimate the semantic transfer in their own writing. Computational analyses confirmed differing ratios of lexical richness. The most striking difference was the learners' reliance on words in the General Service list. This analysis informed the design of the first instructional unit that targeted lexical sophistication.

Perception tests of writing quality with human raters were vital in evaluating the learning outcomes. Two professional raters volunteered their services for the research: one NNS with a higher degree, initial teaching certificate, ten years teaching experience and certification for conducting the official exams of the German Federal Office for Migration; and one NS with a first cycle degree, initial teaching certificate and eight years teaching experience. The raters were given a time limit during the rating activities of both the first and the second instructional units. This was an attempt to replicate the examination context of an

[1] The researcher's learners completed a survey whereby they could ask for support in any area of academic writing. The most common responses were time management, motivation and 'sounding academic'. This study is part of the instruction aiming to satisfy the request for support in 'sounding academic'.

overworked faculty member who must also evaluate the content, presentation and structure of learner academic writing.

Classroom observations monitored learner engagement to serve as a basis for meaningful reflection and further development of instructional units. Systematic sampling was mixed with a naturalistic focus on critical incidents to record the most holistic and objective picture possible. Systematic sampling ensured a reduction in inference and subjectivity. The coding scheme from the Annenberg Institute at Brown University was considered the most relevant as it targets learner engagement. The measures of 'high', 'low' and 'very low' conform to the three scale system of MOLT Classroom Observation Scheme (Guilloteaux & Dörnyei 2008) and the sampling was conducted at each stage of the lesson procedure.

Naturalistic observation was used to record critical incidents. The purpose was not only to record the stimuli of learner behavior but to also counteract the coding scheme's whitewashing effect that could hinder meaningful reflection. The traditional coding schemes are constructed to document the presence or absence of learner behaviors that signify a certain level of learner engagement. The behaviors appear in a hierarchy with the aim of gauging the level of engagement. Subsequently, the positive end of the scale is more nuanced than the negative end of the scale and without cautious implementation instances of high learner engagement can be recorded more frequently. Additionally, without naturalistic observation to document critical incidents of low learner engagement, the data could only state that the instructional unit was suboptimal; there would have been no data to aid development of an improved unit.

The first instructional unit was conducted with nine volunteer learners. The second instructional unit was integrated into an academic writing skills course (see explanation of genesis). On the day of instruction, fifteen learners were present. Barring one ERASMUS student, all learners were full-time English Studies students at Chemnitz University of Technology and they all needed to complete substantial academic writing projects in the time periods surrounding the instruction. The ERASMUS student was an English major at her home university who was completing academic writing projects issued by the faculty at Chemnitz University of Technology in the time before and after the instruction.

Owing to the international composition of the learner groups, learners had differing levels of experience with academic writing, differing levels of proficiency in English and different L1s. The majority of the learners have also been active in academic writing cultures outside Germany submitting work in both English and the respective country's official language (in some instances this was the learners' L1 and in some instances a further L2). The overwhelming majority of participating learners were EFL learners but there were two ESL learners and one NS. The length of time spent at university ranged from 1.5 semesters to 10.5 semesters.

3.1. Results

To ease reading and provide an overview of the lesson procedure, the following section represents the results in chronological order of acquisition.

3.2. Instructional unit 1

3.2.1. *Lesson procedure and classroom observation*

Stage Description	Stage Aim	Learner Behavior
Lead-in		
Ss are given a handout with non-academic/pseudo-academic sentences and are asked to rate on a scale of one to ten how academic the sentences sound. Solowork – Ss must commit to a number on paper. T monitoring for concept checking. T collects handouts – Ss are unaware of peer answers.	Introduction to expectations of academic language. Informal diagnostic test of baseline knowledge about sounding academic.	Ss work actively on task. Ss follow directions. Ss struggle with different sentences – no obvious pattern to the judgments individual learners find challenging. All Ss identify the least academic sounding sentence as such and give it a correspondingly low rating. Ss consider the other non-academic/pseudo-academic sentences as slightly academic (4) to academic (8).
Language focus		
Ss are given a noticing exercise handout with parallel sentences: sentences from the lead-in and their academic equivalents. Ss re-rate the first sentence using the one to ten scale. Ss look for the linguistic differences between the academic and non-academic versions of the sentences.	Inductive presentation of the complexity and density of academic writing and activation of existing knowledge.	Ss follow directions. Ss giggle at their previous answers – T abandons re-rating exercise as it seems this learning outcome is resolved. Ss work actively on the analysis task. Ss cooperate. Ss discuss rigorous content with peers.
Reporting and presentation of resource		
Ss report the differences they have found. T collates results on board. Ss and T discuss the results briefly. T paraphrases the complexity and length to sophistication. Ss and T discuss the CEFR level that should be displayed in university writing. T uses projector to introduce the Ss to the English Vocabulary Profile.	Language focus of the textual devices that lead to sounding academic with the emphasis on the features of words used. Language focus on lexical sophistication. Learners hear that university writing is CEFR B1/B2 at Bachelor level and CEFR B2/C1 at Masters level. Introduce the functionality and operation of English Vocabulary Profile.	Ss report complexity, length, density and nominalization. Ss display high volunteering. Ss display high participation. Ss copy from projection.

Controlled practice		
Ss receive a passage of student writing and use the English Vocabulary Profile to grade the passage in pairs or threes.	Ss practice using the English Vocabulary Profile. Ss practice assessing the CEFR level of writing. Concept checking.	Ss work actively on task. All but 1 Ss follow directions. Ss perform appropriate manipulation of materials. All but 1 student report correct level (A2). S who didn't follow directions reports A1.
Free practice		
Ss write-up the passage to a higher CEFR level using the English Vocabulary Profile (Ss may use additional resources they have experience in). T tells Ss that they will get feedback from more than one source.	Ss practice changing a non-academic sounding text to an academic sounding text. T motivates Ss by encouraging them to make the most of the feedback opportunity.	Ss work actively on task (2 very slowly – T gives time limit). Ss follow directions. Ss raise questions with T and peers. Ss begin to become frustrated that many sophisticated lexical items from other sources are not in English Vocabulary Profile.
Survey		
Ss receive a survey for ordering tailored lessons on academic writing.	T compensates learners for their participation in her research.	Not part of instructional unit.

3.2.2. *Rater perception test*

The raters were given the CEFR A2 passages the participants had 'written-up' using the English Vocabulary Profile following the instruction's language focus on lexical sophistication. The raters were given an upper time limit of 20 seconds and directed to make more of a snap judgment of the writing quality than a considered appraisal (as may be the case in a high-stakes rating situation, for example a language test for a country's visa). This direction was elucidated by describing the study's context was learner academic writing for faculty and the focus was creating a good impression on the examiner who would ultimately need to read up to fifteen pages of the learner's writing. The raters were not informed of the instructional unit's specific language focus until after the rating had finished. As it happened, 17 of the 18 ratings were issued decisively within five or six seconds. In only one instance did there seem to be any uncertainty. The raters were instructed to use the CEFR levels. The CEFR levels were used for parity and the learners are aware of what approximate proficiency the labelling system refers to.

The instructional unit's nine scripts were rated as follows:

CEFR level rater 1	CEFR level Rater 2	No. of Scripts
C1	C1	5
C1	B2	0
B2	B2	0
B2	B1	2
B1	B1	2

3.3. Instructional Unit 2

3.3.1. *Lesson procedure and classroom observation*

Stage Description	Stage Aim	Learner Behavior
	Lead-in	
T introduces topic of 'sounding academic' on the board. T introduces topic of lexical richness and tells Ss that research shows 'sounding academic' is about lexical richness. Ss translate lexical richness into everyday language. T writes answer (a lot of vocab) on board.	Introducing topic and signposting lesson objectives. Communication of out-of-lesson application for motivation. Eliciting existing knowledge and concept checking.	Ss display high participation. Ss display high volunteering.
	Skills focus	
T tells Ss lexical richness can be divided into lexical sophistication and lexical density, lexical density will be handled later and starts board plan. T elicits features of sophisticated lexis from students. T tells Ss lexis must also have appropriate denotation, connotation and be neutral. T elicits explanations of denotation, connotation and neutral and non-examples.	Emphasizing correctness. Activating existing knowledge. Explaining and concept checking.	Ss display high participation. Ss display low volunteering. Ss volunteer specific, formal, complex, long. Ss need support explaining denotation, connotation.

Controlled practice		
Ss work in groups of four on an exercise finding the non-academic words in learner produced sentences in given problem areas (Ss have both print and electronic resources). T collates the negotiated definitions of denotation, connotation, formality, complexity and neutrality on the board T monitors (lightly as learners will need to do this autonomously for their academic writing and learners are cooperating well and staying on-task with high energy).	Reinforce concepts from boardwork. Provide authentic examples of concepts from boardwork. Practice the skill of identifying non-academic words. Personalizing boardwork to student answers.	Ss work actively on task. Ss follow directions. Ss display high levels of thinking and understanding. Ss display high energy for problem solving despite this being a challenging task for them. Ss discuss rigorous content with peers and T Ss display student-directed enquiry (some Ss are discussing the instruction's content in relation to their own weaknesses, some Ss are discussing what the most useful part has been, some Ss are discussing how the instruction could develop to be more useful and suggesting topics and talking about the knowledge of academic writing that they would still like to acquire)
Reporting and feedback		
T and Ss compare answers in plenary	Producing correct answers	Ss report neutrality and formality first (correctly with high volunteering) Ss display lower volunteering but connotation is offered with only gentle persuasion from T Ss display no volunteering for the remaining two exercises (T instructs class to look at complexity and nominates S to report their answer, first answer is incorrect, second nominee answers correctly but it may have been a lucky guess) Ss and T repeat the same procedure for denotation

Warm-up (homework)		
T highlights lexical density on the board T elicits materials that are dense Ss are given a two-sided passage of learner academic writing and are asked to summarize it in 20 words for homework	Introducing new topic Sensitising Ss to lexical meagerness Non-example of lexical density	Ss display high participation Ss check understanding of assignment
Reporting and lead-in		
T draws Ss attention to board plan from previous lesson Ss report their summaries of homework passage T and Ss discuss the lexical sparsity and repetition of information T and Ss compare lexically meager to lexically rich.	Warm-up and contextualization Producing appropriate answers Discussion about the best practice for academic writing	Ss display high participation Ss display high volunteering Ss taking notes All Ss have done homework
Free practice		
Ss rewrite learner passage in groups of four for lexical richness (learners have electronic and print resources). T distributes a best copy answer sheet. T monitors.	Practicing producing lexically rich texts.	Ss follow directions. Ss work actively on task. Ss autonomously problem solve. Ss apply metacognitive strategies.
Feedback		
Ss receive papers with comments. Ss and T discuss answers.	Motivate and produce correct answers.	Not part of instructional unit.

3.3.2. *Rater perception test*

The raters were given the previously lexically meager passages that had been rewritten by the groups of participating learners following the instructional unit on lexical richness. The raters were given an upper time limit of 30 seconds and directed to find lexical errors. The raters could find no lexical errors within the time limit. During the feedback phase, the researcher did notice one connotation error as part of a much closer analysis. There was also one proofreading error (word repetition) that the raters were not instructed to locate but can be included here for completeness. There were no further linguistic errors present in the five scripts the groups had produced. One script, however, did not include any lexical items not in the original. On account of this script's lexical density, clarity and correctness, it can still be said that the quality of the writing is higher but unfortunately it is not possible to claim that the instructional unit's language focus of lexical richness is represented in this script. On a more positive note, the remaining four scripts do include sophisticated lexical items not present in the lexically meager text and the lexical items have been used correctly.

4. Discussion

At the beginning of the first instructional unit the participating learners were unable to recognize lexis that was not sophisticated enough for academic writing as such. During the free practice phase of the unit participating learners were in a position to select sophisticated lexical items and apply them in their scripts. It would therefore seem that teaching word choice as a writing subskill is possible. Additionally, the perception tests with human raters appear to confirm that lexical sophistication in academic writing will be credited despite the presence of lexical errors as all of the scripts rated at CEFR C1 included at least one lexical error.

These results were obviously encouraging but shortcomings were evident in the first instructional unit that must be discussed when considering the further pedagogical implementation of this instruction. Firstly, the observational data document a decline in learner engagement during the free practice phase. It would appear that learners do not find using this resource for free academic writing agreeable and grow frustrated quite rapidly when a resource is perceived as inadequate (the learners voiced their frustration after only one or two instances of searches bearing no results). This clearly invites the presumption that learners are unlikely to use this resource during their independent academic writing projects. This finding need not detract from the value of schooling learners in the functionality of the resource as a proofreading tool for independently identifying the CEFR level of their writings but serves as a limitation in the context of teaching word choice.

Secondly, although the presence of lexical errors does not seem to have a particularly adverse effect on human judgments of writing quality, it unfortunately raises some difficult issues for the EAP practitioner when trying to issue feedback. Providing meaningful feedback on all of the lexical errors would be overwhelming for the learner whilst correcting all of the lexical errors would be demotivating. It was for this reason that the decision was made to include lexical appropriacy in the instruction that was planned as a second cycle of the study.

The positive results from the first instructional unit regarding the viability of teaching word choice as a subskill advocated expanding the scope and time investment of the second instructional unit. A more deductive approach was chosen as it would have been impossible for learners to have been aware of the original conceptualization of lexical richness. Following the instructional unit, the learner scripts displayed writing gains but more importantly, in comparison to the first instructional unit, learner engagement was a lot higher during the second instructional unit.

The learners concentrated on the controlled practice task without coercion for almost fifty minutes and even when the learners' stamina began to wane, their discussions still revolved around the lesson's content without exception. Furthermore, during the free practice stage the learners took the initiative to implement metacognitive strategies and provide peer feedback. Moreover, the learners have continued to mention the instructional unit's content in subsequent lessons and they continue to use the terminology correctly. Learners have even rewritten passages of completed academic writing projects to improve the lexical richness.

An obvious further step for the study would be to try and isolate the reason for the higher learner engagement in the second instructional unit. Additionally, it seems wise to now develop a rating scheme that would facilitate repeating the perception tests with faculty examiners to establish how similar or different the judgments of learner writing quality are. The far-reaching benefits to EAP syllabi of teaching word choice as a subskill certainly justify a continuation of the investigation.

5. Conclusion

This study attempted to explore the viability of teaching word choice as a writing subskill by designing and implementing two instructional units that conceptualize the lexical features that faculty examiners appreciate in free academic writing. Both instructional units lead to writing gains perceivable to human raters but learner engagement was higher in the second instructional unit. To conclude, this study's data indicate that creating agreeable vocabulary instruction with learning outcomes relevant to English Studies majors is possible and can support gains in writing proficiency.

References

Crossley, S. A. & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing, 18,* 119-135.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, *4 (2),* 139-155.

Fritz, E. & Ruegg, R. (2013). Rater sensitivity to lexical accuracy, sophistication and range when assessing writing. *Assessing Writing, 18,* 173-181.

Guilloteaux, M., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL QUARTERLY*, *42 (1),* 55-77.

Hien, T. (2009). Why is action research suitable for education?. *VNU Journal of Science: Foreign Languages*, *25,* 97-106.

Hunston, S., Francis, G. & Manning, E. (1997). Grammar and vocabulary: Showing the connections. *ELT Journal*, *51 (3),* 208-216.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19 (1),* 57 – 84.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16 (3),* 307-322.

Llach, M. P. (2007). Lexical errors as writing quality predictors. *Studia Linguistica*, *61 (1),* 1-19.

Lloyd, M. (2008). 'It just sounds strange, and here's why...'. *IH Journal of Education and Development, 24*. Retrieved from http://ihjournal.com/it-just-sounds-strange-and-heres-why.

Ma, Q. (2013). Matching vocabulary learning process with learning outcome in L2 academic writing: An exploratory case study. *Linguistics and Education*, *24*, 237-246.
McNamara, D. S., Crossley, S.A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27 (1),* 57-86.
Roberts, F., & Cimasko, T. (2008). Evaluating ESL: Making sense of university professors' responses to second language writing. *Journal of Second Language Writing*, *17*, 125-143.
Santos, T. (1988). Professors' reactions to the academic writing of non-native-speaking students. *TESOL QUARTERLY*, *22 (1),* 69-90.
Vann, R. J., Meyer, D., & Lorenz, F. O. (1984). Error gravity: A study of faculty opinion of ESL errors. *TESOL Quarterly, 18 (3),* 427-440.

Interplay of Teacher and Peer Online Feedback for Academic Writing: Preliminary Insights

Elena Ončevska Ager (Ss Cyril and Methodius University-Skopje)

Abstract

This paper reports on a selection of preliminary insights drawn from a small-scale classroom exploration into some of the effects of using a custom-made e-learning platform for the development of students' academic writing skills. The aim of this paper is to look into the outcomes of peer review made public to the class and the teacher joining the student peer community in giving feedback, as evidenced by the students' online behaviours. One such outcome was that the teacher, apart from moderating the students' online exchanges via elicitation, guidance and correction, also implicitly modelled teaching behaviours such as giving feedback for the students who study to be teachers themselves upon graduation. Also noteworthy were the instances of the teacher merely monitoring the students' work, with the students mostly autonomously, though collaboratively, meeting the academic writing syllabus targets.

1. Introduction

In this paper I will report on a small-scale informal investigation of peer feedback going public to the class community (the teacher included) and some of the consequences of such an instructional intervention. After reviewing the literature for insights on peer feedback and learner autonomy, I will provide background information to my classroom research project and introduce the online platform specifically designed for my academic writing courses. I will then go on to present and discuss some of the more notable findings, suggesting implications for the teaching of academic writing at a tertiary level.

2. Online peer feedback

Peer feedback, one of the features of the process-based approach to writing (Hedge 2014), has long been perceived by students as useful and valid, specific (compared to teachers' often general comments on students' writing) and student-friendly: peer audiences are often viewed as more sympathetic than the teacher. Peer feedback is also believed to provide the conditions for students to develop their critical thinking skills by engaging in meaningful encounters in writing (Rollinson 2005). While my teaching context had enabled me to perceive the above positive effects of peer feedback in my classroom, I found that despite the extensive training in peer review my students received, the majority of them continued to produce limited and poorly composed peer feedback. Their comments were often reduced to one-liners such as "excellent" or "poor", which made their feedback less than helpful and constructive.

Reading the work of Mitra (2005), who claims that students function as self-organised systems when they care about the task at hand, sometimes even in teacher-free environments, I set out to observe if and how the quality of my students' peer feedback will change when it is made public to their peers and myself, i.e. when peer feedback is not limited to two readers only (myself and the peer reviewer).

Getting the students to post their work and comment on their peers' work in a specifically designed online forum resulted in surprisingly rich and complementary peer feedback, which at times left me, the teacher, virtually redundant. It can be claimed that teacher redundancy represents the ideal in education, with the students willingly taking full charge of their own learning – very much along the lines of Holec's definition of learner autonomy (1981). In view of the fact that students often struggle to take ownership of their learning when the learning task is so complex as to entail developing a new, academic literacy, considered by Hyland (2012) to be a 'foreign culture' to students, peer feedback going public became a welcome addition to my writing course: the students seemed to feel empowered by being part of a supportive online network which had as its goal acculturating them to the new, academic discourse community. Finally, online environments lend themselves to experiments like this one because they "have the potential for knowledge construction (not just information dissemination)" (Salmon 2003: 45).

3. Background to the study: teaching context and online platform

The classroom exploration project was embedded in a state university setting in the Republic of Macedonia during the academic 2013/14 and involved a total of 30 third-year English language undergraduates who attended 90-minute face-to-face writing classes once a week. The students' class work was supplemented by online work. A complementary e-community was built as an app on Heroku, a free for small-scale projects cloud platform, and was entitled The Bar. The students used The Bar to (1) post their academic writing tasks in a forum, virtually collaborating with their peers and myself and (2) upload their work as part of an electronic portfolio. Indeed, a major part of the class work was geared towards preparing the students to contribute effectively online. Such preparation entailed discussing key notions as a class or the students working in small groups to prepare their work before publishing it on The Bar. On a few occasions, face-to-face classes were substituted with online sessions, with the students watching relevant video material and engaging in out-of-class work to produce the desired written output, to be later posted on The Bar.

4. The classroom exploration project

Observing the students' work on the online platform and myself engaging in providing feedback on their work, I had ample opportunity to notice patterns of student learning and social behaviour. I decided that the most notable areas to look at for the start of what I hope will turn into a more formal and systematic exploration are the following:

- the ways in which the students phrased their feedback
- the ways in which the students engaged in peer and self-teaching
- the opportunities the students took for non-compulsory meaningful interaction.

I will discuss all of the above by analysing exchanges on The Bar. All the students' names were anonymised and their language is presented as it appeared on the platform, i.e. it did not undergo any error correction. My name (Elena) is the only one that was not replaced with a pseudonym. The excerpts analysed were taken from two online tasks:

- Task 1: an individual summary of spoken text, i.e. a TV student debate
- Task 2: a group selection of texts for summarising on topics of the students' choice.

4.1. Phrasing of feedback

Before starting to work on the platform, I exposed the students to a popular article on online netiquette (WebWise 2012) which outlined the importance of being clear, concise and polite when interacting online. The article alerted the students to the intricacies of online communication which arise partly from the limited means of expression compared to face-to-face interactions. The students, therefore, received some broad suggestions about how to behave in the online environment to avoid communication problems.

What they did not receive, however, was advice about any specific feedback strategies as I relied on their common sense to produce constructive feedback. This approach on my part soon proved problematic. The comments provided even by some of the most competent students in the group read somewhat abrupt, less than tactful and were poorly laid out, as can be seen in Ada's feedback (Example 1) below.

(8) your thesis statement focuses on students' rights and institutions who look after their interests whereas the debate was more about whether students fight for their rights and needs and how
(9) another issue that was discussed… -> try to put this issue in the thesis statement
(10) the situation is really bad -> I don't think he said this, be more precise
(11) make it shorter, 50 words at least

Example 1: Sample early feedback by Ada

This called for a note to all students asking them to make sure they provide some positive feedback alongside directions for improvement. The students soon started to mimic my own feedback strategies of not only providing balanced feedback, but also addressing the author of the text in friendly terms, for instance by greeting them first and/or using their first names. Ada soon switched to producing paragraphs of feedback, rather than notes only and phrasing it in a more friendly yet professional way (Example 2).

I think that you three as a group chose a very difficult and complicated topic, which is good because we tend to turn a blind eye to such topics instead of discussing them. There are many arguments pro and against abortion and It isn't possible to present them all, so I think you made a good choice of texts dealing with some crucial aspects of it.

Example 2: Sample later feedback by Ada

4.2. Peer and self-teaching

Another aspect of the students' interaction on The Bar worthy of discussion was the richness of feedback that the authors received, usually from more than one reviewer. Example 3 contains an excerpt from one such exchange where the two peer reviewers comment on different, yet complementary aspects of the author's text, leaving little to myself to add. Indeed, in my comment I merely expand a little on the peer reviewers' observations, but do not bring any new suggestions for improvement. It is noteworthy that both reviewers remembered to phrase their feedback in such a way as to include both positive and negative criticism while suggesting areas for improvement. What does not go on in the interaction captured in Example 3 is inter-reviewer communication in the sense of Sarita commenting on aspects of Sid's feedback – a move which could have made the discussion look more natural.

SID: I must say that it's good that you managed to make a summary which has no more than 100 words. The beginning is good and promissing, but the second part and the ending can be improved. The first mistake is THUS which is (if I can say so) misussed. Try to change something there. In the penultimate sentence you say that Sara can't see ANY positive changes, which is not true. You don't need MOREOVER in the last sentence and Monica doesn't point out that protests are not a solution, but only that protests should be the last choice. Eventhough the second part of the last sentence beggins with BUT, it doesn't contrasts the first part and it's unclear WHO should be more aggressive.

SARITA: Lia, I think that you don't have a clear thesis statement on what you're going to talk about. You should clearly points out the main topics being discussed in the debate. According to the word limit, it's good. The second and the third point are not supported. Please go through your summary once again and try to give examples.

ELENA: Hi Lia, you're one lucky student getting such constructive feedback from your colleagues! :) Sid and Sarita's comments summarise mine, I guess. :) No need to use so many transitions, especially as "furthermore" and "moreover" don't logically belong in the text. The thesis statement definitely needs revising in order to include all the points that you set out to develop.

Example 3: Peer feedback on a summary of spoken text followed by my feedback

The interaction in Example 4 illustrates yet again the multifaceted nature of the peer feedback provided on the platform, this time featuring more peer voices. Unlike in the exchanges presented in Table 3, here some attempts are being made by Sid to establish inter-reviewer communication; namely, he requests clarification from Frida about her feedback. In this case, too, I am at a loss to add any further directions for improvement. What I do in my response is merely summarise Ivette, Ivy and Sid's contributions: the author might find such a concise summary more manageable to work with and less overwhelming. However, there is room for improvement regarding the quality of the feedback provided: Frida's contribution is virtually Ivette's comments rephrased and Sid repeats the point about the length of the authors' texts. More research is needed to establish the reason(s) for such repetition of ideas and more training seems to be necessary to prevent much repetition in online exchanges from happening.

> IVETTE: Girls, you did choose an engaging topic! Being physically active can help us lead healthier lives. However, too much exercise can lead to overtraining, and parents should know how much is too much for their children. I think that you could've mentioned something about this issue. I would also like to say that the excerpts are rather long.
>
> IVY: I noticed some mistakes in referencing too. I hope that you'll take this comment into consideration.
>
> FRIDA: Hi, girls! You're topic is very interesting because as you have mentioned sports are good for life and health. However, I think that you should include excerpt dealing with the negative effects of sport. I agree with Ivana that your excerpts are too long (they should be 100-150 words) and you should pay attention to the referencing as well.
>
> SID: Great topic, I agree with 95 percent of the things mentioned, however the texts are too long, especially Kay's - it contains 300 words. And also, I'm not saying that there aren't negative effects of sport, but I'd like Frida to specify what she thinks about this.
>
> ELENA: Once again, I'm left without any comments of my own. :) Think all has been mentioned: from length via referencing to mentioning "the other side of the coin".

Example 4: Peer feedback on a group post of selected texts for summarising on the topic of sportsmanship followed by my feedback

The exchange in Example 5 illustrates peer teaching of a citing convention. Namely, the students selecting texts for their group requested help with referencing sources with no author and I pointed them to the appropriate conventions – hence Bea's confidence below that they did well referencing their "authorless" source. Parenthetically, this was not a convention that had been discussed in class, which may be the reason for Lia's confusion. Peer feedback in this case effectively substituted teacher instruction: my comment regarding referencing merely endorses Bea's explanation. From the exchange in Example 5 it looks as if the students are very capable of problem-solving via cascading tips that some of them got from myself upon request.

> LIA: Hi Elba, Bea and Sandra, yours is a very interesting topic too, which we debated recently. :) I think you've mentioned all the important points concerning this topic, and I think each of your texts expresses various aspects of the topic (the typical characteristics of a serial killer, their brain, other reasons for their behaviour...). I also have comments on your reference; although Stefani mentioned that the authors were unknown to you, perhaps you could have found some other texts that do have an author! :) All in all, great topic!

BEA: As far as I know, reference is possible with the author being absent. Therefore, we didn't have to search for a text that would have an author. We followed the APA style and managed to write down the reference. :)

ELENA: Hi there Serial Killers Group :), Kudos from myself for your choice of topic and yes, APA has conventions for quoting "authorless texts" - thanks for illustrating that for us. :) I'm worried that not all texts have an accompanying magazine/newspaper reference. As you know, the title of the article only is not enough.

Example 5: Peer feedback on a group post of selected texts for summarising on the topic of serial killers followed by my feedback

Kaia's feedback in Example 6 similarly reinforces content that was discussed in class, namely students' indiscriminate use of transition signals. Her main source of positive feedback comes from the observation that transition signals were not overused – a "trap" students tend to fall in possibly as a result of too strong a focus on such linking devices on their academic writing courses. Such recycling of learning points on the part of the students can help the teacher to gain insight into students' understanding of content material, while providing opportunities for the teacher to denounce some of their own teaching in favour of peer teaching. The phrasing of Kaia's feedback may well be conducive to a motivational boost in the recipient of her feedback; however more research is needed in order to ascertain the impact of well-phrased student feedback on peer authors.

KAIA: Hey Khiara, I read a lot of summaries but I decided to comment on yours. I guess I really could not notice anything that I would like you to change. I really liked that you mentioned everything in such an organized way, with a proper thesis statement and examples. And your summary is a proof that you can nail it without any transition signals. I really like it.

Example 6: Peer feedback on a summary of spoken text

The excerpt below (Example 7) supplies a self-teaching aspect to the student learning process: observing the exchanges on The Bar apparently aided Kaia in the process of noticing too general a wording of the thesis statement in Lia's summary – an oversight she herself had made in her own text.

KAIA: Hi Lia. I agree with all mentioned above, but, as I noticed in my summary too, I think that the thesis statement is too general. Besides this I must admit that you pointed out the most important topics of the debate, and managed to mention them in the given word limit, a thing which all of us had a struggle with. All in all, good job!

Example 7: Peer feedback on a summary of spoken text

4.3. Non-compulsory meaningful communication

Finally, the exchanges on The Bar seem to have provided additional, not previously planned opportunities for meaningful interaction. Some participants, myself included, reported welcoming the opportunity to read about the interesting topics that some groups had selected texts on. This resulted in comments on the content of the texts as part of the feedback they provided – a requirement which was never explicitly made (Examples 8 and 9).

NILDA: I really like your topic. Global warming is so exhausted theme, and it is very interesting to see all this from a different perspective. I was not even aware that it could be some kind of trick or political toy. Thanks for the enlightenment :)

Example 8: Nilda's comment on a group post of selected texts for summarising on the topic of global warming

ELENA: I'm impressed by your selection of texts, Gender Stereotypes Group!! I hadn't been aware about the term "face-ism" before, thanks for introducing me to it. :) Good referencing, too. (Apologies for not being able to format the text properly when you post it as a comment - I'll get that fixed as soon as possible!).

Example 9: Elena's comment on a group post of selected texts for summarising on the topic of gender stereotypes

5. Discussion of findings and implications for academic writing instruction

This informal classroom exploration of making peer feedback public with the teacher joining the peer community at a visible level confirmed Rollinson's (2005) understanding of peer feedback as useful, valid, specific and student-friendly. Indeed, it tapped into students' critical thinking capacities and was by no means reduced to questioningly constructive one-line observations. More research is certainly needed to ascertain the reasons for the breadth of comments the online platform yielded, though some speculations include the change of audience, the novelty of the medium, the accessibility of the medium (say, from the comfort of the students' homes and at times suitable to the students, with some of them accessing the platform well after midnight), the prospects for enhancing students' self-reliance and confidence as writers, etc. as likely factors for the richness of the peer feedback. While most of the students did need teacher guidance for improving their own and/or others' work, some students proved to be perfectly self-sufficient and competent enough to apply the assessment criteria, at times even substituting the teacher. Also, the students took the opportunities to engage in non-obligatory but nonetheless meaningful interaction and improved their feedback strategies following direct and indirect instruction.

I would like to voice a note of caution, though, to any similarly-minded colleagues who are looking to embark on a similar blended academic writing course experiences. First, online netiquette training seems necessary for teachers to put in place at the outset of any blended programme. If this kind of training is to be successful in the sense of proceduralising students' declarative knowledge, it is best designed in an interactive fashion, rather than by merely exposing students to information about online netiquette. Analysing successful and less successful online encounters might also be a good start to such preparation for an online writing context. Second, students seem to need explicit instruction in phrasing feedback in order to prevent misunderstandings and/or online hostility. Implicit instruction in this experiment did work provided it was coupled with explicit appeals for shunning some and embracing other peer feedback strategies. Third, students seem to need more training in how they interact with their peers' prior contributions. Namely, most of my students commented on their peers' work on The Bar failing to acknowledge the comments the authors had already received, which resulted in repetition of ideas. More research is needed to ascertain what the reasons for such redundancy are. Apart from actively deciding not to read what had been written before, the students may have felt at a loss to add any new content to the feedback already provided and decided to make a post simply to meet the course requirements and/or to avoid embarrassment. Indeed, teachers need to intercept such practices which can ultimately be conducive to student frustration. One way to go about it is to assign students specific criteria to apply when writing their peer feedback. Another is to encourage group solidarity by urging students to be sensitive to the others' personal online space and advising them against monopolising the online environment when providing peer feedback.

6. Conclusion

With this small scale investigation I attempted to shed some preliminary light on the consequences of peer feedback going public, with teacher redundancy and direct and indirect feedback instruction being some of its most notable effects. More research is needed to ascertain the reasons behind such online behaviours as redundancy in peer feedback. I hope to have outlined some of the intricacies of online peer feedback, which can prepare the ground for comprehensive and systematic approaches to investigating academic writing online instruction in the tertiary context in the Republic of Macedonia.

References

Hedge, T. (2014). *Teaching and learning in the language classroom*. Oxford: Oxford University Press.

Holec, H. (1981). *Autonomy in foreign language learning*. Oxford: Pergamon.

Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse*. Cambridge: Cambridge University Press.

Mitra, S. (2005). Self-organising systems for mass computer literacy: Findings from the 'hole in the wall' experiments.In *International Journal of Development Issues, 4 (1),* 71-81.

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*. *59 (1),* 23-30.

Salmon, G. (2003). *E-moderating: The key to teaching and learning online*. London: Taylor and Francis Books.

WebWise Team (2012, October 10). *What is netiquette? Webwise: Making the most of being online*. Retrieved from http://www.bbc.co.uk/webwise/guides/about-netiquette.

Teaching Academic Writing to EFL Students: A Process Approach

Bisera Kostadinovska, Irina Petrovska (University St. Kliment Ohridski–Bitola)

Abstract

One of the first things that EFL students are being told is that they have to learn to communicate well in the target language: to understand each other, to be able to communicate among each other, read what they have written and to write. The basic language skills reading, listening and speaking are, therefore, the most exploited ones.

But what is as important as developing these three skills, is learning how to communicate with the other person when he/she is not present. This is where emphasis on writing is being put. Needless to say, teaching writing with the sole purpose of communication is not what learning to write in the target language is all about. Learning how to write appropriately in the target language also helps students to strengthen their grammar, they can "play" with the language, "step out of their comfort zone" and the effort to express ideas and the constant use of the hand and brain is a unique way to reinforce learning. When students are engaged in writing, they often discover something new to write about or a new way of expressing their ideas - they come "face-to-face" with the real need to find the right word and the right sentence. The relationship between thinking and writing is what makes writing a valuable part in any language course. The methodology for teaching writing is not a clear-cut one and most of the approaches for teaching writing overlaps. All of the approaches for teaching writing have a different focus, the sentence, the paragraph, the grammar – syntax, communication and the process approach. In this paper, the focus will be put on the Academic Writing classroom where the Process Approach was being used with 1st and 2nd year undergraduate students. This approach focuses on the process of writing rather than the written final products.

1. Introduction

Teaching, as well as learning a foreign language has to encompass all of the relevant skills in order to obtain beneficial results. Teaching the English language as a foreign language is being done according to the already established teaching methodology. The four language skills are said to be equally important for students to be able to become a fluent speaker/user of the target language he/ she is learning. Reading, writing, speaking and listening are the four language skills that must be present in the teaching of English as a foreign language. If we take a look at the structure of the resources that are being used for this purpose, we will see that the above mentioned skills are involved equally in both teaching and learning resources.

When students are being taught a foreign/second language, they are taught to communicate with other people, that is, to understand them, to be able to exchange ideas, to convey meaning, to read what they have written and to write to them. An integral part of the proper functioning in a new cultural setting is learning how to communicate with people that are not in the immediate surroundings.

But learning how to write properly has many other aspects that are worth the effort. First of all, writing reinforces the grammatical structures, vocabulary that the students have learned. When writing, the writer has the liberty to express herself/himself while being very adventurous with the language. Third, when students write, they become very involved with the new language; the constant thinking how to express a certain idea is a great way to reinforce learning.

What has to be distinguished are the differences between speaking and writing which has to be done in order to justify the importance of teaching writing:

- Speaking is universal and inherent; everyone acquires a native language in the earliest stages of their lives but not everyone learns to read and write;
- The spoken language has dialect variations, whereas the written language demands following the standards of grammar, syntax and vocabulary;
- Speakers use their voices and other non-verbal "means" of communication to convey meaning and express their ideas. Writers rely on their words to do the same;
- Speakers use pauses and intonation, whereas writers use punctuation;
- Speakers pronounce, whereas writers spell; and, most importantly
- Speaking is spontaneous, while writing is a planned activity.

Because of the last mentioned reason, in the process of learning a foreign language, writing has to be present equally as the other three language skills and the emphasis should be put on the final product as well as on the process.

But, teaching writing did not receive the needed and appropriate attention until the second half of the 20th century. Harmer points out that the writing skill has finally been recognized as an important skill for language learning. He lays stress on the essentiality of the writing skill saying "The reasons for teaching writing to students of English as a foreign language include reinforcement, language development, learning style, and most importantly, writing as a skill in its own right" (Harmer 1998: 79). Writing is indeed becoming necessary not only in L2 in school settings, but also in our daily life, particularly owing to the prevalence of information technology, such as writing e-mails, or overseas business letters. The reason that it was not emphasized so far is that students' writing was only looked at as a ready-made product; students were being taught that the only thing that was going to be evaluated was the final product and not the process or the acquisition and mastering certain skills that they needed in order to have a review-ready written product.

Having this in consideration, through extensive research, a number of approaches and techniques have been provided regarding EFL or L2 writing during the last decade. Unfortunately, it is not easy to introduce these techniques directly into ESL classes. In particular, the process approach, which several researchers began to emphasize as the key in L2 writing, has been influential in the classroom. Until then, the focus of writing instruction was the written product. The process approach, on the other hand, focuses on the process of writing (Onozawa 2010: 154).

2. What is the process writing approach?

The Process Writing approach is an approach to writing, which focuses on the process by which learners produce their text rather than on the produced texts themselves. In the end, learners surely need to and are required to complete their products, yet the writing process itself is stressed more. By focusing on the writing process, learners come to understand themselves more, and find how to work through the writing. They may explore what strategies conform to their style of learning. Brown (2001: 336) states that writing is a thinking process; a writer produces a final written product based on their thinking after the writer goes through the thinking process. In addition, Brown quotes Elbow (1973) as saying that writing should be thought of as an organic, developmental process not as a way to transmit a message but as a way to grow and cook a message (Brown 1973: 14-16). Kroll (1990) also quotes Applebee (1986) as saying that the process approach “provided a way to think about writing in terms of what the writer does (planning, revising, and the like) instead of in terms of what the final product looks like (patterns of organization, spelling, and grammar)” (Kroll 1990: 96).

In recent times, ESL writing moved from a language-based approach to the process approach. It is not clear what brought the process approach to ESL. Reid (2001) claims that it arose for two reasons: researchers’ recognition of the newly developing field of English native speaker composition and teachers’ realization of the needs of English L2 students in the academic environment. (15) ESL writing classes were grammar-oriented up until the second half of the last century. Since then, various approaches and suggestions have been developed through laborious studies and research. More recently, some researchers have presented the post-process approach for L2 writing (Atkinson 2003; Matsuda 2003), which adds more social dimensions to writers (Fujieda 2006: 68), but the process approach seems to remain the preferred and approved approach (Onozawa 2010: 156).

This writing approach is considered to be very effective. First, in the product approach, the focus is on the end result of the learning process, and the learner is expected to perform as a fluent and competent user of the language. The process approach, in contrast, stresses the process that writers go through in composing texts (Nunan 1991). Brown (2001: 335) states that [in the product-oriented approach] a great deal of attention was placed on “model“ compositions that students would emulate and how well a student’s final product measured up against a list of criteria that included content, organization, vocabulary use, grammatical use, and mechanical considerations, such as spelling and punctuation. That is, students convey their messages to the readers in written form through the complex writing process; prewriting, drafting, revising, and editing. (Onozawa 2010: 157).

In the Process approach, students do not write on a given topic in a restricted time and hand in the composition; rather, they explore a topic through writing. The latter one is made by using many techniques that develop the writing process- prewriting techniques, writing and post- writing techniques that help the students develop their ideas, explore new and abandon redundant ideas for their own piece of writing. This process also gives them the sense of liberty in their own composition- they determine which way they will guide their written piece.

The Process Approach gives students the two most crucial supports: time and feedback. The feedback is twofold: the one from the teacher and the one from the peers. This approach to writing steps away from the Audio-Lingual Method- which was the only method implemented in teaching writing- and the controlled writing. It also fosters cooperative learning.

For the purpose of this research, an emphasis has been put on the pre-writing techniques as part of the implementation of the Process Approach in an EFL classroom.

Developing pre-writing skills is essential to every piece of writing the "writer" does. While the end result is also important, achieving the skill of an efficient writing process will help the student become a skillful writer. Optimal pre-writing strategies eliminate confusion and minimize writer's block while actually writing. Prewriting activities can be employed later on in every stage in the writing process! The "free time" at the starting point of the writing process is essential.

Prewriting refers to exercises that help learners start thinking about their writing task by measuring what they know, identifying new ideas, and indicating areas requiring further research. It is a way of putting critical thinking into action. Prewriting can help sharpen learner's skills of observation and evaluation. Prewriting can help save time by quickly determining which ideas are worth developing and which should be discarded. Prewriting puts critical thinking into action.

During pre-writing, many students find it helpful to create lists as a way to generate topics that they might write about. Students should write down possible topics. It may be easier for students to number or double space between each idea so that they can go back and revisit their ideas. This allows them to easily add to these ideas and hopefully form a coherent topic. Students can also free write; after which they can discuss their possible topics with their peers. Through this process, students can get feedback about their writing and constructive criticism that will help them in their finished work. Once the students have chosen a topic the next step of an efficient pre-writing process would be free writing.

3. Method and research

The research conducted for the purposes of the paper was made with first year students and second year students in their undergraduate studies (the second year students attend the Academic Writing class for the first time) at the Faculty of Education in Bitola, Macedonia, who are students of English as a Foreign Language at the department for English Language professors. The students were at beginner stages of Academic Writing but B2 level of English Language knowledge. The research was focused on developing pre-writing strategies in the academic writing process. The introduced techniques were the following:

(1) Activity “tell and Transform”- Employing knowledge telling vs. knowledge transforming
(2) “Silent Brainstorming”
(3) Listing
(4) Brainstorming
(5) Mapping (also known as mind mapping)
(6) Clustering
(7) Free writing
(8) Questioning

But for the sake of the research only two were chosen for further implementation and research:

(1) Activity “tell and Transform”- Employing knowledge telling vs. knowledge transforming
(2) “Silent Brainstorming”.

The methods used for the collection of data for the research were the following: in the first phase, teaching and practice, in the second phase: monitoring, asking questions, observing students’ work and interviews with the students.

3.1. Activity “tell and transform”- employing knowledge telling vs. knowledge transforming

Knowledge telling vs. *knowledge transforming;* the teacher gives the topic; in an unstructured way during the telling-all-you-know phase.

Knowledge telling means that students just have to write down everything that comes into their mind about a certain topic. This is probably the only approach that is applicable to beginners’ stages of teaching and learning academic writing. Knowledge transforming means that basic knowledge is developed, shaped and organized so that the student’s own structure is imposed on the material (White 1987: 11). What is important to mention here is that several attempts might be needed to go from the first phase into the second, or to go from unstructured knowledge to an acceptable draft that transforms into a concept.

3.2. Silent Brainstorming

This activity helps students to generate their own ideas for topics that they have chosen but need further thinking on the topic. This technique fosters cooperative learning and peer input analogous to brainstorming: its freedom, creativity and lack of inhibition are of great value in prewriting. By doing this activity, every student gets an input from his/her peers and every student contributes in an intensive way.

In this activity, each student gets a card on which he/she writes the topic they are planning on writing about, and a brief statement about any aspect of the topic that might be developed in their essay. After each student has done this, the process of silent brainstorming begins. The teacher shuffles the cards, and hands them out to the students in the class. The teacher only serves as a facilitator in this process. Each student has to write a statement about the topic they see on the

card. The activity is finished when there are enough comments on each card which at the end go back to their originators.
It is important to tell the students that the teacher will not read or grade these cards, so that they are ensured that they are giving their input to their peers and not the teacher so that they would feel more inhibited.

3.3. Analysis

During research, both techniques proved to be easily explicable and accessible by the students. After each lesson where these techniques were used, students were asked for a feedback about the Academic writing lessons and the techniques used. Students were asked if they thought whether each of the techniques was helpful and whether they thought they can use it further on in the writing process. It was important to find out whether they thought it was better to be "evaluated" by the teacher or by their peers.

Separately, about *knowledge telling* and *knowledge transforming* activities students were asked whether it was easy for them to go from phase I to phase II or whether they needed further practice. It was important to have feedback for this particular point about the usage of this technique because it is considered to be a more elaborated and more demanding pre-writing technique.

These two techniques were used and researched on from two main points of view: the one was being evaluated by the teacher and the other one offered more independence for the students. Also, the *knowledge telling and knowledge transforming* technique required more attempts to be made by both the students and the teacher in order for the technique to be "mastered", while the *silent brainstorming* technique was simpler to use and required no further explanation.
Here are the answers that students gave when asked about these two techniques:

- 50% of the students said it is helpful, but they need help from the teacher
- 30% said that they can re-use this technique in every stage in the writing process
- 20% said it is hard to go from phase I into phase II

Figure 1.1. Knowledge telling and knowledge transforming

- 80% of the students found it helpful, useful, and possible to use it more than once. The fact that it is not evaluated by the teacher, made them write more comfortable.
- 15% thought that they only needed two or three comments

Figure 1.2. Silent brainstorming

Asked whether they would use these techniques again during the process of writing, the students provided the following answers:

- Strongly Agree: 55%
- Agree: 18%
- Neutral: 27%
- Disagree: 0
- Strongly Disagree: 0

Figure 2.1. Knowledge telling and knowledge transforming

- Strongly Agree: 75%
- Agree: 15%
- Neutral: 10%
- Disagree: 0
- Strongly Disagree: 0

Figure 2.2. Silent brainstorming

4. Discussion

From the conducted research, it can be seen that the pre-writing techniques are of the utmost importance when dealing with students who enter the Academic writing class for the first time or are in their beginning stages. The Process Approach allows close monitoring of the process that students go through when writing and until they hand in their final draft. As it was previously mentioned, this approach to teaching writing gives students time to do their work. These techniques that were used in the writing classes allowed the teacher to monitor students' work without obliging them to hand in their work and without pressuring them with evaluation.

From a teacher's perspective, the *Knowledge Telling* and *Knowledge transforming* techniques proved to be a more advanced technique and a technique that requires more practice but which is very useful for both students and teachers.

The *Silent Brainstorming* was said to be an activity that proved to be very productive since the teacher is only a mediator; the students felt more independent and responsible for their own writing; the decision making is left to them. It takes a lot of time and effort to write, and so it is only fair that student writing is responded to appropriately. Positive comments can help build student confidence and create a good feeling for the next writing class. It also helps if the reader is more than just the teacher, but peers also, which only helps to make the writer (the student) more involved and engaged in his/her piece of writing.

5. Conclusion

The main idea of this research conducted with EFL students at the Faculty of Education in Bitola, R. of Macedonia was to show the importance of the pre-writing process in Academic writing. As it was previously mentioned, this was made within the framework of the Process Approach to teaching writing. This research has explored this approach from the perspective of a teacher experimenting with it in the field of incorporating reusable pre-writing techniques that can be used throughout the entire writing process.

As it can be seen from the results from this research and the feedback from the students, they all agreed that the implemented techniques, *Knowledge Telling* and *Knowledge Transforming* and *The Silent Brainstorming* can be used over and over again in the writing process whenever the need arises. These two techniques also foster cooperative learning, which can be said that it has a social dimension.

This Process Approach is focused on the process of writing, but that does not mean that it neglects the final product; furthermore, it shows that the evaluation process in the Academic Writing should encompass both the process and the final product. We hope that EFL teachers benefit from this research and the results achieved and that they will become more interested in this approach to teaching writing in their classrooms.

References

Applebee, A. N. (1986). Problems in process approaches: Toward a reconceptualization of process instruction. In A. R. Petrosky and D. Bartholomae (Eds.), *The teaching of writing*. Chicago, Ill: National Society for the study of Education, 95-113.

Atkinson, D. (2003). L2 writing in the post-process era: Introduction. *Journal of Second Language Writing, 12,* 3-15.

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy.* NY: Pearson Longman.

Elbow, P. (1973). *Writing without teachers.* New York: Oxford University Press.

Fujieda, Y. (2006). A brief history sketch of second language writing studies: A retrospective. *Kyoai Gakuen Maebashi Kokusai Daigaku Ronsyuu, 6,* 59-72.

Gatz, M. (2004). *Pre-writing strategies* (unpublished paper). Illinois State University, Normal, Illinois, USA.

Harmer, J. (1998). *How To Teach English*. Harlow: Longman.

Kroll, B. (1990). *Second language writing: research insights for the classroom.* Cambridge: Cambridge University Press.

McKensie, L. & E. Tomkins, E. G., (1984). Evaluating students' writing: A process approach. *Journal of Teaching Writing, 3 (2),* 201-212.

Matsuda, P. K. (2003). Second language writing in the twentieth century: A situated history perspective. In B. Kroll (Ed.) *Exploring the dynamics of second language writing.* Cambridge: Cambridge University Press, 15-34.

Nunan, D. (1991). Communicative tasks and the language curriculum. *TESOL Quarterly 25/2,* 279-295.

Onozawa, C. (2010). A study of the process writing approach: a suggestion for an eclectic writing approach. *Proceedings of Kyoai* Gakuen *College, Japan, 10,* 153-163.

Reid, J. (2001). Writing. In R. Carter & D Nunan (eds.). *The Cambridge guide to teaching English to speakers of other languages.* Cambridge: Cambridge University Press, 28–33.

VanderPyl, G., (2012). *The process approach as writing instruction in EFL (English as a Foreign Language) classrooms* (unpublished Master's thesis). SIT Graduate Institute, Washington, Washington DC, USA.

White, L. (1987). Against Comprehensible Input: The Input Hypothesis and the Development of Second-Language Competence. *Applied Linguistics, 8,* 95-110.

Attaining *Academic Literacy*: University Language Instruction Reforms for Polish Doctoral Students

Magdalena Rogozińska (Jagiellonian University Kraków)

Abstract

This contribution presents an outline of expectations of doctoral language instruction in Poland in accordance with European requirements, Polish national laws together with individual university-level decisions. Focus is on the instruction of modern languages, in particular English, and demonstrates the discord between government and university language requirements from doctoral students and the actual instruction they obtain. It shows there is no set curriculum in this area; academic teachers, who are predominantly not practicing academics are allowed to use their own intuition and free will to choose materials for their classes or even entirely exclude the academic component from their language curriculum. *Academic Literacy* goes beyond the written word and some skills such as library skills, editing skills, public speaking, awareness of publishing processes, writing applications for research grants/scholarships, or even forging liaisons with international academic associates may exceed the capabilities of a single language instructor and should be offered in collaboration with other competent practitioners.

1. Critical observations by the European Commission concerning higher education in Europe

The European Framework For Research Careers was introduced in 2011 by the European Commission in response to the observation that despite all the efforts and funding towards the Bologna Process, the European research market lacks unity; academia, industry and other sectors do not use their full potential to collaborate with each other (European Commission 2012: 1) The primary role of the Bologna Process has been to modernize higher education, enabling, as a consequence, a better quality of instruction and research by not only encouraging international cooperation, but also stimulating competition. The European Union has strived to improve scientific potential emerging from its member countries in order to merge academic and business sectors by creating a free-movement European research area where scientific knowledge and technology are freely available and competition is enhanced. Simultaneously, the European Union has pledged in its treaties to promote research those activities it perceives to be of value (European Commission 2012: 1). However, to date, the European research market has not become coherent and the division between careers in academia, industry and other sectors has continued to exist. In addition, European academia has remained hermetic. The system seems to be stagnated by early career choices that are almost irreversible and with little flexibility to swap sectors at a later stage. What is more, the market lacks clear and transparent internal career structures which could be equated with other member states. Consequently, it brings about various predicaments for both researchers themselves and their potential employers. On the one hand, business opportunities may be invisible to researchers whilst on the other, employers themselves may not be aware of the spectrum of competences that researchers can bring to their businesses. To address the aforementioned issues, the European Union has attempted to describe and classify the research career in widely understood terms. *The European Framework for Research Careers* has been introduced as a "voluntary transparency instrument" whose main objective is "to make research career structures comparable across employment sectors and countries."(ibid.) Since the Framework is not focused on any particular sector, the proposed descriptors can serve as guidelines to all researchers, irrespective of their employment base and help them examine the various characteristics they may possess.

The idea of creating a framework is not a novelty. Frameworks have already been forged for various sectors at national and institutional levels. What makes this Framework unique is the attempt to go beyond a given sector and provide researchers and business from various fields and countries with unbiased and volitional guidance. The Framework does not impose any particular, centralised set of rules or expectations. The European Union highlights the voluntary nature of the Framework application by its stakeholders. By no means are researchers required to match all the criteria enlisted in the Framework, nor must the sequence be compulsory.

> *First Stage* Researcher. PhD candidates, up to the completion of their PhD
>
> *Recognised Researcher*. Those successfully obtaining a PhD degree but who are not fully independent researchers yet.
>
> *Established Researcher.* Academics who have amassed a significant level of expertise and independence in their field.
>
> *Leading Researcher.* Distinguished specialists who are capable of leading research in their area or field.
>
> Adapted from *The European Framework for Research Careers 2011: 2*

Since this contribution evaluates university language programs mainly at PhD level, the focus is on the First and Second Stage Researcher only. All the other descriptors can be found in the Framework.

The First Stage Researcher (R1) describes individuals and PhD candidates who conduct research under supervision in either research institutes, universities or industry. The desired competences such people should acquire involve developing language, communication and environment skills, with a particular emphasis on international context and involve both hard and soft. Hard skills would be the ability to undertake supervised research, proven excellent understanding of the field of study together with a good working knowledge of producing necessary data under supervision. Not only should PhD candidates have the proficiency to glean knowledge, they should also demonstrate the aptitude to convey this knowledge to other academics and stake holders. Finally, the Framework stresses the necessity for competent critical analysis, evaluation and synthesis of new and complex ideas in conjunction with the drive to expand knowledge of research methodologies and discipline.

The Recognised Researcher (R2) encompasses both specialists who hold a doctoral degree and those researchers who, despite the lack of a PhD degree, have corresponding experience and competence in a given field. To achieve the fundamental difference in the desired competences for this stage is the necessity to acquire specialised knowledge by going beyond ones own field of expertise to other employment sectors, industries and communities by way of comprehending the agenda of industry and other related employment sectors, fully understanding the usefulness of one's research outcome to the relevant stakeholders together with communicating and promoting science. Added to this, *Second Stage Researchers* are able to give assistance to *First Stage Researchers* to achieve success in their research and development.

The necessary competences for this stage include all competences mentioned in the First Stage Researcher including a wide array of new skills. Apart from a proven track record of expertise in an area of study, special focus is placed on research. The researcher must be able to use the needed skills to create, design and carry out research in their chosen field. Originality that goes beyond what is already known is a key aspect of the work. The uniqueness of such research will result in registering a patent or a peer reviewed publication on a national or international scale. Moreover, the researcher undertakes an active role in academia by participating or running various workshops, conferences and contributing to the field literature. Liaising with peers is another feature of the *Second StageResearcher*. They not only critically analyse, evaluate and synthesise new and complex ideas, but also make their work available and to explain the results of their research, thus, indicating its value to the research community. Finally, the researcher manages their own career path that demonstrates a clear and realistic outlook on future career aims and steers their career in such a way so as to enhance employability.

The primary aim of forming this framework is to acquaint researchers and their potential employers with some guidelines which could have substantial practical purposes for various stakeholders. It can assist researchers with finding themselves on the market and identifying the career paths they would like to opt for, be it in academia or industry. Understanding what is required from them in academia on a global scale and what this career entails is where the Framework can assist. It can stimulate more cross border activity and establishing transparent and bias free guidelines for employers concerning employing staff, promotion, training and remuneration. Public authorities may gain from the Framework by being advised how many researchers should be trained locally to meet their regional and national research and development targets. If needed, the Framework might be used by policy makers as an international comparison or benchmark for the research population. Finally, society may better understand and therefore value researchers' achievements and it can have valuable input in creating a single market for knowledge, research and innovation.

Adapted from *The European Framework for Research Careers 2011: 2-3*

2. Polish academia

2.1. Doctorial studies and forms of assessment

According to government statistics from 2013[1], there are 38,496 people who inhabit Poland. There are 474 universities and schools of higher education in Poland including 140 public schools and 334 private.This number comprises of 42 public universities and 2 private run by the Catholic Church. The number of research units in public institutions is 778 and 665 in private. Additionally, there are 119 research institutes spread all over Poland the biggest of which is the Polish Academy of Sciences with 70 institutes. Humanities subjects comprise of 21.3%, Medical subjects including health and physical education 16.5%, Environmental Sciences 6.3%, Agriculture, Forestry and Veterinary subjects 6.9%, Social Sciences 16.3%, Science 9.3%, Technical 9.3% and Art 3.8%. Altogether the number of academic staff totals 76069, including 10,652 professors, 14,272 habilitated doctors and 51,145 doctors. As far as students are concerned, there were 1,764,060 registered students in the academic year 2011-2012 and the astonishing number of 40,263 PhD students, including 29,943 full time, 10,320 part-time. Poland spent approximately 8.77% of its public budget on science in 2012 (Nauka w Polsce 2013: 7-37).

PhD studies in Poland are regulated by the government that adheres to the higher education law from 27 July, 2005.[2] The duration of the studies can last from 2 to 4 years. PhD students who study full time at public universities are exempt from tuition fees. Before students are allowed to register for the PhD title conferment procedure, they are required to successfully complete their courses and have produced at least one publication. Those who are exempt from tuition fees also have to undertake and complete teaching practice for their university.

The PhD student's achievement record (Karta Osiągnięć Doktoranta[3]) is a document which illustrates every PhD student's activity in the following academic dimensions: the effectiveness of university doctoral instruction together with progress with the PhD thesis, scientific, didactic and organizational involvement in a given academic year. For every achievement the student accrues a set number of points. These points will be used to create ranking lists for scholarships, grants and other benefits. Students are obliged to submit the record by 30 June. However, should the student take part in an activity considered to be of merit by the university, this has to be submitted by no later than 30 September. Therefore, within the Polish system, PhD students will need to complete four such forms. Although there might be slight deviations in the document itself amongst universities, the core criteria mentioned below consistently appear on the list.

---

[1] Polish Main Statistical Office:
http://stat.gov.pl/cps/rde/xbcr/gus/L_podst_inf_o_rozwoju_dem_pl_do_2013.pdf
[2] http://isip.sejm.gov.pl/DetailsServlet?id=WDU20051641365
[3] http://www.inp.uj.edu.pl/documents/3905854/5711193/Karta_osi%C4%85gni%C4%99%C4%87_doktoranta_18.03.2014.pdf

2.2. Progress with the PhD thesis

Successful graduation from the four-year PhD program is not sufficient to obtain the title of PhD in Poland. Doctoral students need to start the formal procedure which will eventually lead to PhD thesis completion and defense after which they are granted the title. However, recent government regulations have imposed a requirement of a minimum of one publication prior to registration as a prerequisite of starting the procedure. The language of the publication may be Polish or a modern language.

The ability to obtain grants is another criterion that appears on The PhD student's achievement record. PhD students are expected to continually enhance their research activities also through external funding. Grants are divided into two categories. The first is called *big grants* and includes grants which are seen by the Ministry of Science and Higher Education as more valuable and they are allocated by institutions such as the National Science Centre, the European Economic Area (EEA) and Norway Grants, Scientific Exchange Program (SCIEX), Scholarships "START" and "DOCTUS". The latter category is called *small grants* and involves being given financial support for selected research activities such as conference attendance, summer school participation or national scholarships deemed to be of lesser value, e.g. the university scholarship, Krakow City scholarship.

Obtaining international grants has more points allocated for PhD students. They can attempt to apply individually or in collaboration with other scientists in which case the number of points is divided by the number of participants in the project.

In addition, The International Research period is valued higher than any national research period. PhD candidates are therefore encouraged to apply internationally and build their knowledge base from both national instructions alongside international. Participation in a summer/winter school or a training whose program was for no less than one week whilst directly related to the subject matter of the PhD, is also taken into consideration bearing in mind that any international journey will result in gaining more points for the doctoral student.

Finally, the PhD supervisor evaluates the PhD candidate's knowledge and performance and grades them accordingly.

2.3. Ratings of publications

Publishing is part and parcel of existing in academia. Doctoral students are fully cognizant of the popular saying "publish or perish". Nevertheless, despite immense efforts from budding academics to appear in peer-reviewed journals, not every publication is rated equally. The Ministry of Science and Higher Education has strived for parameterization of science and has introduced the following categorisation of research journals.

According to the guidelines that concern only humanities and social sciences implemented by the Ministry on 13 July 2012, publications are divided into the three sections[4]. The first one comprises publications in scientific journals which are included in the Journal Citation Reports (JCR) data base and which contain the Impact Factor (IF). 10804 scientific journals meet these criteria and have been allocated from 15-50 points depending on the journal. The second part enlists publications in scientific journals which do not possess the Impact Factor (IF). 2,623 scientific journals matching the above criteria are on the list and publishing in them can bring from 1 to 10 points. Finally, the last part mentions publications in scientific journals which are in the European Reference Index for the Humanities (ERIH) base. 4,262 scientific journals are on this list and are valued at 10 points.

Each of the journals from the Government's List has a given number of points. The authors can access the list and see the number of points the journal has to offer prior to deciding to submit their work for publication. It is worth mentioning that all types of publications written in a foreign modern language are scored higher than those written in Polish irrespective of the quality of the contribution.

2.4. Controversies concerning parametrisation of publications

It needs to be pointed out that this assessment applies not only to doctorial students but also practicing academics. The Ministry of Science and Higher Education claims that parameterization must be used to assess entire faculties and research units collectively and not individually. In practice, many institutions use the criteria and the amassed points as an argument for promotion or additional benefits an academic may receive such as research funding (Ministry of Science and Higher Education 2014).

Recently, such a parametric approach to assessment has faced a wave of criticism from academics. Some of the objections concern the criteria that tend to be frequently changed by the government. Some of the benchmarks were established after the implementation of this document. Moreover, academics disagree with some of the decisions concerning the classifications of journals e.g. the decision to allocate more points for any article written in English is highly controversial (Czubaj 2011, Kulczycki 2014). The pure language of publication does not necessarily prove its content quality. Publishing in selected small but widely known journals have a lower point value purely due to their size, disregarding their excellence. Furthermore, academics who conduct their research in collaboration are assessed worse than those who do it individually. The aforementioned arguments are the mere examples of mounting criticism towards strict parameterization of Polish scientific achievements.

[4] http://www.nauka.gov.pl/komunikaty/komunikat-ministra-nauki-i-szkolnictwa-wyzszego-w-sprawie-wykazu-czasopism-naukowych.html

2.5. Other forms of PhD student assessment

Those PhD students who undertake their studies full-time and free are obligated to conduct, depending on availability, up to 90 teaching hours of instruction in their field of study.This didactic involvement is rewarded by a number of points the university has allocated. Moreover, it is natural that good grades are essential in meeting the requirements for a scholarship. In the Polish system grades are allocated from 2.00, being the lowest, to 5.00 the top mark. It also possible that a student can be awarded ½ a point.

Participation in conferences is also an indespensible element of working in academia and in this field the assessment is based on the location of the conference (national versus international) and the conference language (Polish versus a foreign language). This division has caused an additional wave of criticism since the scoring criteria is rigid and assumes in advance that foreign-based conferences are always of higher value solely because they are held abroad and exclude content related excellence.

Finally, The Ministry of Science and Higher Education acknowledges that the nature of work in academia does not limit itself with doing research and publishing. Active involvement in the organizational aspect of university is also considered of significant importance. As a result, it is expected that doctoral students are actively involved in university organizational activities such as administrative help, assistance with recruitment, organizing national and international conferences, exhibitions, workshops, educational projects; help with the university website, staff hours in the library or IT room, coordinating various projects and involvement in PhD student associations. Interestingly, should any activity done in a foreign language, the students will garner more appreciation from the points system.

3. How are foreign languages taught in Poland? A case study from the Jagiellonian Language Centre, Cracow.

The Jagiellonian Language Centre was founded on 1 October 2005 and is an inter-faculty unit of the Jagiellonian University in Cracow whose primary aim is to teach languages and support students with the complex process of acquiring a foreign language. It offers courses to university students in modern languages such as English, German, Spanish, Italian, Russian, French as well as Lithuanian, and Latin. All the above mentioned languages together with Chinese, Norwegian, Esperanto, Japanese and Catalonian are on offer in commercial courses open to students, university employees and external learners. Currently, approximately 140 language instructors are employed by the unit, the majority of whom teach at the university. The unit cares for mainstream students together with students with various disabilities by providing them with specialised methodology and equipment.

3.1. Organisation

The languages are divided into the following language sections: English, Latin, German, Romance Languages (Spanish, Italian) and Russian. Needles to say, due to its huge popularity, English is the most widely taught language at the University and English is the largest section. Each section has its own coordinator and all the sections report to the Director of Studies and the Executive Director. The regulations implemented by the unit come from the Ministry of Science and Higher Education, Jagiellonian University laws as well as internal decisions within the unit.

The Ministry of Science and Higher Education in Poland imposes an obligation on higher education students to master a foreign language competency, which is no lower than B2 level, according to CEFR, for the first cycle of study and no lower than B2+ for the second cycle. Since CEFR does not include B2+ level in its original reference and the Ministry does not require C1 level, challenges arise concerning interpretation of this level by universities and educational institutions and confusion with establishing language syllabus/curriculum. However, it should be stressed that acquiring language mastery does not equal obligatory course participation for those students who possess a suitable international certificate which can exempt them from taking classes. Moreover, students who do not wish to participate in the courses can take a university exam at the end of an academic year and if they pass, the language requirement will be met. Some students opt to attend private language schools or private tuition and take this opportunity to dispense with university teaching. The minimal number of hours is 120 teaching hours in the first study cycle, 60 in the second cycle that total 180 minimum. However, where a program not divided into cycles, following the Bologna Process implementation, has to last five consecutive years and again 180 hours minimum. Successful completion of each module results in obtaining a given number of ECTS points: 4 ECTS for the module in BA studies, 2 ECTS for MA studies and 6 ECTS is the program consisted of five years. The Jagiellonian Language Centre additionally issues its own certificate of language competence. What is more, MA students and students in a five-year program are obliged to undertake one course in a foreign language, not related to language classes run by the Jagiellonian Language Centre. All language courses start in the second year.

3.2. Didactic approach of the Centre

The teachers have considerable autonomy in planning and running their language courses, especially for MA programs. They need to prepare their students for the final examinations thus, certain criteria remain yet how they do it is of their volition. The Centre firmly believes that the best method of achieving its objectives is task-based teaching and learning. The didacting approach used by the Jagiellonian Language Centre places the student in the centre of the didactic process with the teacher serving a mentoring function. The student, on the other hand, takes a role of an expert and a partner in their own field and plays an active role in enriching their language competence with the assistance from their teacher. The content of teaching shall be useful for the students and lead to acquiring and/or developing intended competences and strategies. It is of paramount importance that the language instruction is adequately adjusted to students' capabilities and language competences. It should not be too easy nor challenging which will force the language instructor to constantly monitor their students' progress. If needed, it ought to be individualized and followed by a reflection on the learning process.

Teaching and learning is perceived as an individual endeavor and, in the context of higher education, shall not be treated like a mere language service. The Centre has applied the Post Method Pedagogy Model (Kumaravadivelu 2001: 537-538), which propagates no prescriptive dominant method of teaching and emphasizes individual and social dimension of learning. The learning process is deemed as an adjustment period based on personalized experience and action (Dewey 1967) or socio-cultural approach to learning propagated by Vygotsky (1978), in which learning is understood as an integration process that takes places in the sphere of cognitive development (Stawicka 2014: 128). Such an approach places teachers in a new role. They cease to have a dominant role in the didactic process and are instead expected to apply a frontal position of a learning coordinator and/or facilitator assisting students to gain learning objectives by means of their own decisions and actions (Stawicka 2014: 129). The Centre's intention is to provide its students with sufficient language abilities to communicate freely in foreign languages in private life, university and business contexts. They should also be adept at using source books and authentic materials. Although mastering grammar, vocabulary and pronunciation is considered to be vital in foreign language acquisition, the emphasis during instruction is put on communication. The Centre has decided to create groups with no more than 16 students to enable free communication. Teachers have plenty of multimedia equipment at their disposal and a university-run internet platform called *Pegaz*[5]*,* where they can submit additional resources, design multimedia quizzes, tests, uploads videos or run forum discussions. E-learning is gaining popularity and the teachers are increasingly expected to deliver some of their language instruction online. Also according to the laws imposed by the Ministry of Science and Higher Education, no more than 20% of the total amount of hours can be delivered using this method. The online instruction is widely used by teachers especially before or after public holidays when the attendance record tends to be poor. The cover for teachers on sick leave is also partially organised in the distance learning mode.

In MA programs an indispensable component of every language course is the vocabulary related to the field of study undertaken by students. The teachers are instructed to use authentic materials which reflect the area of expertise of their students. Teachers running courses for MA students do not normally use traditional course books but are expected to carry out a needs analysis in their

[5] http://pegaz.uj.edu.pl/

groups and adjust their curriculum accordingly. What is more, at present the Centre is cooperating with all university faculties, requesting them to nominate a person who could inform the teachers about faculty needs and provide assistance with selecting the best resources for their MA students. Other compulsory elements, treated very seriously by the Centre, are academic English and presentations. Academic English encompasses various forms of writing. Students are obliged to write a minimum of one written essay per term. Every level has a recommended list of writing types with the final choice and sequence left to the teacher. Examples include a letter of application, essays, reports for the higher C1 level, discursive letters (opinion, for and against, suggesting a solution to a problem,) for B2 level, formal emails, essays for B1 level and finally, informal emails/letter, internet comments for beginner levels A1/A2. Other advisable types include a proposal, an article and a review. In an attempt to increase its standards and meet the students' expectations, the Centre has recently carried out a needs analysis survey in which students were requested to suggest writing forms they need for their studies or future career. Some responses are presented here. Students in the Polish Language Faculty suggested practicing more critical essays, short stories, a screenplay, synopsis, summary and reviews. Students in the History Faculty wanted to study writing biographies, introduction to writing historical essays as well as more specialised writing skills such as presenting primary and secondary sources, writing bibliography in various styles, constructing a thesis - to name but a few. Students in the Science Faculty need skills to write a laboratory report, summary, research project/project description, critical literature review, and thorough understanding of a scientific publication. Students irrespective of the faculty feel the need to learn describing visual information, summaries of a BA/MA thesis and more practice on paraphrasing. It has to be noted that many of these forms or writing are practiced by teachers but it appears not sufficiently.

Another aspect of language communication that the Centre deems of paramount importance is rhetorical functions. Regardless of the type of texts taught in the courses, it is recommended that teachers introduce the following list: compare, describe, explain, express opinions, justify, persuade, recommend and/or suggest, hypothesise, advise and evaluate.

Finally, the Jagiellonian Language Centre believes that being able to prepare and deliver a professional scientific presentation is an indispensable element of academia that every single university student needs to master. As a result, students are obliged to prepare a presentation which is relevant to their field of study a minimum of once per academic term. Prior to students' presentations, exhaustive preparation is given which includes discourse management (structuring a presentation with an appropriate introduction, plan, main body, summary and conclusion), presentation skills (knowing your audience, eye-contact, body language), the language of presentation (formal terminology, dealing with visual aids, complex questions) and general impression such as language fluency, pronunciation and voice projection. The presentation has to be based on source or at least authentic materials and a concise summary of it has to be submitted to the teacher. Occasionally students prepare short exercises whose aim is to check comprehension, explain complicated words or terminology.

3.3. Student assessment, standardisation and accreditation

For numerous years teachers have had only minimal guidelines concerning the form of the final examination. The generally agreed consensus was to test all language skills proportionally and adequately to CERF descriptors and the oral exam had the form of an academic presentation followed by a discussion. That meant every teacher prepared their own exam based on their experience and intuition. However, recently the Centre has resolved to implement the process of standardization in most aspects of its activity and apply for an accreditation from an international organization called EAQUALS[6]. The Centre has observed numerous drawbacks in the old system and set its goal to invite an external, international and reputable organization to carry out a language audit, evaluate its work and offer useful contributions. The application process is due to be completed by the end of 2016. At present the Centre is welcoming the EAQUALS advisory body and is adjusting its operations so as to comply with its standards. One of the crucial criteria is to standardize assessment. Teachers will cease to have so much free will in deciding about the form of the final examination and joint exams for all students in a given faculty will have already been organized.

3.4. Teaching languages at doctoral level

In 2013, there were 2,968 PhD students at the Jagiellonian University and 3,220 when students from Collegium Medicum are included[7]. Doctoral students are eligible to participate in a 60-hour foreign language course. The Ministry of Science and Higher Education in Poland imposes a requirement on every PhD candidate to possess a foreign language competence at the minimum C1 level in CEFR. There are special classes open exclusively to PhD students. The Centre forms separate groups dedicated to students from humanistic and scientific departments. Additionally, the offer extends to modular courses covering topics such as, for the English language courses, business English, academic writing, English for natural sciences, reality reflected in literature and art, for German courses, German in the language of media, legal German and for Russian business Russian.

The doctorial foreign language exam can only be conducted by the most experienced teachers who have the title of *starszy wykładowca* (senior lecturer). Before the exam the Centre offers free consultations and advice on their language

[6] http://eaquals.org/

[7] Based on the University statistics from 31 December 2013.

level as well as the scope and structure of the examination. The candidate is expected to be well versed in the following areas: presenting their own profile with particular emphasis on their scientific interests and achievements, being able to present the topic of their research and overall content of the dissertation, with a special emphasis on the scope of PhD research, research hypothesis and methodology applied in it, results and conclusions. Even though for most of the time, the exam is in the form of a monologue, the candidate shall expect to be requested by the examiner to clarify a concept, idea or explain something in greater detail. The proficiency in using academic discourse is investigated by the candidate's ability to explain, compare, contrast, forecast or conclude to name just a few skills. Finally, the candidate is asked to involve in a conversation concerning their future career plans, any intended or undertaken international coopration, projects, or implications of their research.

The exam lasts approximately 30 minutes and is in the oral form only. The criteria for assessment include the ability to discuss general topics, PhD dissertation, vocabulary and grammar range, pronunciation and intonation. In order to maintain a high level of standardization in assessment, a general annual meeting takes place every year and the exam is held in the presence of Assessment Committee (Komisja Egzaminacyjna).

3.5. Curriculum for doctorial students

Surprisingly, there is no established curriculum for these courses. Every teacher is expected to prepare their own teaching materials and the only requirement is that students must have "an academic component". There is no syllabus but to compensate for it there are annual meetings where teachers can share their peer knowledge and exchange good practice. Teachers' monitoring is limited to students' questionnaires once per term and one observation visit during the class per year from a senior lecturer. Unfortunately, if we look at the profile of the teachers, only a few of them are PhD holders and the remaining have no or limited experience in academia. They are proven, experienced and very dedicated teachers but it does not alter the fact they are not practicing academics and may be ignorant of the skills needed in this profession. The Centre realises this drawback and has recently organized a series of training sessions to acquaint all the staff with newest approaches to teaching and special workshops were devoted to academic components such as academic writing. At doctorial level it is understandable that there are no course books aimed specifically at PhD students. The students present a wide array of backgrounds, having various interests and profiles. Since they are mature students who have voluntarily undertaken their further studies, one can assume that they will reach for the relevant, necessary knowledge themselves without undue teacher encouragement. Therefore, the didactic approach of task-based teaching and learning may serve its purpose in guiding PhD students through their individual language needs. What is more, PhD students are welcome to actively participate in workshops run by

ArsDocendi[8] which is a university based unit whose aim is to improve the didactic skills of academic teachers. The workshops are free and conducted by eminent professors although it has to be stressed that most of them are run in Polish.

Lastly, teachers are encouraged to monitor PhD students' needs and incorporate them into their instruction. They are requested to be on full alert to respond to any new changes that might relate to language competence of their doctorial students.

3.6. Other forms of innovative language instruction

In addition to standard lessons which follow the aforementioned didactic approach, all the teachers are encouraged to create their own tailor-made courses. This provides an alternative to coursebook based instruction addressed to mainly MA and PhD students as well as students who have extra credits. The objective of such "Thematic Courses" (Kursy Tematyczne) is to develop students' interest by means of instruction in the English language. Every semester new courses are developed and those on offer grow in diversity. Some of the courses which are on offer include Legal English, Financial English, Academic Writing, English in the Natural Sicences, Reality Reflected in Art and Literature, English through Short Stories, Words Are Born Every Day, Read and Think Critically and Listen Read and Debate.

4. Conclusions

Based on the information presented in this contribution, there appears to be some misgivings towards the teaching foreign languages policy at PhD level in Poland. BA and MA level students have a wide array of courses to choose from and are adequately taught to pass necessary state exams whereas if we compare the requirements from a PhD students' assessment card and the actual instruction they obtain, there is clearly discord. Lack of a set curriculum will invariably lead to unequal standards of instruction. It appears doctoral students have not received adequate attention to date. The argument that PhD students are deeply self motivated individuals who can learn on their own is valid. However, the university should also take some responsibility for language preparation given to people who might well represent its name in future. Dedicated and experienced teachers are undeniably an asset of the Centre but even in case of native speakers, relying entirely on one's intuition may seriously compromise the quality of teaching (Schmied 2013: 27). As a result, in 2013 the Centre presented a suggestion to the faculties to nominate one representative who would act as a consultant providing the teachers with assistance as to the choice of materials and the exact skills students from every faculty require. The cooperation has

[8] http://www.uj.edu.pl/doktoranci/warsztaty-ars-docendi

commenced the academic year 2014/15 and has already resulted in obtaining valuable feedback that can be incorporated into teaching.

Running classes for doctorial students is a daunting task that may exceed competence of just one instructor. The variety of skills needed that go beyond academic writing imposes a string of challenges and in certain areas, only collaboration with other practicing academics will bring expected results. It has to be pointed out that not every university in Poland has implemented an "academic component" whereas some language centres still teach only general English. In this respect, the Jagiellonian University has made a great effort to critically evaluate its program and introduce desired changes. The modern didactic approach provides a wealth of opportunities to shape the instruction in such a way most needs are addressed. In addition, vigorous self assessment, internal and external, implemented by the Centre provides good ground for evaluation and reflection on its current activities.

References

Czubaj, M. (2011, October 10). List otwarty dr hab. prof. SWPS Mariusza Czubaja do Minister Barbary Kudryckiej. Retrieved from http://www.petycje.pl/petycja/7999/list_otwarty_dr_hab._prof._swps_mariusza_czubaja_do_minister_barbary_kudryckiej.html

Dewey, J. (1967). Wybór pism pedagogicznych. przeł. J. Pieter. Wrocław: Zakład Narodowy im. Ossolińskich

Diagnoza Stanu Szkolnictwa Wyższego w Polsce. (2009, November 1). Retrieved from http://www.nauka.gov.pl/g2/oryginal/2013_05/fa5b19e372e1bed45db817b8380c868.pdf

European Commission, Education and Training (2015, February 2). Retrieved from http://ec.europa.eu/education/policy/higher-education/bologna-process_en.htm

Kulczycki, E. (2014, September 4). Ocena czasopism a parametryzacja humanistyki. Retrieved from http://ekulczycki.pl/warsztat_badacza/ocena-czasopism-a-parametryzacja-humanistyki/

Kumaravadivelu, B. (2001). Toward a postmethod pedagogy.*TESOL Quarterly 35,* 537-560.

Nauka w Polsce. (2013, January 1). Retrieved from http://www.nauka.gov.pl/g2/oryginal/2013_08/010df9372ca1a016e6e9bf7746817b38.pdf

Rozporządzenie Ministra Nauki i Szkilnictwa Wyższego 1) z dnia 13 lipca 2012 r.w sprawie kryteriów i trybu przyznawania kategorii naukowej jednostkom naukowym. (2009, November 1). Retrieved from http://dziennikustaw.gov.pl/

Schmied, J. (2013). English for Academic Purposes: Contrastive Perspectives in the Curriculum. In J. Schmied & C. Haase (eds.). *English for Academic Purposes: Practical and Theoretical Approaches.* Göttingen: Cuvillier, 19-40.

Stawicka, M. (2014, April 3). Dwupodmiotowość w nauczaniu projektowym na lektoratach w szkole wyższej. Retrieved from http://www.jcj.uj.edu.pl/documents/15103613/37328696/konin

Vygotsky, L.S. (1978) *Mind in Society: The Development of Higher Psychological Processes.* Cambridge: Harvard University Press.

Connecting Adverbials in Student Writing

Sven Albrecht (Technische Universität Chemnitz)

Abstract

This study investigates the differences in the use of connecting adverbials, which are grouped into four functional categories, in timed and untimed student writing. Analysing the frequency can reveal different usage patterns, which then can be evaluated in their significance by carrying out chi-square test. The main source of data for analysis is the *ChemCorpus*, containing learner texts written by German native speakers collected at Chemnitz University of Technology. It is then compared to the reference corpus, the *International Corpus of Learner English (ICLE)*, which contains texts written by speakers with different first languages. The analysis will focus on the variables of timedness, gender, native language and prototype to examine what influences the students' use of connecting adverbials. Furthermore, issues that arose during the analysis, especially concerning data compatibility will be discussed.

1. Introduction

Academic writing in English differs from other forms of written language. It has been claimed that writing academically has to be learnt by students of English and non-English speaking backgrounds alike (Schmied 2011: 2). Thus, it is important to analyse expert academic writing and novice academic writing and to compare the results to deduce implications for teaching academic writing. However, as Schmied (2011) and Hüttner (2007) point out, student writing can be seen as apprentice texts to become recognized members of the discourse community and thus is very likely to exhibit differences to expert texts of the genre. Therefore, analysing the student texts and comparing them as their own genre can also provide useful insights on the variables that influence the students' language use. Possible influential factors can be sociolinguistic, such as mother tongue or gender, but also due to other conditions, such as text type or department. Thus, the analysis of student writing can provide useful insight into which of the factors is most influential.

The aim of this study is to investigate possible differences in the use of connecting adverbials by novice writers of academic English depending on whether the time for writing the text is limited or not. Connecting adverbials operate at sentence level and "serve a connective function making the relationship between two units of discourse explicit" (Lenker 2010: 37).

The connecting adverbials that were included in the study as well as the functional categories[1] used to group them have been based on previous studies and

[1]additive: *besides, actually, alternatively, regarding, similarly, likewise, namely, in addition, incidentally, thus, for instance, in other words, on the other hand, for example* adversative: *nevertheless, although, yet, though, but, however, in fact, instead, rather* causal: *because, therefore, consequently, hence, then, in this*

research by Hůlková (2011), Bolton, Nelson and Hung (2002), Milton and Tsang (1993), Biber et. al. (1999) and Halliday & Hasan (1976). Furthermore, the influence of sociolinguistic variables, namely gender and native language (L1), will be analysed. Additionally, the prototypical connecting adverbials of the four functional categories will be compared to the rest of the respective categories. The findings from the main data source, the *ChemCorpus*, will always be compared to the *ICLE* for reference. Moreover, the theoretical framework that forms the basis of this study will be presented, as well as a detailed description of the applied methodology.

1.1. Research questions

There are four major research questions that cover everything that has been analysed in the two corpora. The first three research questions comprise the variables of timedness, gender and mother tongue (L1); the fourth research question deals with the prototypical connectors in the functional categories:

RQ 1: Do L2 learners of English use more connecting adverbials in timed or untimed writing?
RQ 2: Do women use more connecting adverbials than men?
RQ 3: Do German L1 writers use more connecting adverbials than L1 writers of other Germanic languages?
RQ 4: What are the prototypical connectors for each of the four functional cate gories?

The first research question investigates whether more connectors are used in timed or untimed writing. Since there is hardly any research that investigates the influence of time on writing (Gregg, Coleman, Davis & Chalk 2007), it is difficult to predict possible findings. My hypothesis is that, as there is more time available for revision and editing, it is more likely that a higher number of connecting adverbials can be observed in untimed writing.

The second research question deals with the variable of gender. It will be investigated whether female or male students use more connectors. The initial assumption is that female authors use more connecting adverbials than male authors. While previous studies on different genres have found considerable influence of gender on the use of language (Kortmann 2005) and cohesive devices in particular (Livia 2003), the situation in academic writing seems to be different (Hůlková 2011).

The third research question deals with the mother tongue of the writers. The influence of the L1 on student writing has been researched quite extensively,

respect, for this reason, on account of this, as a result, on this basis, whence sequential: firstly, secondly, thirdly, previously, afterward(s), eventually, finally, lastly, anyhow, anyway, next, at this point, to sum up, in short, in the end, ultimately, at last however there are studies that claim little or no effect (Gilquin & Paquot 2008), as well as studies that claim a considerable effect (Martin-Martin 2005). Hence, it is interesting to investigate what can be found for the data sets at hand. Therefore, German L1 writers will be compared to L1 writers of other Germanic languages, such as Dutch or Swedish, and it will be possible to see whether German writers use more connecting adverbials than the writers with other Germanic languages as their L1. It is important to consider that the ICLE, which will be used for comparison, only contains texts written by authors with two other Germanic native languages, Dutch and Swedish.

The last research question steps away from the social variables and investigates which connecting adverbials are prototypical for each of the four functional categories. Prototype semantics generally assumes that the boundaries of semantic categories are "much more flexible and fuzzy than is suggested by traditional componential semantics" (Kortmann 2005: 209). Due to this fuzziness, there are "central or typical members of a category" (Saeed 2009: 37), which are the best representation of the underlying cognitive concept. Consequently, the prototypical connectors will be identified and then compared to the rest of the connectors in this category. It is hard to make a prediction here, but I would guess that the prototypical connecting adverbials make up a large percentage of the connectors in the respective functional categories.

After answering all of the four research questions presented, an extensive picture of connecting adverbial usage by non-native writers of English will be visible. Furthermore, possible areas for future research can be highlighted.

2. Methodology

The methodology employed in this study is a quantitative comparative analysis of two corpora, one of them being the main object of the research and the other serving as reference.

2.1. Data: *ChemCorpus*

The *ChemCorpus*, which has been compiled at Technische Universität Chemnitz (Schmied 2015, table 3 above, this volume) contains texts produced by students in the Bachelor and Master Program in English and American studies as well as some older text produced by students of the Magister Program, which, however, is no longer offered. Hence, all the texts have to be considered apprentice texts, even though they have been produced at the end of the programs.

For the purpose of this study, not the whole corpus was used but subset grouped by text types, which are written magister theses (MagTheses) and written magister exams (MagWritten). The corpus has been compiled from the year 2001 onwards, with the most recent texts used for the analysis in this paper, dating to 2011. The present corpus contains 1,709,983 words, and the complete corpus consists of four text types, i.e. written magister exams, magister theses, bachelor theses and master theses. The sub-corpus used in this paper consists of

1,059,263. The Magister theses in the corpus are further divided by area, namely Culture and Literature on the one side and Linguistics on the other. However, this division is neglected, since this variable is not investigated. Nevertheless, it is important to keep in mind that the area in which the papers have been written might also have an impact on the writing style and ultimately on the use of connecting adverbials (Hůlková 2011). Moreover, the texts are not evenly distributed into the subcategories, which makes the corpus not ideally stratified and might lead to minor skewed data. Furthermore, it is essential to note that the disproportion in the figures originates in the different lengths of the texts in the categories. The texts in the written exams section contain on average 1,987 words, with a standard deviation of 593 words, whereas the texts in the theses section contain 29,379 words on average, with a standard deviation of 9,990 words. This may be an issue, since the untimed part of the corpus is multiple times as large as the untimed component.

2.2. Data: Monitor corpus

For this paper, the *International Corpus of Learner's English (ICLE)* is used as a monitor corpus. It contained timed and untimed "argumentative essays written by higher intermediate to advanced learners of English from several mother tongue backgrounds" (UCL - ICLE 2012). Similarly to the *ChemCorpus*, not the whole corpus was used but only a subset, since for a considerable number of texts the metadata was incomplete and contained no information on the timedness. The two categories of timed and untimed texts contain 639,673 words and 1,684,555 words respectively, totalling 2,324,228 words. The untimed texts are on average 700 words long, with a standard deviation of 295 words, whereas the timed texts average 619 words with a standard deviation of 302 words. There is not much difference in length and deviation between the two categories, which is in stark contrast to the *ChemCorpus*, where the difference in length between timed and untimed writing is more than a factor ten.

2.3. Corpus compatibility and issues

When comparing two corpora, the question of data compatibility arises. During the analysis of the two corpora used in this paper, some issues arose by looking at the numbers of texts and the word count. While the corpus that is primarily researched contains roughly 1 million words, the monitor corpus contains approximately 2.3 million words, which is more than twice the primary corpus. Thus, this issue has not much weight in terms of comparability, as for comparing the two corpora the absolute values can easily be transformed into normalized, relative values. However, when taking a closer look at the texts comprising the corpora, another issue becomes apparent. Table 1 shows that the texts in the timed components of the *ChemCorpus* and the *ICLE* differ by a factor of more than three. The differences in text length are even more extreme for the untimed components, where the texts differ by a factor of more than 40. In general, the *ICLE* is stratified in terms of text length, whereas the *ChemCorpus* is not.

	***ICLE* timed**	***ICLE* untimed**	***ChemCorpus* timed**	***ChemCorpus* untimed**
Average Text Length in words	619	700	1,987	29,379
Standard Deviation	302	295	593	9,990

Table 1: Average text length of corpora

But since the overall number of words is well above one million in both corpora, there should be no problem in the analysis.

2.4. Text Type

When discussing the comparability of the data, the different text types are also important. While the *ChemCorpus* contains exam essays and theses, the *ICLE* contains argumentative essays. The comparability of these text types, however, is still subject to discussion. They differ not only in length, as seen above, but also in the communicative strategies utilized. While the argumentative essay usually tries to convince the reader to something by presenting a structured argumentation, the thesis presents an academic subject. Henceforth, the linguistic strategies to achieve the communicative purposes differ as well. Consequently, it is necessary to compare these two text types to assess whether they are comparable or not. Theses generally follow the Introduction – Method – Results – Discussion (IMRD) structure of research articles (Samraj 2008: 57), as they can be seen as apprentice texts of this type. However, Samraj shows that, depending on the department, sometimes different strategies are utilized. As a result, it can be expected that student's theses are written in a style mostly similar to research articles, implying a use of connecting adverbials similar to research articles.

The argumentative essay, however, has a different purpose, i.e. "to persuade the reader of the correctness of a central statement" (Hyland, 1990: 68), resulting in a different structure. The text is structured in three stages. Firstly, there is the thesis, which is accompanied by an attention grabber and background information on the topic, as well as a short evaluation, which briefly supports the thesis. Secondly, the main argumentation, which presents a number of claims that are furthermore supported by evidence, follows. Lastly, the conclusion, which rounds up the argumentation and reaffirms the thesis, completes the essay (Hyland, 1990). Resulting from this structure, the connecting adverbials used will differ too. Since it is the purpose of an argumentative essay to persuade the reader, the connecting adverbials used to create cohesion will be mostly additive and causal. Adversative connecting adverbials will most likely be used less frequently, since they would not match the communicative purpose of relating arguments to each other.

The two text types are quite different in structure and communicative purpose. Thus, the results of the analysis have to be examined very carefully and critically, since the effect of the different text types in the two corpora cannot be reliably predicted.

2.5. Gender

Another issue is the stratification of the corpora. As Schmied (2011) points out, there is a severe lack of male students within the field of English Language Studies at universities. This problem is mirrored in the data used for this study. The *ChemCorpus* contains only nine texts written by male students at a total of 85 texts, with 15 texts for which there is no information concerning gender. For the *ICLE* corpus there are 2833 texts written by female students, 601 texts written by male students and nine texts without available gender information. These numbers show a clear skewedness towards texts written by female students. While this is an issue affecting the stratification, it might not have any influence on the results, since Hůlková (2011) has shown that gender has no significant influence on the usage of connecting adverbials and that possible minor differences in the statistics are merely idiosyncratic (Hůlková 2011: 137).

2.6. Department

Furthermore, the *ICLE* does not contain information regarding the department which the students who produced the texts in the corpus belong to. The *ChemCorpus* on the other hand differentiates between Linguistics texts and Culture and Literature texts, but only for untimed texts and with an unequal distribution. Furthermore, the influence of this distinction on the actual use of connecting adverbials is unclear and needs further investigation, since previous studies on the impact of the department have only differentiated between major fields and not within the fields (Hůlková 2011). Thus, it is debatable if department, as represented in the *ChemCorpus*, has an influence because all the texts belong to the field on humanities. Since the availability of the department metadata is very limited in the two corpora, it had to be omitted as a factor in the analysis.

All in all, the two corpora at hand differ in some ways, but the skewedness in the structure of the data seems to be similar. This fact, and the relatively large amount of data, totalling approximately 1 million and 2.3 million words for the *ChemCorpus* and the *ICLE* respectively, still makes the corpora a reasonable data base for analysis.

3. Results

3.1. Quantitative comparison

The statistical analysis of the data resulted in figures that exhibit differently clear tendencies for the foci of this study. The first aspect of the corpus that has been analysed is timedness. Figure 1 shows the average total of connectors according to timed – untimed writing in the two corpora side by side.

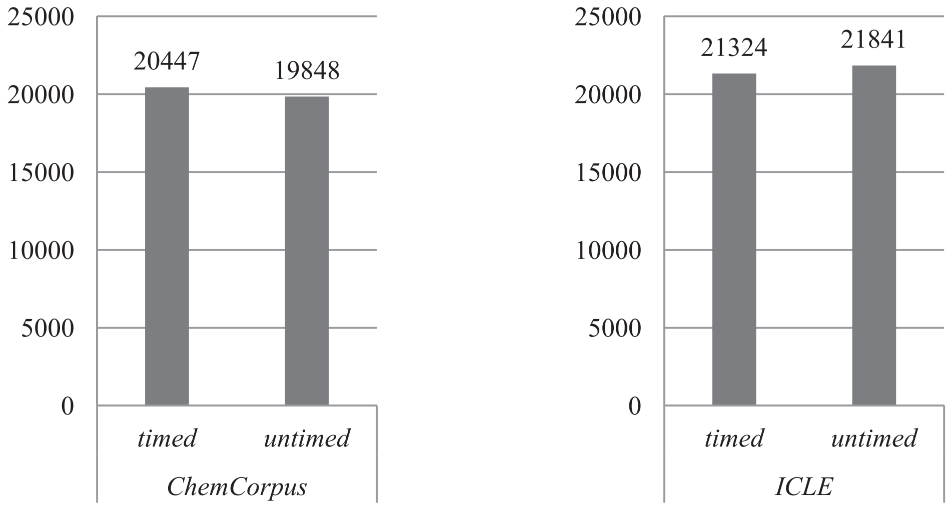


Figure 1: Comparison timed – untimed in *ChemCorpus* and *ICLE* (per one million words)

The figure shows that the numbers are very close together for the two branches within each corpus, as well as for both corpora in comparison, averaging at around 20,000 connectors per one million words, which equals two percent of the words. It is furthermore curious that there are more connecting adverbials in the timed section of the *ChemCorpus*, whereas in the *ICLE* there are more connecting adverbials in the untimed section. When calculating the p-values it can be seen that the difference is rather significant for the *ChemCorpus* at 0.002845 whereas for the *ICLE* it is less significant at 0.01283. Since both numbers are not highly significant, it is necessary to look at the statistics for the functional categories to see if they follow the same overall trend.

Figures 2 and 3 the statistics aggregated by functional category and timedness for both corpora. As the average total has already suggested, there is hardly any apparent difference. Especially in the adversative, casual and sequential categories of the *ChemCorpus* the difference is strikingly low. While for the functional categories in the *ICLE* the difference is more visible it still remains low. To verify the assumption that the differences are rather insignificant, chi-square tests have been calculated and resulted in the p-values in Table 2 (below). They confirm that the difference is mostly insignificant and only the additive category exhibits a significant effect in both corpora. Furthermore, the sequential category for the *ICLE* data shows a significant effect.

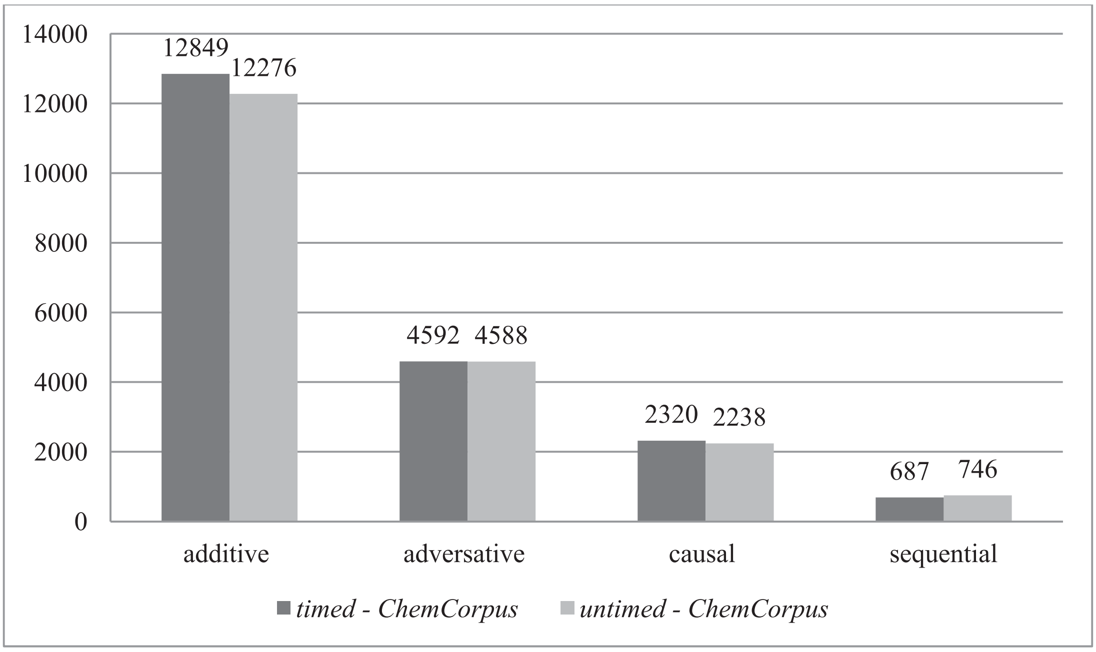


Figure 2: *ChemCorpus* connecting adverbials by functional category (per one million words)

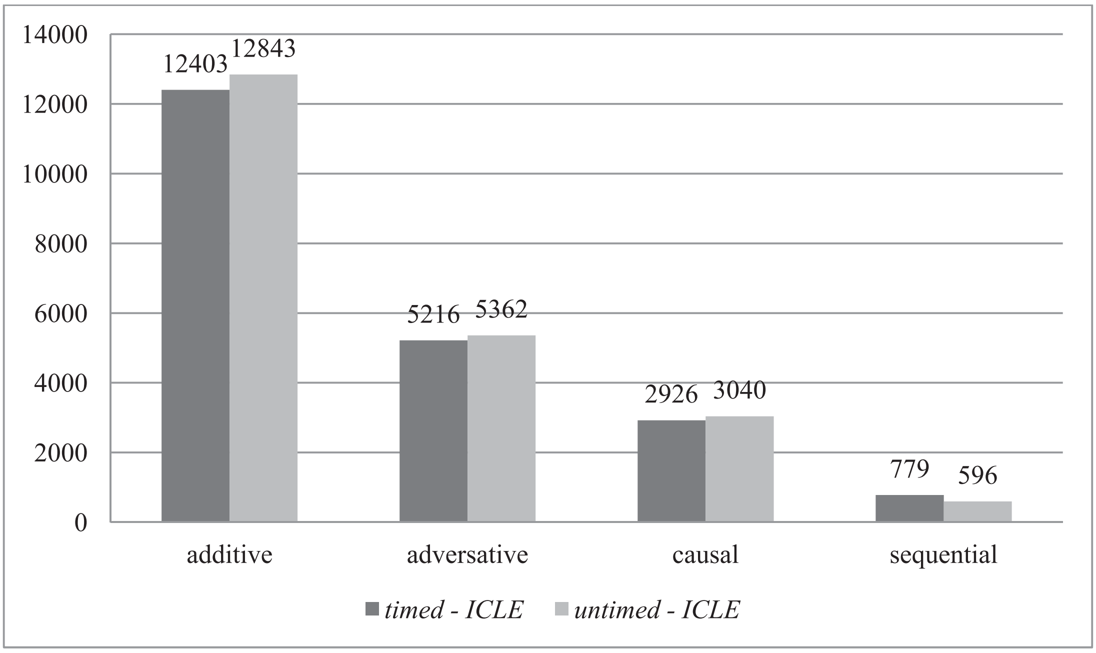


Figure 3: *ICLE* connecting adverbials by functional category (per one million words)

In the following, I want to shift the attention to the variable of gender. Even though Hůlková (2011) finds that in her data gender did not have any influence on the usage of connecting adverbials, I decided to include gender in my analysis to see whether her findings could be confirmed. Again, the analysis splits up the connecting adverbials into the four functional categories to provide a more detailed view. Figure 4 below shows the results.

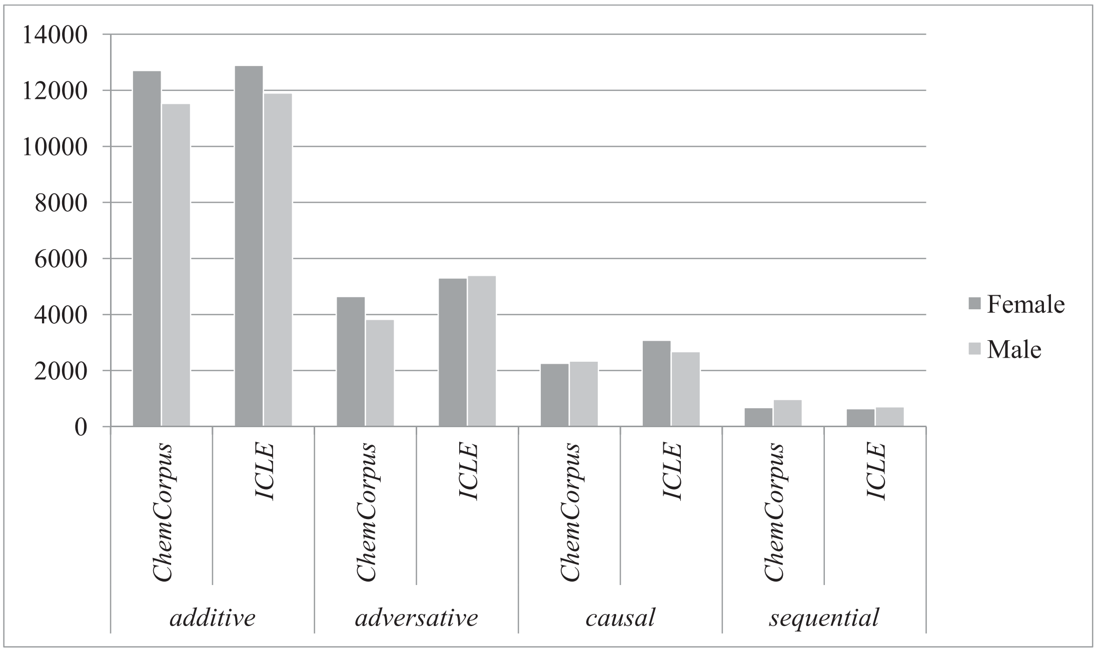


Figure 4: Functional categories by gender (per one million words)

The bars in the chart already suggest that female writers use connecting adverbials more frequently than their male counterparts, but the difference is not as very obvious for all categories. While women use connectors in total roughly 1,600 more per one million words for the *ChemCorpus* and 1,200 per one million words in the *ICLE*, the functional categories do not exhibit the same usage pattern throughout. The *ChemCorpus* data shows that women predominantly use more additive and adversative connecting adverbials than men, whereas it is the other way round for the causal and sequential category. For the *ICLE* data, the usage patterns are even less visible. Female authors use more additive and causal connecting adverbials, whereas male authors use more adversative and sequential connectors. However, the differences for the sequential and especially for the adversative category are very low.

The next factor that has been analyzed is the mother tongue of the authors. However, the parameters for the analysis are far from ideal, since the *ChemCorpus* only contains texts by German native speakers, whereas the *ICLE* consists of texts written by students from a wide range of L1 backgrounds. Hence, only the native speakers of other Germanic languages in the ICLE have been compared to the *ChemCorpus* data.

Figure 5 shows that authors with Swedish as their L1 use additive and adversative connecting adverbials more frequently than others, whereas authors with Dutch as their L1 use more causal connectors. It is furthermore notable that there is a discrepancy between the usage of connecting adverbials by German L1 authors of the *ChemCorpus* and the *ICLE*. The *ICLE* texts exhibit a higher usage of adversative and sequential connectors, which may be due to the different text types of the two corpora. Overall, the authors with Swedish or Dutch as their L1 use roughly 2,000 connecting adverbials more per one million words than their German colleagues.

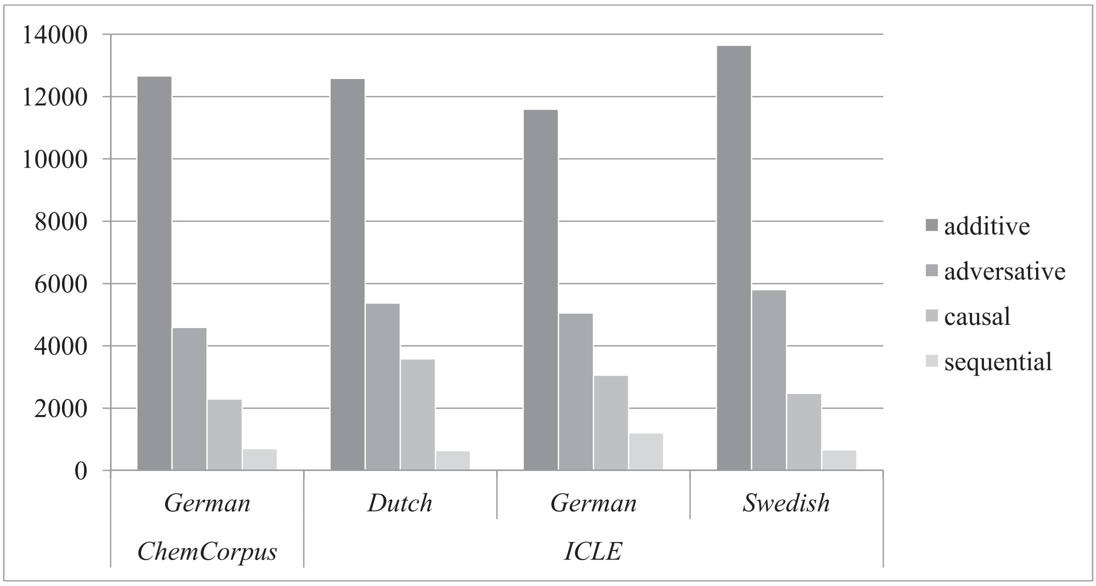


Figure 5: Functional categories by L1 (per one million words)

The last aspect that will be analyzed in this section is prototype. Here, the most frequent connecting adverbials were treated as prototypical and compared to the rest of connectors in the respective categories. The following connecting adverbials were selected as prototypical: *and* (additive), *but* (adversative), *because* (causal) and *next* (sequential). Figure 6 and 7 show the plotted statistics for the comparison prototypical connecting adverbials against the rest of the respective categories.

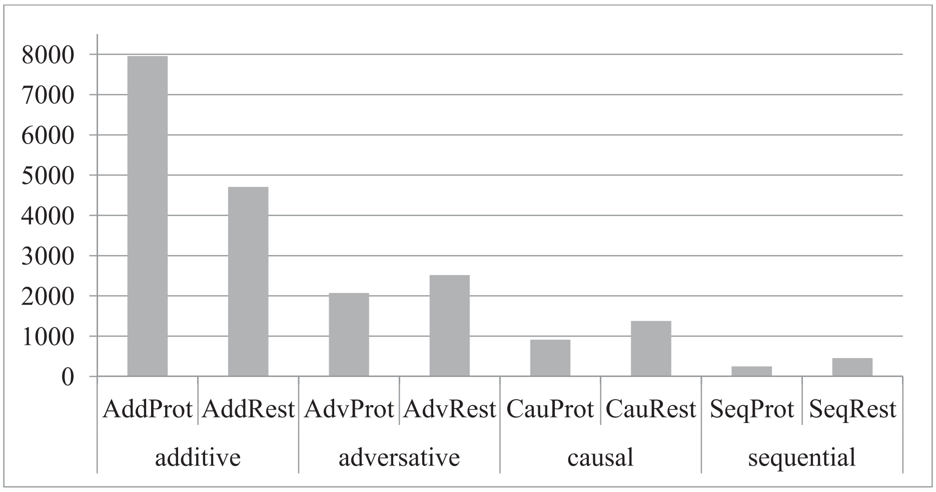


Figure 6: *ChemCorpus* functional categories by prototype (per one million words)

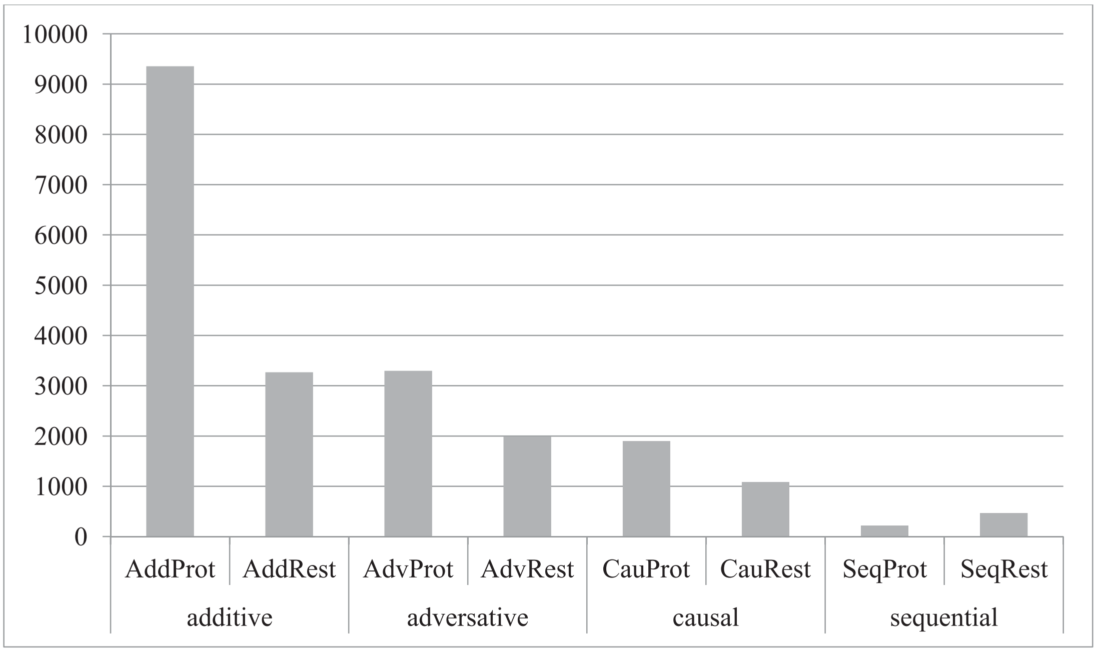


Figure 7: *ICLE* functional categories by prototype (per one million words)

It is striking that the results for the two corpora are almost completely different. The statistics for the *ChemCorpus* show that the additive prototype occurs more frequently than the other connecting adverbials in the category, whereas for the other three categories the frequency of the prototype is lower than the rest of the connecting adverbials combined. In contrast, the statistics for the *ICLE* data coherently show that the prototypical connector in three of the four categories occurs more often than the rest of the connectors, however, the sequential prototype occurs less frequently than the rest of the connectors in the category combined. It is interesting that the statistics for the *ICLE* data show basically the reverse tendencies that could be observed for the *ChemCorpus* data. Moreover, the difference between prototype and rest is bigger for the *ICLE* data, which hints at a lower diversity of connecting adverbial used in the texts.

Finally, the statistical significance of these results has been tested by carrying out chi-square test. The results will confirm, whether the differences are really due to the parameters or only by chance. As Table 2 below shows, the effects for gender, L1 and prototype are clearly genuine and not by chance. However, for timedness, the results are not as clear as expected, since there seems to be a very significant effect for the *ICLE* data, whereas no significant effect can be proven for the *ChemCorpus* data.

	Timedness	**Gender**	**L1**	**Prototype**
ChemCorpus	> 0.05	< 0.001	< 0.001	< 0.001
ICLE	< 0.001	< 0.001	< 0.001	< 0.001

Table 2: Significance of timedness, gender, L1 and prototype on *ChemCorpus* and *ICLE* (p-values)

3.2. Interpretation

Firstly, the analysis revealed that there seems to be no definite answer to the question of whether the timedness of a text has a significant influence on the use of connecting adverbials. When comparing the overall connecting adverbial use in the timed and untimed branches of the corpora, the numbers are on similar levels. Nevertheless, overall the phenomenon is statistically significant for the *ICLE* data only and not for the *ChemCorpus* data, as shown by the p-values from the chi-square test. Thus, the first research question cannot be answered definitely.

Secondly, there is a clear pattern of use for gender. The statistics have shown that female student writers use more connecting adverbials in total, albeit this is not exactly the same for all four functional categories. This overall difference was even more prominent in the *ChemCorpus* than in the *ICLE*. The gender specific use of connecting adverbials for the four functional categories exhibits a pattern in the *ChemCorpus*, where female authors use more additive and adversative connectors, whereas male authors use more causal and sequential connectors.

Thirdly, the analysis has shown that the L1 of the authors has an influence on the use of connecting adverbials. It could be seen that writers with Dutch or Swedish as their L1 use more connecting adverbials in general. In particular, Dutch writers of English use more adversative and causal connectors whereas Swedish writers of English use more additive and adversative connectors. When calculating the significance for these numbers, it can be seen that overall, the L1 has a significant effect on the use of connecting adverbials. These findings are quite astonishing, as the initial assumption was that German authors would use more connecting adverbials than writers with other Germanic L1's.

Lastly, there is a very clear pattern observable for prototype. The data shows that the prototypical connecting adverbial, meaning the most frequently used, accounts for a big, and for parts of the data even major, share of the overall connector use in the respective functional category. At this point, it is necessary to differentiate the results for the two corpora. The *ChemCorpus* shows that the prototypical connecting adverbials make up a large share of all of the connecting adverbials in the functional categories, but not the majority, except for the additive category. The *ICLE* data shows the opposite. Here, the prototype accounts for the majority of connecting adverbials in the functional categories, except for the sequential category. This shows that the *ChemCorpus* has a greater diversity of connecting adverbials, whereas the *ICLE* is dominated by the prototypical connecting adverbials. Despite these differences, the results are highly significant for both corpora.

4. Discussion

The results presented in the previous section have to be treated carefully, as they are particular to my study and are likely to be influenced by its parameters.

One of the key questions that has already been touched on earlier concerns the compatibility of the data sets. The problem here results from the different text types in the two corpora and the very different lengths of the individual texts. Additionally, these deficits become especially apparent when discussing the influence of gender. In my study, gender has a statistically significant influence in three out of the four functional categories for the *ChemCorpus* data and in two of four categories for the *ICLE* data. These findings directly contradict the study by Hůlková (2011) who found that, for her data, gender did not have any influence on the use of connecting adverbials. However, despite being statistical significant, the results for the *ChemCorpus* have to be treated very carefully since male students are strongly underrepresented in the data. Thus, the decision whether gender really affects the use of connecting adverbials cannot be answered at this point since it is unclear if the results have been altered by the aforementioned deficits.

Furthermore, looking at the results for the timedness factor, the possible effect of the data insufficiencies can be seen ever more clearly. Here, a statistically significant effect can be observed for the *ICLE* data, which contains highly compatible texts of the same text type and similar length in both the timed and untimed category. For the *ChemCorpus* data on the other hand, there seems to be no statistically significant effect. This might result from the composition of the *ChemCorpus*, as the texts in the timed and untimed categories belong to different text types and vary greatly in length. Therefore, to answer the research question on the effect of timedness more precisely, it would be necessary to carry out another study with data that is more compatible and thus allows for a better comparison of the results.

A general issue that arises is the question about the significance of p-values. Statistical significance alone merely indicates that there is an effect, but does not allow any conclusions on the size of the effect.
Consequently, paying special attention to these factors is crucial before drawing conclusions on implications for teaching and future research.

The problems that have been discussed here, especially those related to the comparability of the data sets, highlight the importance for the creation of well stratified learner corpora.

References

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

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

Bolton, K., Nelson, G., & Hung, J. (2002). A corpus-based study of connectors in student writing Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics 7:2*, 156-182.

Gilquin, G., & Paquot, M. (2008). Too chatty Learner academic writing and register variation. In *English Text Construction* (1 ed., Vol. 1, pp. 41-61). John Benjamins Publishing Company.

Halliday, M. A., & Hasan, R. (1976). *Cohesion in English.* New York: Longman.

Hůlková, I. (2011). Conjunctive Adverbials in Academic Written Discourse: Conjunctive Adverbials in Academic Written Discourse. In J. Schmied, *Academic Writing in Europe: Empirical Perspectives* (pp. 129-142).

Hüttner, J. I. (2007). *Academic Writing in a Foreign Language - An Extended Genre Analysis of Student Texts.* Frankfurt am Main: Peter Lang.

Hyland, K. (1990). A Genre Description of the Argumentative Essay. *RELC Journal*(21), 66-78. doi:10.1177/003368829002100105

Kortmann, B. (2005). *English Linguistics: Essentials.* Berlin: Cornelsen.

Livia, A. (2003). Linguistic Approaches to Gender. In J. Holmes, & M. Meyerhoff, *The Handbook of Language and Gender* (pp. 142-158). Oxford: Blackwell.

Martin-Martin, P. (2005). Scientific Writing: A Universal or a Culture-Specific Type of Discourse? In *Revista de Lenguas para Fines Especificos* (Vols. 11-12, pp. 191-203).

Saeed, J. I. (2009). *Semantics* (3 ed.). Oxford: Wiley-Blackwell.

Samraj, B. (2008). A discourse analysis of master's theses across disciplines. *Journal of English for Academic Purposes*(7), 55-67.

Schmied, J. (2011). Academic Writing in Europe: a Survey of Approaches and Problems. In J. Schmied, *Academic Writing in Europe: Empirical Perspectives* (pp. 1-22).

Swales, J. M. (1990). *Genre Analysis - English in academic and research settings.* Cambridge: Cambridge University Press.

UCL - ICLE. (2012, 12 19). Retrieved from http://www.uclouvain.be/en-cecl-icle.html

"It is probably the reason why…" Hedging in BA and MA Theses by German ESL Students

Dana Beyer (Technische Universität Chemnitz)

Abstract

Among linguists (Lakoff 1973; Hyland 1998; Nuyts 2001 & 2005) hedging in academic writing of native as well as non-native speakers has been of interest ever since. The phenomenon of hedging devices, used in an academic environment to downtone and mitigate scientific claims in order to gain credibility and acceptability, has been researched under various distinct aspects. The present paper tries to investigate how German students of English and American Studies at TU Chemnitz apply epistemic adjectives and adverbs in their BA and MA theses. Ten BA and ten MA theses were selected from the *ChemCorpus* (non-native academic student writing in English) by the same students to ensure the comparability and possible improvements in the usage of hedges from the BA to the MA level. The adjectives *probable*, *likely*, *possible*, *apparent*, *presumable* and *unlikely* and the adverbs *probably*, *possibly*, *apparently*, *presumably*, *seemingly* and *relatively* were analyzed according to their frequency in the theses, the preference of students toward the usage of either epistemic adjectives or adverbs and whether there is a difference in the choice over more tentative than less tentative adjectives and adverbs. The investigation showed that the frequency of epistemic adverbs is indeed higher than the occurrences of epistemic adjectives. Additionally, students employed more hedges in their MA theses than in their BA theses and there seems to be a preference for less tentative hedges, such as *probably*, *likely*, *relatively* and *possible/possibly*. These findings indicate that the students are well aware of their field research and its outcomes and how to apply hedges in their academic writing. The higher number of adjectives and adverbs used in the MA theses also hints at a progress in writing and the greater awareness of the importance of hedging devices in academic writing.

1. Introduction

Academic writing for non-native speakers in English still seems to be problematic in terms of how to find the right balance between objectivity and subjectivity. There is not only the issue of writing in a language different from the mother tongue, but also that the writing cultures of distinct countries might vary to a great extent from one another. One linguistic phenomenon of academic writing that has received quite a lot of attention in recent years is hedging (Hyland 1998, Nuyts 2001, Schmied 2011).

Hedging in the academic community refers to the act of using language in a more indirect and cautious manner with regard to research findings. As stated above, it is the writer's decision to find a balanced way of subjective and objective writing. Since hedges have been of great interest to linguists, it is comprehensible that there exist manifold classifications of hedges within the framework of epistemic modality (Schmied 2009). The discussion on classifying these devices ranges from where does epistemic modality start to which lexical items can be considered a hedge and which items are no hedges. Due to these manifold theories, it seems essential to create a working definition of hedges which applies to the aims of this paper. The definition and classification of hedges will be given in chapter 2.

The paper intends to elicit possible differences in the usage of hedging devices among German students of English and American Studies at TU Chemnitz (*ChemCorpus*, Schmied, this volume). Therefore, ten BA and ten MA theses, written by the same students, were selected and investigated for epistemic adjectives and adverbs. The corpus analysis is supposed to shed light on whether the occurrences of the selected hedges differ from BA to MA level and whether students prefer adjectives and adverbs with distinct degrees of likelihood (cf. Nuyts, 2001). These viable differences may show whether the students progressed their writing skills with regard to hedges from their BA to MA studies.

The subsequent chapters will give an overview of the literature on epistemic modality and hedging as a part of it. Beginning from chapter 3, the methodology, the results and the interpretation of the results of this corpus analysis will be presented.

2. Modality

2.1. Epistemic modality

As academic L2 student writing has sparked linguists' interests, modality is merely one linguistic concept that has received relatively much attention in recent years (cf. Nuyts 1993, 2001; Hyland 1998; Van linden 2012). Although many linguists have tried to find a coherent model of modality and its distinct forms, it seems to be impossible to agree on one universal classification. As Nuyts (2005) puts it modality "remains one of the most problematic and controversial notions: there is no consensus on how to define and characterise it" (Nuyts 2005: 5). What seems to be clear when we talk about the modal characteristic of an utterance is that it is "the grammaticalization of the speakers' (subjective) attitudes and opinions" (Palmer 1993: 51). The semantic domains of modality have clear cut boundaries and can be divided into probability, possibility and necessity. But if the conceptual categories of modality are taken into consideration, several proposals have been published. The conceptual categories by scholars, such as Coates (1983), Bybee et al. (1994) and Van der Auwera & Plungian (1998), are not as detailed as Palmer's (1993) and Nuyts' (2005) concepts (cf. Van linden 2012: 26). Overall, modality can be subdivided into three conceptual categories, namely dynamic, deontic and epistemic. Dynamic modality is concerned with the subject-participant's ability and capacity in the clause, i.e. the subject is able to do something that is expressed by the clause's main verb (Nuyts 2005: 7). Contrary, deontic modality mainly revolves around the notions of obligation and permission (Nuyts: 2005: 9). Nevertheless, as has already been hinted at before, these concepts of modality have more or less detailed subcategories depending on the preferred linguistic framework. But due to the focus on epistemic modality in the course of this paper, these short definitions of these two remaining categories are sufficient for the following discussion. The main focus of this paper is that of the notion of epistemic modality, which is best described by its characteristic of "the degree of probability of the state of affairs" expressed by the modal auxiliary, verb, adjective and/or adverb (Nuyts 2005: 10). In other words, it is the writer who judges the truth of the proposition in terms of possibility, probability and certainty (cf. Hyland 1998: 44). As Nuyts' definition already indicates epistemic modality can have distinct degrees, which means that these degrees can be seen on a scale. This scale ranges from a positive side of absolute certainty to probability, possibility towards an implicitly negated side of improbability and absolute uncertainty that the state of affairs is unreal (Nuyts 2005: 10). Figure 1 displays this scale of epistemic modality.

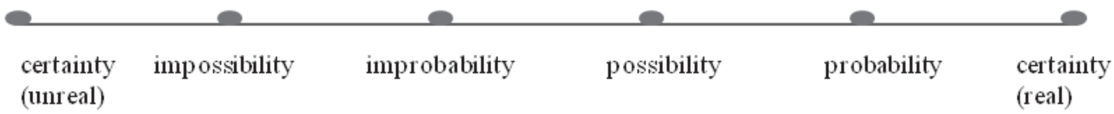


Figure 1: Scale of the likelihood of the occurrence of an event (Nuyts 2005: 11)

Although for some approaches of epistemic modality, such as Kratzer (1978), this scale may not be acceptable. However, the English language provides a wide variety of modal adjectives and adverbs that facilitate the expression of certainty (real and unreal), impossibility, improbability, possibility, and probability (cf. Nuyts 2005: 11). A further degree of the gradability of probabilities of states of affairs are grading expressions usually before modal adjectives, for example *very probable*, *rather certain*, *almost certain*, etc. (cf. Nuyts 2005: 11). This shows that epistemic modality relates to lexical items that display confidence and can as well refer to a lack of confidence in the truth value of the utterance(s) (cf. Hyland 1998: 44). On the one hand, the confidence and the lack of confidence also indicate the writer's commitment to the state of affairs, i.e. how convinced the reader/writer is of the state of affairs (cf. Van linden 2012: 20). One the other hand, it also illustrates the subjective character of the reader's/writer's set of beliefs, which in turn means that epistemic modality is also non-factual (cf. Hoye 1997: 43).

The discussion on modality and its subcategory of epistemic modality can be extended by far, but due to the brevity of this paper the summary above should be sufficient.

2.2. Hedges

Hedges can be considered one aspect of epistemic modality (cf. Hyland 1998: 44). The term "hedge" was coined by G. Lakoff in 1972 and describes "words whose job it is to make things fuzzy or less fuzzy" (Lakoff 1973: 471). The difficulty of Lakoff's definition is whether it is feasible that hedges are able to fulfill two opposing conversational tasks: to mitigate a claim ("make things fuzzy") and to communicate a claim directly ("less fuzzy"). Crompton (1997) also indicates that a less fuzzy expression rather "seem likely to increase rather than avoid the utterer's commitment" (Crompton 1997: 272). Judging by its name, hedges are expressions that intend to "hide something", i.e. the writer/speaker avoids their commitment to a statement or proposition instead of saying something in a direct manner (cf. Crompton 1997: 271). The concept of "boosters", which strengthen rather than mitigate claims, are probably a more fitting expression for less fuzzier statements. Although there have been numerous studies on hedging expressions in academic writing (Hyland 1998; Hinkel 2002), a uniform definition for this linguistic phenomenon does not exist. This may be due to the fact that modality is a very subjective concept, which does have clear cut boundaries. But the term hedge might also be hard to define since these expressions are not restricted to a certain linguistic item, which renders it possible that every lexical item expressing some kind of tentativeness, mitigation, subjective commitment, etc. can be counted as a hedge. A more detailed definition of hedges is provided by Hinkel (2002) who states that

> it [hedges] represents the use of linguistic devices to decrease the writer's responsibility for the extent of the truth value of propositions/claims to show hesitation or uncertainty, and/or to display politeness and indirectness in order to reduce the imposition on the writer or the reader (Hinkel, 2002, p. 148).

Hinkel (2002) includes a two-dimensional characteristic of hedges – they do not only save the writer's/speaker's face, but they also influence the judgment of the claims by the reader and avoid face-threatening of the reader (cf. Hyland 1998: 177). It seems that hedged propositions gain more acceptance among peers than non-hedged propositions because it shows the writer's degree a credibility in their research field (cf. Huebler 1983: 18; Hyland 1998: 178). In addition, reader-oriented hedges show that the writer/speaker is willing to discuss and negotiate their scientific findings, whereas direct assertions leave little or no room for negotiation (cf. Hyland 1998: 178). It can even be suggested that more important scientific claims are less hedged than less important claims, which Swales confirms by saying that "high-level claims are likely to be important but risky, whilst low-level claims are likely to be trivial but safe" (Swales 1991: 117).

Clemen (1998) goes one step further by saying that hedges always depend on the context and are not easily classified, which illustrates again that there are no set criteria and no fixed linguistic items that can be attributed to hedges (cf. Clemen 1998: 237). Nevertheless, for the following analysis of L2 writing, a definition is needed, which will be adapted from an earlier ESL student writing analysis on hedges:

> Hedges are any linguistic devices that a) express a writer's uncertainty of/tentativeness towards a proposition/claim, in which this proposition/claim is transformed into an opinion rather than a fact, and/or b) increase the politeness and the social acceptability of propositions/claims to avoid conflicts between the writer and the reader (Beyer, 2009, p. 3).

The above definition subsumes the most important features and functions of hedges. Foremost, it mitigates the writer's statements by using hedging devices.

Furthermore, it shows the non-factual and rather subjective character of hedges, which has also been emphasized in the previous chapter about epistemic modality. By weakening claims, hedges also act as markers for politeness and acceptability towards the reader. Here the two dimensions become again obvious: writer-oriented hedges save the writer's face and protect their reputation by diminishing the commitment towards their findings. Reader-oriented hedges are the second dimension helping to establish a writer-reader-relationship and minimizing the degree of criticism towards scientific peers that may arise when the same research field is studied (cf. Hyland 1998: 174-179).

2.3. Epistemic modal adjectives and adverbs

According to Nuyts (2001), modal adjectives and adverbs can be seen as the "purest expressions for epistemic modality" in West Germanic languages because they most clearly mark the degree of likelihood of the state of affairs (cf. Nuyts 2001: 55). As the scale of the degrees of likelihood in section 2.1 already indicated, the modal adjective/adverb *certain(ly)* is found at the extreme positive end, whereas *probable/probably* are in the middle of the positive side and *possible/possibly* are near the neutral point (cf. Nuyts 2001: 55). Depending on the context, these degrees of likelihood can naturally change and the speaker/writer can additionally scalarize their position by using grading expressions, e.g. *rather certain*, *very probable*, *almost certain* (cf. Nuyts 2005: 11). The classes of epistemic adjectives and adverbs overlap, but they are by no means identical (cf. Nuyts 2001: 55). The class of modal adverbs is larger than that of modal adjectives, which is explained by the fact that adverbs are more frequently applied than adjectives (cf. Nuyts 2001: 56). Since modal adverbs have a "floating" character, they are able to appear in several distinct positions within a sentence without influencing the meaning relation between clause and adverb (cf. Hyland 1998: 134; Nuyts 2001: 57). Example (1) and (2) illustrate the usage of *it is probable* and *probably*. With the scope of the modality being the only deciding factor, adverbs can either stand at the beginning of the utterance and function "thematised" or they can be inserted intrasententially (Hyland 1998: 134). However, the clause initial position of the modal adverb may highlight the hedge and influence the reader's understanding of the entire sentence (cf. Hyland 1998: 135). Adverbs can either refer to the entire clause or they allude to a phrase only (cf. Hyland 1998: 138). The cleft-like construction *it is probable* suggests something similar as the clause initial position for epistemic adverbs, i.e. "the modal qualification acquires some special informational status" and is therefore put into focus at the beginning of the utterance (Nuyts 1993: 938ff.). As can be seen in example (1), modal adjectives cannot change their positions within an utterance – they are restricted to one position.

(1) It is probable that she is going to work.
(2) She is probably going to work.

Another feature of modal adjectives is that they can occur in interrogatives, whereas modal adverbs cannot be questioned (compare examples (3) and (4)). But there seem to be exceptions, such as example (5):

(3) Is it probable that she is going to work?
(4) *Probably she is going to work?
(5) Does she perhaps go to work?

Sentence (4) clearly illustrates that the modal adverb *probably* cannot be questioned. If we look at (3) and (5), these sentences are perfectly fine. However, in (5) the adverb *perhaps* does not have an epistemic meaning, but "rather speech act modifying elements" (Nuyts 2001: 58). The usage of *perhaps* in sentence (5) is derived from the epistemic meaning, but it does not imply the degree of likelihood of the state of affairs. The adverb rather alters a neutral question into a tendentious one (cf. Nuyts 2001: 58). It is also not the adverb that is being questioned in the sentence, but whether *she goes to work*.

One last attribute that differentiates modal adjectives from modal adverbs is that the latter ones do not have negative forms.

(6) It is improbable that she is going to work.
(7) *Improbably she is going to work.

Unlike modal adverbs, modal adjectives can be negated, which is shown by sentence (6). For example (7), it is impossible to negate the adverb *probably*. One reason could be that modal adverbs reflect the speaker's/writer's attitude toward the proposition, which indicates that they cannot negate their own attitude (cf. Nuyts 2001: 60). This is also true for the non-questionability of epistemic adverbs since the writer cannot question their opinion. On the contrary, modal adjectives belong to the proposition, i.e. they are integrated and "qualify the state of affairs referred to by the utterance" (Nuyts 2001: 60).

2.4. Hedging in ESL student writing

Mitigating claims by applying hedges in academic writing by ESL students still seems to be a rather scarcely researched field. Nevertheless, it is crucial for a writer of academic texts, no matter whether these texts are being published or not, to establish a credible and suitable academic writing style. Hedges are one linguistic means to positively connect to their readership and to convincingly state their research aims and findings. Although there have been ongoing discussions about the phenomenon of hedges, non-native speakers still seem to be either unaware or insecure how and when to use hedging devices. In order to become conscious and confident in this field, the application of hedges in academic texts has to be included into the curriculum. However, Hyland (1998) remarks that there is still no sufficient amount of adequate literature on hedging available (cf. Hyland 1998: 218). Moreover, epistemic meanings can be expressed in many different ways in English, which might also be reason why students are rather cautious. Intercultural differences in academic writing certainly also play a major role in acquiring a suitable style of writing. For instance, German and Czech academic writing is said to be more direct than in English, i.e. hedges are less often applied in German and Czech (cf. Hyland & Milton 1997: 186). On the one hand, different countries have different writing styles – L2 students learn to write differently in their culture (cf. Hyland 1996: 482). On the other hand, this could be an artifact of the demand that science and academic writing is "intended" to have an objective point of view instead of a subjective one. An impersonal writing style may be more common in hard sciences, such as physics, mathematics and geology, than in soft sciences. Hyland (1996) raises concerns about an overtly impersonal and confident style of writing "because even the most assured propositions have an inherently limited period of acceptance" which may be risky in the academic world (Hyland 1996: 478).

Certainly, for a coherent and credible style of academic writing the inclusion of an adequate amount of hedging devices is essential and illustrates that the writer has understood the conventions and practices of taking stance in the academic world.

3. Methodology

3.1. Data collection

The data for the analysis was taken from the *ChemCorpus* which includes distinct levels of student writing, such as Magister theses, written Magister exams and Bachelor and Master theses term papers at Bachelor and Master level. The compiled corpus for the present analysis contained ten Bachelor and ten Master theses from the fields of English linguistics (eight BA and eight MA theses) and British and American Cultural Studies (two BA and two MA theses). The importance is that the ten BA theses were written by the same ten students as the selected ten MA theses. This ensures the comparability of possible changes of the usage of hedging devices from the BA to the MA level. Table 1 illustrates the word counts of the BA and MA theses. The total corpus size is 447,105 words. Each thesis is written according to the IMRD structure which means that they consist of an introduction, a literature section, a methodology, discussion section and a conclusion. The 10 selected students have a German background and studied English and American Studies at TU Chemnitz for their Bachelor as well as for their Master degrees. The BA theses were handed in between 2009 and 2011, whereas the MA theses were finished between 2011 and 2014. The topics of the theses as well as the gender of the writers were not of interest, since both were not a variable for the corpus analysis.

BA thesis	*word count*	*MA thesis*	*word count*
BA09Ft_UA	12,728	*MA13Ft_UA*	31,758
BA09Ft_DB	17,545	*MA12Ft_DB*	34,660
BA11Ft_LH	22,810	*MA14Ft_LH*	39,853
BA11Ft_AH	17,157	*MA14Ft_AH*	15,581
BA10Ft_JH	18,532	*MA13Ft_JH*	27,519
BA09Ft_SK	14,302	*MA13Ft_SK*	28,818
BA09Ft_CN	13,983	*MA11Ft_CN*	20,946
BA09Mt_SS	14,768	*MA12Mt_SS*	27,561
BA09Ft_CW	17,536	*MA13Ft_CW*	25,731
BA11Mt_PB	13,583	*MA14Mt_PB*	31,734
total	**162,944**		**284,161**

Table 1: Word counts of BA and MA theses

3.2. Data analysis

The texts were investigated with the concordance program *AntConc 3.2.1w*. The theses from the *ChemCorpus* (Schmied, this volume) are available as .txt files, which is the format that *AntConc* supports. Furthermore, in all 20 theses table of contents, tables, figures, charts, appendices, direct citations, references and pictures were omitted to ensure full running texts and to avoid biased results. The theses were uploaded to the program and parsed for the modal adjectives *presumable*, *probable*, *apparent*, *likely*, *unlikely* and *possible* and also for the modal adverbs *presumably*, *probably*, *apparently*, *seemingly*, *possibly* and *relatively*. The adjective *apparent* and the adverb *apparently* were analyzed in differently than the other adjectives and adverbs because both can have an epistemic and a non-epistemic meaning. *Apparently*/*apparent* in the sense of *evidently* or *obvious* convey a very certain writer commitment to the truth value of the claim, whereas a low degree of commitment is expressed by the adverb/adjective in the sense of *seemingly*. Sometimes these different usages become only clear when the sentence, they occur in, is regarded in context.

(8) […] it became apparent where misunderstandings and obscurities were most likely to occur and the concerned questions were amended accordingly. (MA14Ft_LH)

(9) It seemed that in many cases paramilitary organisations were *apparently* viewed as a possibility to show the disaffection with the current situation. (MA13Ft_CW)

Sentence (8) shows the usage of the non-epistemic meaning of the modal adjective *apparent*. Here *apparent* is used in the sense of *visible* which can be derived from the verb *became* in front of the adjective. In (9), the modal adverb *apparently* is obviously applied in its epistemic meaning of *seemingly*, which is emphasized by the lexical verb *seemed* at the beginning of the utterance.

The investigation of the epistemic adjectives and adverbs were supposed to find out:

i. whether there were any notable differences in the frequencies of the adjectives and adverbs between BA and MA level?
ii. in addition, the analysis is intended to shed light on whether students become more aware of the usage of hedges, i.e. apply the selected MA students more cautious language than in their BA theses?
iii. thinking back to Nuyts' (2001) scale of likelihood, is it possible to see which degrees of likelihood are preferred in ESL student writings?

I would predict that BA students are aware of the usage of hedges, but they may still express themselves in a more certain manner than two years later in their MA studies when a progress in the style of writing should have taken place. The following chapter will show whether this prediction can be verified.

4. Results

4.1. Epistemic adjectives

The results for the corpus analysis of the epistemic adjectives showed some intriguing findings. Unlike expected, the adjectives *presumable* and *probable* were neither found in the BA nor in MA theses except for one hit for *probable*, which can, however, be disregarded because this finding is by no means significant for the corpus study. Table 2 and Table 3 provide an overview of the absolute and relative frequencies (per 10,000 words).

Although *apparent* was applied in the BA theses, three of the four occurrences had to be eliminated for the analysis because these instances were used in the sense of *obvious* (see example (10)).

(10) The very popular music genre of rap provides access to *apparent* native like English which increasingly is reproduced by young people. (BA10Ft_JH)

In the MA theses, *apparent* occurred almost with same frequency (11 hits) as in the BA theses (13 hits). Contrary to the BA theses, *apparent* was applied in its epistemic (5 hits) as well as in its non-epistemic (6 hits) meaning. The most frequent occurrences were found for the adjective *possible* followed by *likely* and *unlikely*. Nevertheless, the overall hits for *possible* were higher in MA than BA theses. The non-epistemic uses of the adjective were higher in the MA theses (123 hits), whereas *possible* occurred 86 times in its epistemic meaning. For the BA theses, the results were differently: *possible* was found less often in the non-epistemic (21 hits) and more often in its epistemic sense (51 hits). *Likely* and *unlikely* were also more frequently used in the MA theses than in the BA theses. Overall, it can be observed that more epistemic adjectives were applied in the

MA theses (0.159 per 100,000 words) than in the BA theses (0.093 per 100,000 words).

Epistemic adjective	Non-epistemic usage (absolute frequency)	Epistemic usage (absolute frequency)	Overall occurrences (absolute frequency)	Epistemic usage (relative frequency)
presumable	0	0	0	0
probable	0	0	0	0
likely	0	34	34	0.34
possible	21	51	72	0.51
apparent	3	1	4	0.01
unlikely	0	7	7	0.07

Table 2: Absolute and relative frequencies (per 10,000 words) of epistemic adjectives in BA theses

Epistemic adjective	Non-epistemic usage (absolute frequency)	Epistemic usage (absolute frequency)	Overall occurrences (absolute frequency)	Epistemic usage (relative frequency)
presumable	0	0	0	0
probable	0	1	1	0.01
likely	0	56	56	0.56
possible	123	86	209	0.86
apparent	6	5	11	0.05
unlikely	0	11	11	0.11

Table 3: Absolute and relative frequencies (per 10,000 words) of epistemic adjectives in MA theses

4.2. Epistemic adverbs

In contrast to the analyzed epistemic adjectives, the selected adverbs were all found in the ESL corpus. However, the epistemic adverbs *presumably*, *seemingly* and *apparently* had the least frequent occurrences in the BA as well as in the MA theses (see Table 4 and 5 for the overall occurrences of epistemic adverbs). The frequencies for these adverbs were rather low, which renders it almost impossible to draw reliable implications for this study.

Epistemic adverb	Non-epistemic usage (absolute frequency)	Epistemic usage (absolute frequency)	Overall occurrences (absolute frequency)	Epistemic usage (relative frequency)
presumably	0	3	3	0.03
probably	0	25	25	0.25
relatively	0	46	46	0.46
possibly	0	16	16	0.16
apparently	4	7	11	0.07
seemingly	0	5	5	0.05

Table 4: Absolute and relative frequencies (per 10,000 words) of epistemic adverbs in BA theses

Epistemic adverb	Non-epistemic usage (absolute frequency)	Epistemic usage (absolute frequency)	Overall occurrences (absolute frequency)	Epistemic usage (relative frequency)
presumably	0	4	4	0.04
probably	0	65	65	0.65
relatively	0	67	67	0.67
possibly	0	26	26	0.26
apparently	9	18	27	0.18
seemingly	0	3	3	0.03

Table 5: Absolute and relative frequencies (per 10,000 words) of epistemic adverbs in MA theses

The most often applied adverbs were *relatively*, *probably* and *possibly* in the same sequence in BA and MA theses. Looking back at the number of hits for the epistemic adjectives *possible* and *probable*, it becomes apparent that *possibly* is less often used and *probably* is more often applied than the epistemic adjectives. In the MA theses, *relatively* (67 occurrences) and *probably* (65 occurrences) were almost equally often used. Contrary to the BA theses, where *relatively* (46 occurrences) was used nearly twice as much applied than *probably* (25 occurrences).

The case of *apparently* seems to be different than for the epistemic adjectives. Whereas *apparent* was more frequently found in its non-epistemic meaning in the BA and MA theses, the adverb *apparently* appeared twice as often in its epistemic sense in both sub-corpora than in its non-epistemic meaning. If we now compare the frequency of *seemingly* to *apparently*, which both have the same literal meaning in their epistemic sense, we see that *apparently* is more commonly applied in the BA and MA theses than *seemingly*. All in all, it can be said that more epistemic adverbs were used in MA theses (0.183 per 100,000 words) than in BA theses (0.102 per 100,000 words). In the BA theses, adjectives were less often used (0.093 per 100,000 words) than epistemic adverbs. The same is true for the usage of hedges in MA theses: epistemic adjectives occurred less frequent (0.159 per 100,000 words) than adverbs. In general, epistemic adverbs were more often used than epistemic adjectives in the corpus.

5. Discussion of the results

The results chapter of this paper has already provided an insight to the outcomes of this corpus analysis of ESL student writing. One major finding that also confirms Hyland's (1998) and Nuyts' results is that epistemic adjectives are less commonly used than epistemic adverbs (cf. Hyland, 1998; p. 134; Nuyts, 2001, p. 56). The reason for that lies in the fact that adverbs can take distinct sentence positions, whereas adjectives are bound to one position in an utterance. This "floating character" of adverbs does not affect "the meaning relation between clause and adverb" and makes it possible that writers can act more flexibly in their texts (Hyland 1998: 134). A further reason for more frequent application of epistemic adverbs may be that their word class is much more varied and bigger than that of adjectives. It is easier for the writer to choose from a huge pool of possible epistemic adverb expressions than being restricted to a syntactically-bound category of adjectives. The last factor might be the simplest one: the reason of language economy. It may be argued that writers usually prefer adverbs over adjective construction because they are shorter and grasp the epistemic respectively the subjective meaning towards the proposition in the fewest words imaginable. Figures 2 and 3 illustrate the number of occurrences of adjectives and adverbs in BA and MA theses per 10,000 words.

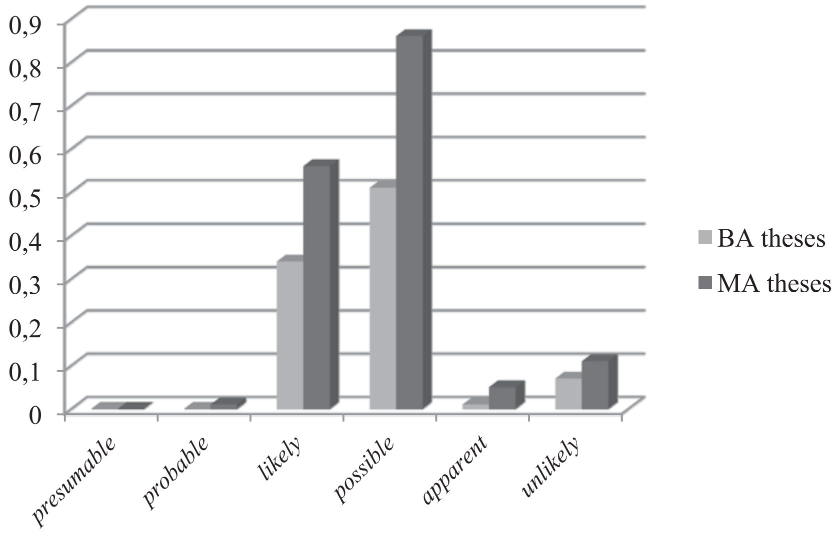


Figure 2: Usage of epistemic adjectives in BA and MA theses per 10,000 words

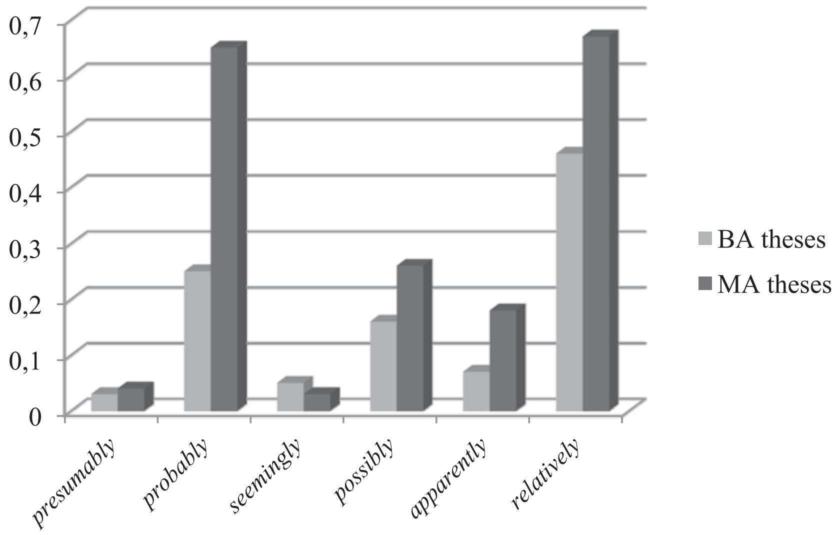


Figure 3: Usage of epistemic adverbs in BA and MA theses per 10,000 words

One further finding of the analysis was that the adjective *apparent* was applied more often in its non-epistemic meaning of *obviously*, but as an adverb *apparently* was more often found its epistemic sense. Although the overall frequencies of both, the adjective and the adverb, were not significant this tendency became quite visible. Sentence (11) shows the non-epistemic usage of *apparent* and (12) indicates the epistemic application of the adverb *apparently*.

(11) The origin of languages is depicted in a so-called language tree by means of which their respective historical roots become *apparent*. (BA11Ft_LH)

(12) With respect to Ireland, the dependency theory might *apparently* be the most suitable one because Ireland does clearly not belong to the group of core countries and its economical dependency to former colonial ruler Great Britain […]. (BA11Mt_PB)

Furthermore, the findings also implied that BA students employed the adjective and the adverb more often in its non-epistemic meaning than at their MA level. A possible explanation for these distinct usages might be that the style of academic writing progresses from the BA to the MA level, i.e. it may be probable that a student at the BA is unaware of *apparent(ly)* having two contrasting meanings. As Lyons (1977) points out the more certain usage of *apparent(ly)* may also have to do with the distinction of subjective and objective modality (cf. Lyons in Nuyts 2001: 60). The question is whether epistemic modality can be objective after all or whether it always is subjective by definition.

Looking at this differentiation from another angle, it can be assumed the adverb *apparently* is rather subjective implying that this hedge is more writer-oriented and confers the "I-say-so" meaning to the proposition. The adjective *apparent*, in contrast, may belong to the part of objective modality and is more content-oriented, which can be interpreted as the more factual "it-is-so" state of the proposition (cf. Lyons in Nuyts 2001: 60). Basically, the writer commits more to the facts than to their own opinion about the statement. To be able to compare and elicit more reliable findings for the above interpretation, a bigger corpus of ESL writing may be helpful for additional analyses.

The more tentative epistemic adjectives and adverbs, such as *presumable/presumably*, *seemingly* and *apparent(ly)*, were not as frequently used as the less tentative ones, e.g. *possible/possible*, *relatively*, *probably* and *likely*. The exception was *probable* which was found merely once in the corpus and can therefore be excluded from the interpretation of the results. Although hedges are used in the theses, the indirectness of a sentence can be expressed on a scale ranging from relatively certain to almost improbable. Nuyts' scale of the likelihood of an event was presented in chapter 2 of this paper.

(13) The metaphor use is *probably* motivated by the wish to highlight the competence of EU institutions and politicians, as well as their willingness to help. (MA13Ft_UA)

(14) Second, ScLV, a traditional feature of Scottish English, is *seemingly* not very common in contemporary urban speech of educated people. (BA09Ft_CN)

(13) and (14) demonstrate the two degrees of likelihood, namely the writer of the first utterance is turns the statement into a more direct one in comparison to the more tentative sentence (14). The writer of utterance (13) displays relatively certain that the reason for the metaphor usage was the emphasis of the EU's competence. In (14), the indirect *seemingly* illustrates that it could also be possible that ScLV can occur in educated people's speech although the findings imply something different. Both sentences were found in the discussion sections of the theses, which consist of the writer's subjective interpretations of results and findings. Naturally, these research paper chapters are hedged more often than for instance methodology sections. The usage of the less tentative adjectives and adverbs in the BA and MA theses would also confirm Hyland and Milton's assumptions that academic ESL writing is less hedged than native speaker writing (Hyland & Milton 1997: 186). However, if we take a look at the number of hedges used in the BA and MA theses, it becomes apparent that ESL writers

do understand that these devices are important for successful academic writing and they seemingly prefer less tentative ones over more tentative devices. The number of hedges is also higher in MA theses than in BA theses, which also indicates that the students use them more consciously than during their BA studies. The question whether ESL and native academic writing can be compared at all since students at a BA and/or MA level do usually not have the experience as "expert writers" of the academic community (cf. Hyland 1998: 218). The number of hedges that were investigated for this study does also not give a full range of the usage of tentative writing devices. To elaborate more detailed on the indirectness of ESL student writing, more hedging devices would have to be included in the corpus analysis.

Connected to the more "direct hedging" of the ESL students, is the following finding: the epistemic adjective *possible* is more commonly used at BA and MA level than its adverbial counterpart *possibly*. Though *possible*/*possibly* is more tentative than *probable*/*probably* the findings reveal something else. As already stated, *probable* was merely once used in the entire corpus, but the adjective *likely*, which can be regarded as a synonym for *probable*, occurred much more often in the corpus. In comparison to *possible*, *likely* had half of the occurrences.

(15) In analogy to the progress of singular they, it seems *possible* that themself will develop into an emancipated if not even the preferred form for gender-neutral antecedents one day. (MA11Ft_CN)

(16) […] it was looked for any striking indications and surprising statements in order to *possibly* reveal unexpected views, methods or other phenomena in the teacher statements […]. (MA14Ft_LH)

(17) Additionally, verbs of emotion are more *likely* to be employed by British magazines than by US American magazines. (MA12Ft_DB)

(18) The second victim was *probably* a Protestant, police stated that his murderers belonged to the UDA. (MA13Ft_CW)

(19) The pattern you and I is observed *relatively* frequently in Twitter discourse, which supports the notion that some that some speakers perceive I as a fixed expression. (MA12Mt_SS)

(15) and (16) show that *possible* and *possibly* are similarly used in the utterances. Both sentences show that there is a 50 per cent chance that something will happen as well as a 50 per cent chance that something will not happen. In (15), *possible* belongs to the sentence-initial construction *it is possible* in which *is* is substituted by *seems* making the claim even more tentative. In (16), *possibly* refers to the verb *reveal*, which could also be exchanged with *it is possible that unexpected views are revealed*. Sentence (17) implies a more direct finding by the use of the adjective *likely*. This is due to the fact that *likely* is connected to a real finding in a study, whereas (15) and (16) refer to future events and predictions. (18) also shows that *probably* is applied to rather emphasize a fact than an opinion, but still holds a certain percentage uncertainty about the victim being a Protestant. Another intriguing result is that the epistemic adjective *possible* was more often used than *likely*/*probable* in BA as well as MA theses, but the adverb *possibly* was less frequently applied than *probably* at both student levels. On the one hand, one could argue that *possible* is more popular since it

usually appears in the *it is*-construction in sentence-initial position where *possible* refers to the entire sentence viz. it highlights the adjective as well as the author's standpoint towards the proposition. On the other hand, *probably* can be used more flexibly within a sentence than an epistemic adjective and is therefore more likely to be used to refer to certain parts in an utterance, which are rather facts than subjective opinions, predictions, etc. This would again prove Lyons' (1977) view of the distinction into objective and subjective modality. However, it is not possible to assign epistemic adjectives and adverbs to either one of the categories. But rather it seems viable that adjectives and adverbs can be present in both classifications, i.e. adverbs and adjectives can be both – subjective and objective. Utterance (19) illustrates the usage of *relatively* which is applied as an approximator and is mostly found on a phrasal rather than a clausal or sentential level. *Relatively* in (19) works as a general downtoner without precisely specifying the frequency in which the construction *you and I* appears.

The results indicate that a more detailed and deeper research on hedging devices in ESL student writing has to be done. These preliminary results can be examined and verified on a much broader scale when taking additional hedging devices, such as lexical epistemic verbs or modal verbs, into account for the next corpus analysis.

6. Conclusion and outlook

The corpus analysis indicated that there are differences in the usage of epistemic adjectives and adverbs as hedging devices in academic ESL writing between the BA and MA level. Firstly, more epistemic adverbs than epistemic adjectives were found in the BA and MA theses, which may be due to the fact that the range of adverbs is much higher than that of adjectives. Secondly, the relatively free positioning of epistemic adverbs within a sentence might facilitate the usage of adverbs for ESL writers. In addition, the investigation showed that the students applied more epistemic adjectives and adverbs in their MA theses than in their BA theses. The implication could be that MA students are more aware of their academic writing style than during their BA studies and therefore use more hedges in their MA theses. Many scholars (Hyland 1998; Nuyts 2001 & 2005; Hoye 2005) have commented on the credibility of the author in their academic text(s) which is not only achieved by watertight results and argumentations, but also by the application of hedging devices. Since research findings are subject to change on a daily, monthly and/or yearly basis, it seems to be reasonable to avoid an overtly direct style of writing in the academic community. On the one hand, more tentative academic writing "shields" the writer from possible criticism towards their work. On the other hand, more indirect writing also acts as a face-saving method for peers and colleagues who work in the same field of research (cf. Huebler 1983; Hyland 1998). Upon a more detailed look at the epistemic adjectives and adverbs, a tendency toward the more direct and less tentative hedges became obvious. The students preferred the adjectives and adverbs *possible*/*possibly*, *relatively*, *likely* and *probably* over *presumable*/ *presumably* and *seemingly*. Although not all of the named adjectives and adverbs can be positioned at the same spot in the scale of likelihood, *presumably* and *seemingly* are on the left side of the scale towards a very tentative statement, *possible/possibly* and *relatively* may be put somewhere in the middle of this scale and *likely* and *probably* on the right end of the scale towards a more certain point of view. Since most of these hedges were found in the discussion sections of the theses, it may be viable to say that students hedge their own findings with less tentative devices. As said before, more important claims are said to be hedged less than less important ones. Basically, the students are convinced of their study and its results and intend to express that with stronger epistemic adjectives and adverbs.

This study was only able to provide a glimpse into the data of the *ChemCorpus*. However, for a follow-up study on German academic ESL writing it would be very intriguing to look at other hedging devices, such as lexical epistemic verbs (e.g. *assume*, *suggest*, *indicate*), modal verbs (e.g. *could*, *would*, *may*) and nominalizations of epistemic adjectives (e.g. *possibility*, *probability*, *likeliness*). Then tentative assumptions that have been made in this paper, for example that students prefer more direct hedges over less direct ones, could be even more precise.

References

Beyer, D. (2009). *Gender and hedging in academic writing: A comparison of gender-preferential usage of hedges in ESL learners' Magister theses* (unpublished Bachelor's thesis). Technische Universität Chemnitz, Chemnitz, Germany.

Bybee, J. L., Perkins, R. & W. Pagliuca (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world.* Chicago: University Press of Chicago.

Clemen, G. (1998). The concept of hedging: Origins, approaches and definitions. In R. Markkanen & H. Schröder (eds*.), Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts.* Berlin: De Gruyter, 237-248

Crompton, P. (1997). Hedging in academic writing: Some theoretical problems. *English for Specific Purposes, 16,* 271- 287.

Coates, J. (1983). *The semantics of the modal auxiliaries.* London: Croom Helm.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features.* Mahwah & London: Lawrence Erlbaum Associates Publishers.

Hoye, L. (1997). *Adverbs and modality in English.* London & New York: Longman.

Hoye, F.L. (2005). ''You may think that; I couldn't possibly comment!'' Modality studies: Contemporary research and future directions. *Journal of Pragmatics, 37,* 1481-1506.

Huebler, A. (1983). *Understatements and hedges in English.* Amsterdam & Philadelphia: John Benjamins Publishing Company.

Hyland, K. (1996). Nurturing hedges in the ESP curriculum. *System, 24,* 477-490.

Hyland, K. (1998). *Hedging in scientific research articles.* Amsterdam & Philadelphia: John Benjamins Publishing Company.

Hyland, K. & J. Milton (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing, 6,* 183-205.

Kratzer, A. (1978). *Semantik der Rede: Kontexttheorie, Modalwörter, Konditionalsätze.* Königstein/Taunus: Scriptor.

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2,* 485-508.

Nuyts, J. (1993). Epistemic modal adverbs and adjectives and the layered representation of conceptual and linguistic structure. *Linguistics, 31,* 933-969.

Nuyts, J. (2001). *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective.* Amsterdam & Philadelphia: John Benjamins Publishing Company.

Nuyts, J. (2005). The modal confusion: On terminology and the concepts behind it. In A. Klinge & H. Høeg Müller, (Eds.), *Modality: Studies in form and function.* London & Oakville: Equinox Publishing Ltd, 5-38.

Palmer, F. R., (1993). *Mood and modality.* Cambridge: Cambridge University Press.

Schmied, J. (2009). Discourse approaches to specialised and popular academic English: Comparing stance, hedges and modality. In M. Dynel (ed.), *Advances in discourse approaches.* Newcastle upon Tyne: Cambridge Scholars Publishing, 76-91.

Schmied, J. (2011). Academic writing in Europe: A survey of approaches and problems. In J. Schmied (ed.), *Academic writing in Europe: Empirical perspectives.* Göttingen: Cuvillier, 1-26.

Schmied, J. (2015). Graduate academic writing in Europe in comparison: A research-based approach to metalanguage and genre. This volume, 1-24.

Swales, J. (1991). *Genre analysis: English in academic and research Settings.* Cambridge: Cambridge University Press.

Van der Auwera, J. & V. Plungian (1998). Modality's semantic map. *Linguistic Typology, 2,* 79-124.

Van linden, A. (2012). *Modal adjectives: English deontic and evaluative constructions in synchrony and diachrony.* Berlin & Boston: de Gruyter Mouton.

Usages of *may* and *will* in Chinese and German Student Writings: A Corpus-Based Comparison

Jessica Küchler (Technische Universität Chemnitz)

Abstract

This study investigates the differences in the usages of the modal verbs *may* and *will* in Chinese and German student writings. Therefore, two corpora (the Sun Yat-sen University-Corpus and the Term Paper sub-corpus of the *ChemCorpus*) have been compiled and analyzed quantitatively and qualitatively. For both modal verbs, adequate categories have been decided on, which were not always clear-cut. All instances have been assigned to one category. Significant differences have been revealed by researching master and bachelor theses as well as term papers written by both groups of learners. Quantitatively analyses show that Germans are more confident in their academic writings when it comes to presenting their ideas, methods and findings, which can be seen in the low number of instances of epistemic *may* and the high number of occurrences of deontic *will* as well as the fairly high number of the 1st person singular subject in connection with deontic *will* compared to the Chinese writings, which feature higher instances of both epistemic modals whereby epistemic *may* was used twice as often.

1. Introduction

"Writing is fundamental to modern societies and of overarching significance in all our lives: central to our personal experiences, life chances and social identities." (Hyland 2010: 191) As my academic career depends on my (academic) writing skills, this topic is especially important for me as a Master student. An essential aspect in writing is the way authors convey ideas and how committed they are to the truth value of their arguments and statements. The best way to express this belief, probability, possibility, necessity and confidence is through modality and especially by using modal verbs. This paper, which is a summarized version of my master thesis, is going to analyze how the modals *may* and *will* are used in academic writings by German and Chinese students and in how far their usages and the epistemic stance of the authors differ.

Everyone has its own way of expressing ideas and putting forward arguments and opinions. However, there are various sources that influence the academic writing style of students and certain patterns that can be found in all student papers. As Germans and Chinese belong to very contrasting cultures and have probably enjoyed different basic and even advanced English education, it is particularly interesting to take a look at their academic writings. The modal verbs that have been chosen for this study are just as diverse as these two cultures. On the one hand, the usages of *may*, which expresses weak modality, and on the other hand the usages of *will*, expressing strong modality, will be researched and compared. Two corpora have been put together to provide the means of a comparative analysis. This enables us as researchers "to depict what is usual in a genre, rather than what is simply grammatically possible, and helps to suggest explanations for why language is used as it is in particular contexts." (Hyland, 2010: 199)

Both German and Chinese English (CE) belong to the expanding circle according to Kachru (2009), including those countries and areas where "English is an 'international language' and [that are] traditionally regarded as societies learning English as a Foreign Language (EFL)" (Kachru 2009: 292). Especially for CE, "the literature to date has largely been restricted to discussions of whether CE exists or how if differs from Chinglish." (Xu 2010: 4) Therefore, a systematic comparative research of this kind into the sociolinguistic and linguistic features of CE for both intra- and international communication is of significance. (cf. Xu 2010: 4)

This study seeks to investigate possible differences in the usages of the modals *may* and *will* in Chinese and German student writings. Before going into detail with the data that has been used and the methodology that has been applied in the second section as well as discussing the results in the third section, however, it is essential to take into account different aspects of modality.

The term modality and the categorization of modal verbs are a much discussed topic, but there is still not one fixed definition and categorization. There exist a lot of different explanations of modality by various authors like Bybee (1995), Biber et al. (1999) and Palmer (2009) to name but a few. According to Portner (2009: 1), who sees modality as a semantic theory, "modality is the linguistic phenomenon whereby grammar allows one to say things about, or on the basis of, situations which need not to be real." In academic writing, theories are explained and discussed, new concepts and ideas are introduced, texts are reflected upon and interpreted and certain aspects of language in various contexts are analyzed and compared. In all of these fields and topics, not only facts are stated, but possible events are discussed, predictions are made and probable reasons and outcomes are examined. Hence, modality plays an important role and modal verbs are often found in academic papers.

Since the author expresses his opinion and discusses findings in his paper, the statements and findings do not necessarily have to be true. To express a certain attitude, modal verbs like *can*, *should*, *may* and *will*, adjectives, nouns, hedges and adverbs like *maybe, probably* or *certainly* can be used. (cf. Biber et al. 1999: 483) Those can be put into three main types of meaning, namely deontic, epistemic (also referred to as 'intrinsic' and 'extrinsic') and dynamic. (cf. Huddleston 2002: 178) Again, there are different opinions on how to read and interpret modality. Whereas a lot of researches neglect the third category, all three types are considered in this study. Deontic "modality refers to actions and events that humans (or other agents) directly control: meanings relating to permission, obligation, or volition (or intention)" (Biber et al. 1999: 485). Epistemic "modality refers to the logical status of events or states, usually relating to assessments of likelihood: possibility, necessity, or prediction" (Biber et al. 1999: 485). Dynamic modality is "concerned with properties and dispositions of persons" (Huddleston 2002: 178) as well as their ability to do something. Different categorizations of *may* and *will* will be explained in more detail in section 1.2.

1.1. Research questions

Four major research questions have been developed in order to compare attitudes of German and Chinese students found in their academic writings:

RQ1: In how far do the usages of the two modals *may* and *will* differ in Chinese and German student writings?
RQ2: Do German L2 learners of English use deontic *will* more often than Chinese learners?
RQ3: Do Chinese L2 learners of English use epistemic *may* more often than German learners?
RQ4: Do German L2 learners of English set themselves more often as the agent in their writings when expressing (deontic) volition?

The first research question is more of general terms and includes all the categories that the modals have been put into. It aims at comparing both groups of students and at getting an overall picture of the differences in the usages of *may* and *will* in their writings. The following two research questions seek to analyze two types of usages of *may* and *will* in more detail whereas the last one seeks to research the subject of the sentence in order to find out which group of students uses the 1st person singular or plural subject more often.

Originally, this study strived towards analyzing if the culturally perceived tentativeness, indirectness and politeness of Chinese students are reflected in their academic writings through the high usage of the modal verb *may* as well as the perceived directness of German students through the high usage of the modal verb *will*. However, there is hardly any evidence for culturally perceived characteristics and differences and it is hence difficult and improper to draw any possible conclusions and come up with probable reasons. "The whole field of contrastive rhetoric and culture needs much more empirical research and rigorous conceptualization." (Schmied 2011: 13)

I nevertheless hypothesize that Chinese students use epistemic *may* more often than German students. Furthermore, I assume that there are more instances of deontic *will* in German student writings and that deontic *may* is generally underrepresented in both corpora. Moreover, I predict that Germans often set themselves as the agent when expressing deontic volition whereas this phenomenon will not be found in Chinese papers.

1.2. Categorization of *may* and *will*

In order to categorize the instances of *may* and *will* in both corpora, the approaches suggested by the *Longman Dictionary of Contemporary English* (LDOCE) and *The Cambridge Grammar of the English Language* (Huddleston 2002: 188 ff.) are regarded as the most appropriate ones. Categories that merely belong to spoken English as well as the ones not represented in my data have been neglected. Both approaches were combined to find the best way of classifying both modal verbs found in my corpora into categories. The lines were often blurred and certain categories overlapped. In the following, the applied categorization will be explained and suitable examples will be given below the explanation.

1.2.1. *MAY*

This sub-section will introduce the selection of categories for the modal auxiliary *may* and will give suitable examples from LDOCE and *The Cambridge Grammar of the English Language* (Huddleston 2002).

a) Epistemic: Possibility
This category applies if there is the possibility that something may be true or may happen, but is not really certain. (cf. LDOCE, s. v. may) It is about the writer's value judgment of the proposition. (cf. Portner 2009: 2) Here, *may* can generally be replaced with *might*.

- *Some chemicals may cause environmental damage.*
- *Your job may well involve some travelling* (=it is fairly likely).

b) Epistemic: Possible to do something
"If something may be done [or] completed in a particular way, that is how it is possible to do so" (LDOCE, s. v. may), this category applies. Here, *may* can be replaced with *can*.

- *The problem may be solved in a number of different ways.*

c) Epistemic: Concessive may
Here, *may* expresses "that even though one thing is true, something else which seems very different is also true" (LDOCE, s. v. may) and can be interpreted concessively. *May* is often preceded by *though*, *although* or *even though* and followed by *but*. (cf. Huddleston 2002:182)

- *I may be slow, but at least I don't make stupid mistakes.*
- *Although this may sound like a simple process, great care is needed.*

d) Epistemic: Purpose

In this category, *may* is used to say “that someone does something in order to make something else possible.” (LDOCE, s. v. may) It is typically used after ‘in order to’ and ‘so that’.

- *The hero sacrifices his life so that his friend may live.*

e) Deontic: Permission

This usage of *may* can be interpreted as “someone is allowed to do something” (LDOCE, s. v. may). Here, *may* can be replaced with *can*.

- *There is a set of rules to show what members may and may not do.*
- *No one may own more than 10% of the shares.*

1.2.2. *WILL*

This sub-section will present the selection of categories for the modal auxiliary *will* and will give suitable examples from LDOCE and *The Cambridge Grammar of the English Language* (Huddleston 2002).

a) Epistemic: Prediction

Will can be used to make confident predictions about the future. (cf. Huddleston 2002: 189) Since the writer is making a statement about something that is going to happen in the future, his knowledge is limited and the statement is not factual. (cf. LDOCE, s. v. may),

- *She will beat him easily.*
- *You will understand when you receive Kim’s letter tomorrow.*

b) Epistemic: Conclusion/ Belief

In this case, *will* is “used to say that you think something is true.” (LDOCE, s. v. will) This category is about explanations, conclusions, assumptions, beliefs and future confirmations. It “is normally restricted to 2nd or 3rd person subjects.” (Huddleston 2002: 188) Here, *will* can be replaced with *must*. Often, *will* is used in connection with *since*, *because*, *due to the fact that* and a like.

- *That will be Tim coming home now.*
- [Knock on door] *That will be the plumber.*

c) Epistemic: Conditional

This category displays the link between conclusion and predictions (sometimes also referred to as futurity). *Will* is most commonly used in *if*-clauses and sometimes in connection with *when* or no conjunction as in the examples below.

- *If they are here, they will be upstairs.*

- *Should this be the case, then they will not win the price.*

d) Deontic: Volition
Deontic modality is concerned with "the speaker's attitude to the actualization of future situations" (Huddleston 2002: 178). The subject-referent, who is on control, is willing to or ready to do something and his utterance shows implicature of commitment. (LDOCE, s. v. will) The speaker set themselves as a deontic source and make promises. (cf. Huddleston 2002: 178) The subject does not necessarily have to be a 1st person subject, especially in academic writing.

- *I will be back before six.*
- *The baby won't eat anything.*

e) Deontic: Obligation
This category is about making things obligatory or necessary. (cf. Huddleston 2002: 178) *Will* is typically used in connection with *have to* and *need to*.

- *You will have to clean the house when we get home.*

f) Dynamic: Habitual
Here, *will* is used to show "characteristics of habitual behavior of animates or general properties of inanimates […] [and] what always happens in a particular situation or what is generally true" (LDOCE, s. v. will). This category is sharply distinct from making predictions. (cf. Huddleston 2002: 194).

- *Oil will float on water.*
- *Accidents will happen.*

2. Data and Methodology

This section presents a detailed description of the data used for this (corpus-based) comparison and the methodology that was employed, namely a quantitative and qualitative comparative analysis of collected student writings.

2.1. The Sun Yat-sen University-Corpus

Within the framework of my study visit at the Sun Yat-sen University in Guangzhou and Zhuhai in fall 2013, data was collected for my Master thesis and the Sun Yat-sen University-Corpus (*SYSU-C*(orspus) henceforth) compiled. As I continued the data collection back in Germany, the papers range from 2010 until 2014. In total, the corpus consists of 186 student papers, whereas most them are MA term papers, and roughly 870,000 words. The exact distribution, text types and number of words can be found in Table 1 below.

***SYSU-C*orpus**	**average length**	**Texts**	**Words**
Master Theses (Linguistics)	16,900	25	422,535
Master Term Paper (Linguistics)	2,500	86	216,278
Master Term Paper (FL Teaching)	2,800	71	200,237
Bachelor Theses (Linguistics)	11,100	2	22,191
Bachelor Papers (Linguistics)	3,000	2	5,933
Total		**186**	**867,174**

Table 1: The *SYSU-C*orpus including respective word totals according to AntConc

Since the distribution of MA and BA papers is simply too unequal, it was not possible to compare those two text types within this corpus. Furthermore, the papers are all written in the field of Linguistics and Foreign Language teaching as "the balance between linguistic, literature, cultural studies and methodology specialists is difficult" (Schmied 2011: 17).

"In many English Departments male students are hard to find" (Schmied 2011: 17), that is why the corpus consists of a lot more papers written by female students than by male students. The same goes for the mother tongue. Even though the native language of people living in Guangzhou and Zhuhai (Guangdong Province) is Cantonese, most of the students at SYSU speak Mandarin as their mother tongue. Hence, both mother tongues are also very unequally distributed in the *SYSU-C*. However, as this paper aims at comparing Chinese and German academic writings, the obtained results are nonetheless significant, even without looking at sociolinguistic factors and text types as well as sections and year of publication.

2.2. The *ChemCorpus* term paper

The *ChemCorpus*, which contains writings by students who study English as their second language, has been compiled at Technische Universität Chemnitz (TUC) (cf. Schmied 2015 this volume). All papers have been produced by students in the Bachelor, Master and former Magister Program in English and American Studies. With over 250 texts and more than 1,800,000 words, the *ChemCorpus* consists mainly of theses and exams. Since term papers were needed for this current study in order to establish a good basis for comparison to the *SYSU-C*, I collected more data in Chemnitz and compiled a sub-corpus (of the *ChemCorpus*), referred to as the *ChemCorpus* Term Paper (CCTP henceforth). The papers were written between 2007 and 2014. In Table 2 below, the exact distribution, text types and number of words can be found.

***ChemCorpus* Term Paper**	**average length**	**Texts**	**Words**
Master Term Paper (various sections)	5,400	10	53,992
Bachelor Term Paper (various sections)	4,550	61	277,658
Bachelor Area Project (Linguistics)	5,460	22	120,105
Bachelor Area Project (Culture Studies)	5,630	67	377,446
Total		**160**	**829 201**

Table 2: The *ChemCorpus* term paper including respective word totals according to AntConc

As can be seen in Table 2, almost all the papers in CCTP are written by BA students and 86 % of them in the fields of Cultural Studies and Linguistics. The remaining 14 % were written in Literature, Foreign Language Teaching and American Studies. 62 % of the papers belong to the section labelled area project. Until last year, all students of English and American studies at TUC had to spend a semester abroad, which was usually in their fifth semester. During their stay abroad, all students had to do research on one specific topic, either on the language or the culture of their host country. As this is not merely a travel report but an academic paper at the very end of their Bachelor studies, those 99 papers were also taken into account for this study. Yet this is not an ideal basis for a corpus-based comparison since the *SYSU-C* consists mainly of MA papers written in the field of Linguistics. That is why the other sections have not been explicitly stated in Table 2 above. However, as has been mentioned before, without looking at sociolinguistic factors and different text types, the results are all the same highly significant and relevant.

2.3. Methodology

To analyze the usages of the modals *may* and *will* in Chinese and German student writings, data has been collected at two different Universities (at SYSU and TUC). Later on, all papers were anonymized and regularized, so that only the main body of the paper was left, and could be compiled as a corpus. In order to conduct a corpus analysis with the concordance program AntConc 3.2.4., all files had to be converted into text files. AntConc is a free toolkit by Laurence Anthony which offers different tools like a file view tool, concordance (plot) tool, a word list tool, key word list tool as well as a collocates and clusters tool. Considering that a great amount of data can be processed within a short period of time, this program is an easy and accessible tool for all students in empirical corpus linguistics.

Both corpora and both modal verbs were analyzed separately. Overall, 3637 instances of *will* and 1485 of *may* have been analyzed. For *will*, I also ran a search on *won't* and '*ll*. After excluding all the quotes, examples in the texts and the names *Will* and *May* as well as the month of *may* and the noun *will*, it came down to 3136 instances of *will* and 1305 of *may*. These instances were analyzed one by one whereby different aspects were looked at. Next to the types 'deontic', 'epistemic' and 'dynamic' as well as type descriptions such as 'deontic volition' and ‘epistemic possibility’, voice and aspect, subject, clause type and time were taken into consideration and noted down. Furthermore, I looked for collocations with mainly adverbs and conjunctions, so as to find possible structures and patterns. All findings were documented in an excel sheet.

Even though the corpora sizes do not vary greatly, the relative frequencies (the number of words per 100,000 words in this case) were calculated in order to have a proper comparison of the results. This was done for all above mentioned categories and different registers. The normalized numbers were then compared and the significant results were marked with an asterisk. Afterwards, the attained quantitative results were interpreted, which will be discussed in the next section.

Other statistical tests like the *t*-test or ANOVA’s were not possible, not applicable or not necessary because in this analysis, means of different groups do not play a role and no independent variables are involved. (cf. Rasinger 2013: 237) Looking at the relative frequencies or even the absolute numbers in this case already forms a good basis for a comparison that is sufficient for this study. The only test that would have been applicable here is *the chi-squared test of independence* (Rasinger 2013: 156). This “test is essentially based on the comparison of the observed values with the appropriate set of expected values” (Rasinger 2013: 157) and shows if our numbers are statistically significant by rejecting the null hypothesis. This can be found out by looking at and interpreting the *p*-value, the number showing the significance level. (cf. Rasinger 2013: 199) However, the chi-squared test only signals that something is significant, but it does not allow drawing conclusions on what is significant. Hence, it is sufficient to merely look at the relative frequencies as they reveal the same.

Another program that could have been applied is R-studio. R is used for statistical computing and data analysis where variables (names) are used instead of numbers. (cf. Gries 2008: 67) This program can easily read excel sheets where all the necessary information has been documented as it has been developed for this research. In our case, it could have been analyzed if, for example, females in their bachelor studies in China use the modal *may* more often than females in Germany or if Mandarin speakers use *will* more often in the field of Linguistics. However, my data does not render such an analysis as the collected papers are unequally distributed when it comes to gender, text types and sections as well as mother tongues in China. Nevertheless, the analysis revealed some significant differences which are discussed below.

2.4. Limitations

As Schmied (2011: 16) stated, the compatibility of data has been a “major problem of comparative research in academic writing”. It is crucial to keep in mind that the text types, meaning bachelor and master papers, and sections such as Linguistics and Cultural Studies as well as gender and mother tongue are not evenly distributed in both of my corpora. Hence, they are not ideally stratified, which could be the reason for minor skewed results. On top of this, the different genres and disciplines in which the papers have been written also influence the style of writing. (Matarese 2013: 178)

Moreover, the concordance program AntConc also has its limitations. When looking at the number of hits in the Key Word in Context (KWIC) tool, it is slightly different from the number of hits in the Word List tool. Apart from that, the program also included 'L1' when searching for instances of the abbreviation *ll* for *will*. Since all hits have been analyzed manually afterwards, this was not a major problem.

As the lines are often blurred, it was not always easy to distinguish between epistemic and deontic modality as well as prediction and conclusion. I tried to find some patterns and collocations (as discussed below) that make the distinction between these two categories clearer. However, other authors might still interpret some sentences differently.

Last but not least, the names of the files are rather unfortunate. As both Chemnitz and China start with the same letters, the files do too, which might lead to confusion at some point. Since the initial idea was to compare Cantonese and Mandarin speakers within the China-Corpus, the name of the files were not paid too much attention to yet. It was only later that the thought of comparing Chinese and German student writings was developed. The new name *SYSU-C*(orpus) seems much more appropriate, not only because the first letters are different, but because all the papers were collected at Sun Yat-sen University in Guangdong Province and not throughout China. Due to the amount of data and work, I kept the old file names of the data in the *SYSU-C* which start with CC (for China-Corpus).

3. Results and discussion

In this section, the obtained results are discussed quantitatively and qualitatively and examples from the two corpora are provided.

3.1. MAY

3.1.1. *Quantitative comparison*

Figures 1, 2 and 3 below show the number of overall occurrences of *may* in both corpora as well as according to the two different types of modality.

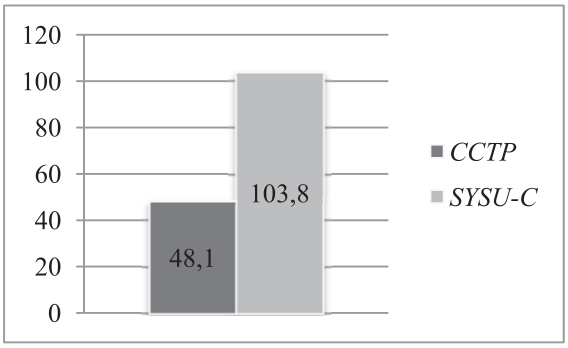


Figure 1: Frequency of *may* in the *ChemCorpus* TP and the *SYSU-C*orpus per 100,000 words.

As can be seen in Figure 1, there are more than twice as many instances of the modal verb *may* in the *SYSU-C*orpus than in the *ChemCorpus* TP whereby almost all of the occurrences can be interpreted epistemically.

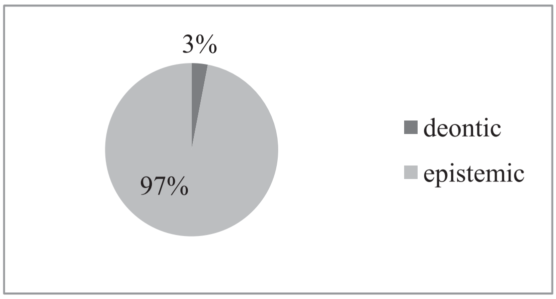


Figure 2: Frequency of *may* in the *ChemCorpus* TP (N=399)

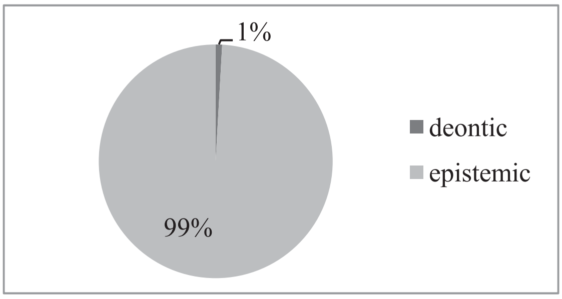


Figure 3: Frequency of *may* in the *SYSU-C*orpus (N=901)

Figure 2 shows that in the *ChemCorpus* TP, they make up 97%, and in Figure 3, it can be seen that epistemic *may* even makes up 99%. As has been expected, the deontic meaning of *may* is underrepresented in both corpora, as it is more common in spoken language.

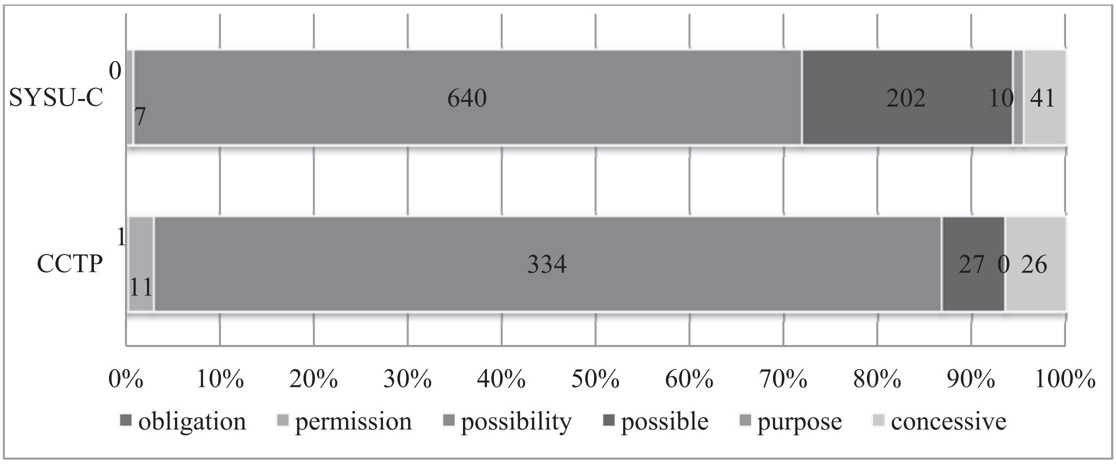


Figure 4: Semantic distribution of *may* in the different categories (N= 399; 901) in percent and per 100,000 words

The modal *may* is most frequently used in the category of epistemic possibility in both corpora whereby it occurs 42 times in the CCTP and 74 times per 100,000 words in the *SYSU-C* as shown in Figure 4. There are only two more categories worth mentioning, namely epistemic possible (to do something) and concessive *may*. With 23 instances, Chinese students used *may* seven times as often as German students to express that something can be completed in a certain way. Examples are given in section 3.1.2. below. Furthermore, concessive *may* was used by both groups of students whereas Chinese used it slightly more often with five instances per 100,000 words.

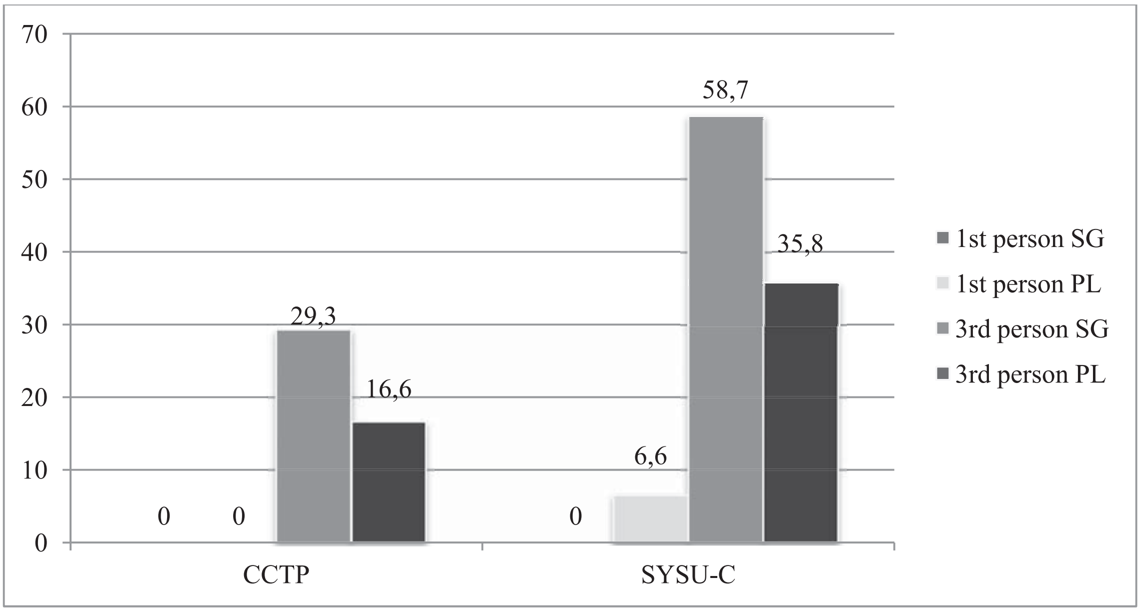


Figure 5: Syntactic distribution of sentence subject in *may-clauses* per 100,000 words

The first thing that is striking when taking a look at Figure 5 is that German students did not use *may* in connection with *I* or *we* at all whereas Chinese students used *we* 57 times, which is more than 6 times per 100,000 words. In both corpora, the 3rd person singular is most often the subject in sentences with the modal *may*. Here, 3rd person singular can be animate (*he* or *she*) or inanimate such as *it*, *this dissertation* or *the reason*. Furthermore, 2nd person singular as well as *one* and *there* have been used as the subject, but were left out of the analysis due to the small number of hits.

3.1.2. *Semantic analysis*

As can be seen from my results and as other researches like Biber et al. (1999) have already found out in their corpus analysis as well, "*may* is extremely common in academic prose, [but] rare in conversation" (487). My analysis revealed that Chinese students use the weak modal verb *may* more than twice as often as German students in order to express epistemic possibility and to show that it is possible to do something in a particular way. (LDOCE, s. v. will) As was expected, deontic permission is rarely found in academic writing. In the following, some examples are given to demonstrate the usages of *may* in both corpora:

(1) This may be one reason for the implementation of the New Economic Policy (NEP) in 1971. (CTP08FBAC_128)
(2) Graduation is concerned with the resources by which the force or intensity of an utterance may be raised or lowered. (CC11FMAT_3)
(3) They may lack time or knowledge necessary to assess technical arguments outside their own area of expertise, but they have an awareness of what they do and do not know, that issues are more complex than they are being told. (CC11FMATP_83)

(4) Maybe, it is possible to think that both systems may appear to be irrelevant to each other. (CTP12FBALJR_23)
(5) A possible reason for that may be that the teaching materials, rather than being based on empirical research, are possibly derived from native speaker intuitions, which may not be completely correct. (CC11FMAT_172)

In the first example, *may* clearly marks logical possibility. The author does not know for sure if this really is a reason for the implementation of the NEP, but there is a certain possibility that it is one. However, as the writer is not completely sure about it, he uses a weak modal to show his uncertainty. Example (2) states that something can be done in a certain way, so it is not so much about saying that it might happen, but that it is possible to raise or lower the force of an utterance. The third example shows a clear case of concessive *may* where the author concedes that they lack of time, but that they are still aware of what they are doing. Here, the writer admits something rather than saying that something might possibly happen. Sentences (4) and (5) are typical examples of hedging and not committing to the truth value of a statement. Both authors, a German and a Chinese one, used the modal verb *may* and the modal adjective *possible* as well as the modal adverb *possibly* in one sentence. This clearly shows that the author does not know if "the proposition is false and put(s) it forward as a possibility" (Huddleston 2002: 181). Collocations with modal adjectives and adverbs like *possibly*, *perhaps*, *presumably*, and *likely* are more often found in Chinese student writings and the two adverbs *certainly* and *clearly* were used more often by German students. When talking about possible events, causes or outcomes, the use of words tells us more about stance of the author to the statement. Hence, German student can be considered as more strongly committed to their ideas and utterances while Chinese students express a certain degree of tentativeness.

Schlosser (2012) claims that "the cause for deciding in favor of *may* might be found in the characteristic features of *may*. This modal could be considered as more formal to a certain group, and therefore appropriate in a written context." (49) This might be true for some of the instances of *may* in the two sets of data, however, most of the utterances clearly show uncertainty, lack of knowledge and tentativeness from the author. In the section of Foreign Language Teaching, some Chinese authors talk about methods that can be good and successful, but do not have to be good. By using the modal *may*, the writer show that they do not really believe in their ideas or proposed methods, but their cultural background does not allow them such negative statements. From the number of instances of *may* in their writings, it can be concluded that Germans are generally more confident and certain towards their statements whereas Chinese are rather tentative.

3.2. WILL

3.2.1. *Quantitative comparison*

Figures 6, 7 and 8 below show the number of overall occurrences of *will* in both corpora as well as according to the two different types of modality.

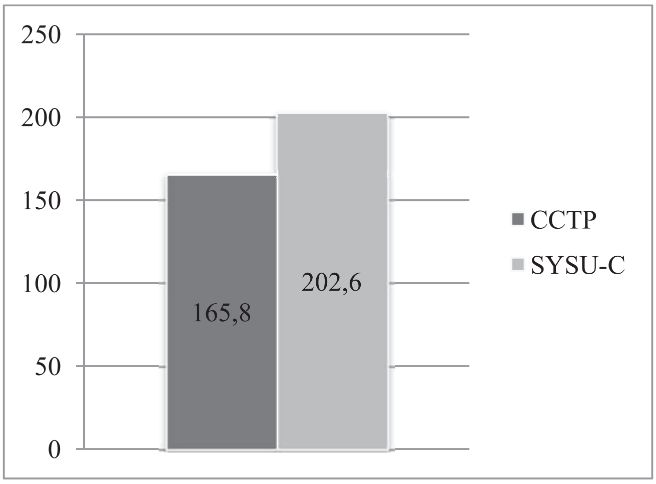
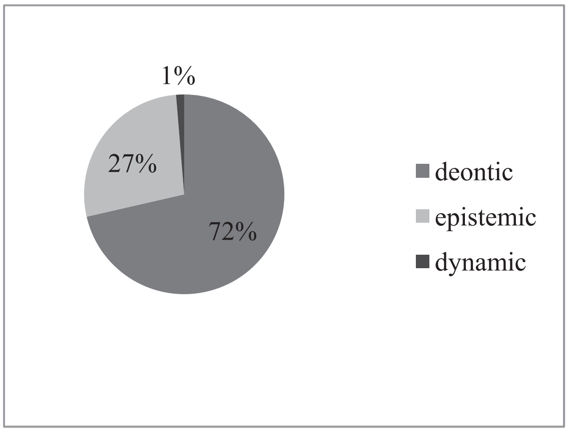


Figure 6: Frequency of *will* in the *ChemCorpus* TP and the *SYSU-C*orpus per 100,000 words



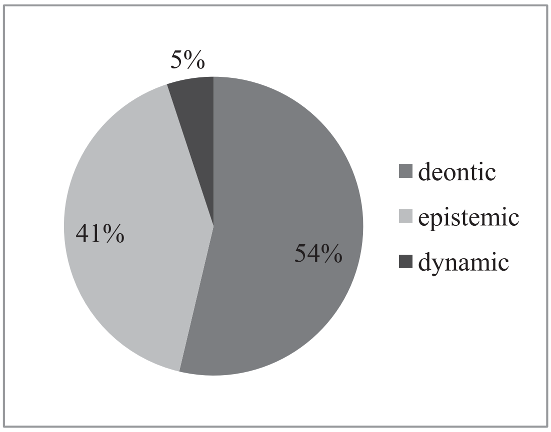


Figure 7: Frequency of *will* in the *ChemCorpus* TP (N=1375)

Figure 8: Frequency of *will* in the the *SYSU-C*orpus (N=1757)

Unlike hypothesized, the modal verb *will*, often associated with clear prediction, is more often found in academic writings by Chinese Students, namely 202 times per 100,000 words, than in writings by German students, as Figure 6 shows. However, when looking at Figure 7 and 8, it can be seen that over 70% of the instances of *will* are deontic in the CCTP, while only half of the instances in the *SYSU-C* are deontic. Out of 1375 occurrences of *will*, 982 are used to mark volition or obligation in German student writings (118 times per 100,000 words).

Dynamic modality which is used to describe habitual behaviour in this case, occurs five times more in the *SYSU-C* with about 10 hits per 100,000 words.

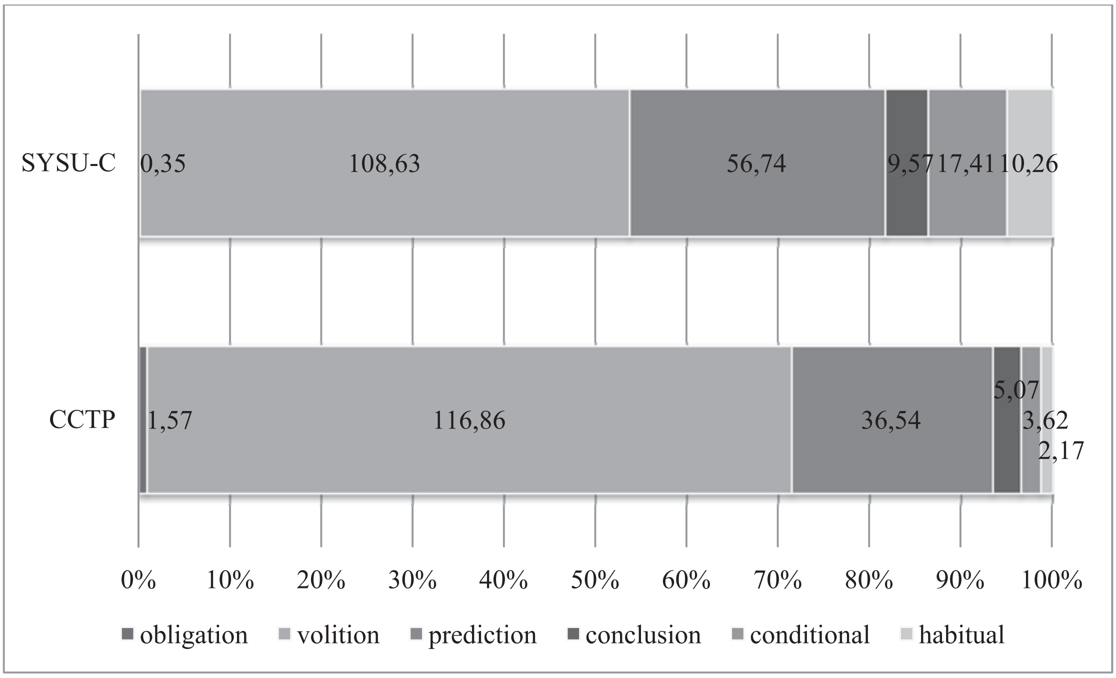


Figure 9: Semantic distribution of *will* in the different categories (N= 1375; 1757) in percent and per 100,000 words

In both corpora, *will* is more often used deontically as can be seen in Figure 9. In the CCTP, *will* marks volition in almost ¾ of the cases whereas in the *SYSU-C*, *will* only has a deontic function in half of the cases. However, only two categories are actually worth considering. Deontic volition and epistemic prediction are most commonly used in both sets of student writings while deontic obligation is almost non-existent and epistemic conclusion is the second least represented category. The remaining categories are more frequently used by Chinese students whereby epistemic conditional is fairly often used in comparison to German students.

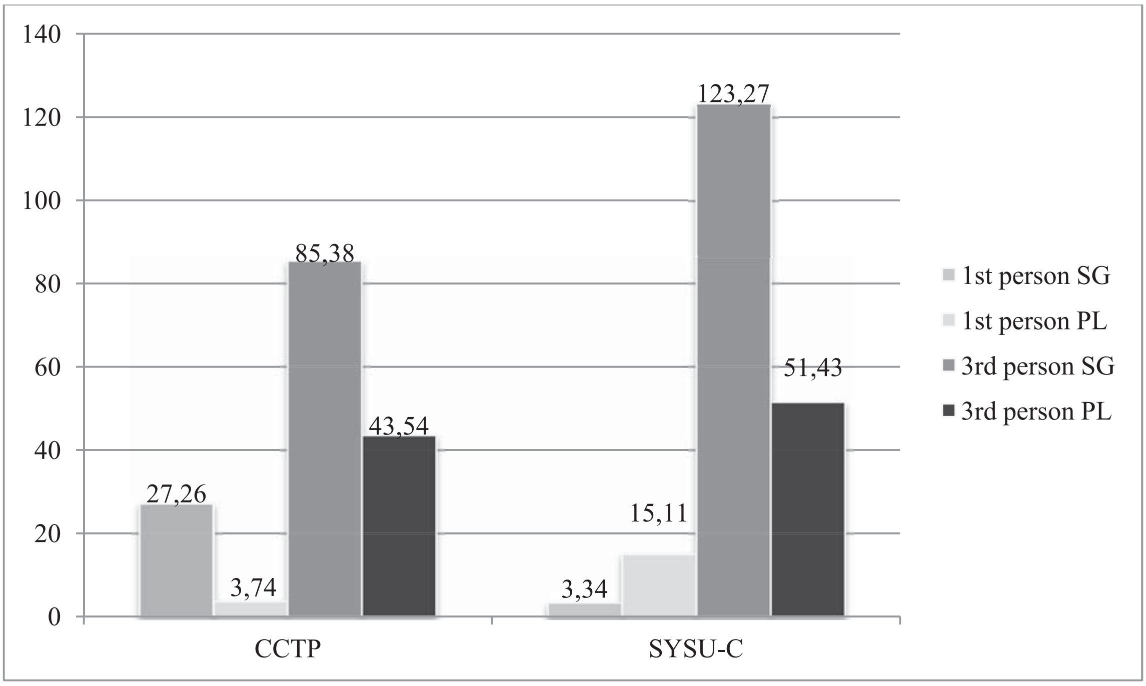


Figure 10: Syntactic distribution of sentence subject in *will-clauses* (N= 1375; 1757)

Figure 10 shows that 3rd person singular, followed by 3rd person plural, subjects are just as prominent in sentences with *will* in both corpora as they are with *may* (Figure 5 above). However, what is worth mentioning here are the instances of 1st person singular and plural: Whereas Chinese students seldom use *I will*, the 1st person singular subject is nine times more common in German student papers. In Chinese writings, *will* is more often preceded by the royal *we* than it is in German writings, but this is determined by local writing conventions. Due to the small number of instances, the 2nd person singular subject as well as *one* and *there* have been neglected in this figure.

3.2.2. *Semantic analysis*

According to Hyland (2010: 199), "*frequency* is based on the idea that if a word, string or grammatical pattern occurs regularly in a particular genre or subset of language, then we can assume it is significant in how that genre is routinely constructed". When looking at the results of the categorization of *will* in my two corpora, it seems surprising that more instances of *will* are found in Chinese academic writings than in German ones. Before proposing possible reasons, some examples of *will* in both corpora will be given:

(6) Some experts predicted that a few years later this number will become even larger than the total population of English-speaking countries.(CC00FBATP_182)

(7) Possibly, this research project will provide some additional findings to the broad theme of climatic, environmental influences on learning behaviour and study performance. (CTP08FBAL_152)

(8) If we carry out the research in a larger scale, we will find that context also has close relation with the three systems of Appraisal. (CC11FMAT_1)

(9) In this section, the author will firstly define two concepts: identity and investment. (CC11FMAT_173)
(10) The findings will attempt to answer the following questions. (CTP00FBAC_112.txt)
(11) **When** English is spoken in different countries or regions, it **will** naturally be influenced by the local language and culture. (CC11FMATP_148)

(6) and (7) are clearly examples of epistemic prediction. The "meaning of prediction created by the modal *will* is closely related to the speaker's confidence in the truth of the proposition" (Lihua 2010: 118). In the first sentence, this becomes obvious because the author even used the verb *predict* and talks about the number becoming larger in a *few years*, which is obviously in the future. The second author is relatively sure that his project will provide more findings, but he weakens his proposition by using the modal adverb *possibly*. As both events are not going to happen for sure, they are examples of epistemic modality.

Will is "often used in the apodosis of conditional constructions, whether the time is past, present, or future" (Huddleston 2002:191). In (8), *will* predicts the outcome (usually consequences) of the condition that we carry out a larger research. As has been shown above, the category of conditional consequences is more often found in my Chinese data. Most commonly, conditions are expressed in *if*-clauses. Sometimes, however, *if* was replaced by *when* or left out completely, in which case the sentence began with *should*.

In examples (9) and (10), *will* shows deontic volition, which is most commonly used by both learner groups. Here, the author set themselves as the agent of the sentence who is intending to define two concepts. This structure is particularly found in Chinese papers, but also common in the German writings. Instead of directly stating that "I am going to analyze something", a lot of students talk about *the paper*, *the data* or *the next section* that "is going to interpret and discuss the results" in order to avoid talking about oneself, as can be seen in (10), a borderline case with findings in the usual AGENT position. According to Biber et al. (1999: 499), *will* "often overtly express[es] the agent of the main verb […] (usually *I* or *we*)". As can be seen in Figure 8 above, German students used *I* as the subject- referent 226 times (26 times per 100,000 words) whereas Chinese students only used it 29 times (3 times per 100,000 words). In comparison to that, Chinese tend to use the 1st person plural subject as the agent when making promises or intending to do something, while Germans rarely use a 1st person plural subject. It can be concluded that Chinese are more collective people than Germans. Nonetheless, it must be noted that both student groups tend to avoid directly talking about themselves as the subject by using an impersonal subject, probably due to the lack of confidence.

In some cases, *will* is neither used to show volition nor to mark prediction. As can be seen in the last example (11), *will* can also describe things that generally happen. The author is not predicting that English is going to be influenced by the culture when spoken in a different country, but he rather talks about something that can be observed every time and hence can be considered as habitual. This category is more often found in Chinese writings, but is generally underrepresented in academic writing.

Since the number of instances of deontic *will* in Chinese papers was surprisingly high, I took a closer look at these papers again. Chinese students conventionally have a small introductory section before each section of their paper where they list what they will focus on and what the reader is going to hear in the following paragraphs. If this is done for all the main parts in a lot of papers, the number of *will* occurrences adds up easily. This characteristic is not found in German papers. Furthermore, one author overused *will* in his/her master thesis with 79 hits per 16,886 words. Strangely, this author speaks of his dissertation and research as something that is going to happen in the future and not of an analysis that he or she has already conducted. These could be two reasons why there are so many instances of *will*, mainly deontic *will* marking volition, in Chinese academic writings.

4. Conclusion

In this article, preliminary results of my master thesis have been presented and discussed. Due to the unevenly stratified corpora, the academic student writings were merely analyzed and compared according to their usages of the modals *may* and *will*. It was not possible to analyze differences between gender, text types, disciplines and mother tongues. Nevertheless, the findings are significant and show great differences in how Chinese and German students convey confidence through using the modals *may* and *will* in their writings.

For the purpose of this study, both *may* and *will* have been divided into semantic classes. This categorization was not always clear- cut, but finally, I decided on five categories for *may* and on six categories for *will*. Some of the categories overlap, especially those of *will* as the borders are often blurred and most of the categories involve futurity in some way. All instances of *may* and *will* have been assigned to one category during the process of data analysis. Afterwards, the relative frequencies have been compared and the findings discussed.

As hypothesized, Chinese students use *may* more often than German students; in this case twice as often. This shows that Chinese are generally less confident when conveying ideas and drawing conclusions in their academic paper. Through the higher usage of *may* in the category of epistemic modality in connection with modal adverbs, the writer marks a lower "degree of certainty about the validity or truth of the statement" (Lihua 2010: 123).

It has furthermore been found out that Germans use the 1st person singular subject *I* more often than Chinese when showing personal volition whereas Chinese incline to use the collective *we* as the sentence subject. In addition, both groups, especially Chinese students, often used an impersonal 3rd person subject such as *the paper*, *the data* or *this study* to avoid directly talking about themselves as the deontic source of the proposition, which can result from a lack of confidence in their research.

In contrast to Biber et al.'s (1999) finding that *will* "rarely marks personal volition" (489) in academic writing, it can be summarized that both Chinese and especially German students used deontic *will* often to show personal volition. "Very interestingly, the confidence conveyed by the use of *will* is manifested by some adverbials projecting certainty" (Lihua 2010: 118) and in some cases weakened by adverbials portraying possibility, which can mostly be found in Chinese papers. In the case of the high number of deontic *will* in Chinese papers, I proposed that they resulted from the summary parts before each section that introduce what will follow.

All in all, it can be summarized that Germans are more confident in their academic writings when it comes to presenting their ideas and findings, which can be seen in the low number of instances of epistemic *may* and the high number of occurrences of deontic *will* as well as the fairly high number of 1st person singular subject compared to the Chinese writings. Further research could focus on the differences between writings of male and female students as well as differences in the different specializations, such as Linguistics and Literature. Even though it will certainly not be easy, further studies could aim at revealing differences between Cantonese and Mandarin speakers in the field of academic writing.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999). *Longman grammar of spoken and written English.* London: Longman.

Biber, D., Conrad, S., Finegan, E., Johannson, S. & Leech, G. (2007). *Longman student grammar of spoken and written English*. London: Longman.

Bybee, J.L., Fleischman, S. (1995). *Modality in grammar and discourse*. Amsterdam: John Benjamins Publishing Company.

Gries, S.T. (2008). *Statistik für Sprachwissenschaftler*. Göttingen: Vandenhoeck & Ruprecht

Huddleston, R. (2002). Mood and modality. In R. Huddleston, & G. K. Pullum (eds), *The Cambridge grammer of the English language*. Cambridge: Cambridge University Press, pp. 172-208.

Hyland, K. (2010). Researching writing. In: B. Paltridge & A. Phakiti (eds.). *Continuum companion to second language research methods.* London: Continuum, pp. 191-204.

Kachru, B. B.; Kachru, Y.; Nelson, C. L. (2009): *The handbook of world Englishes*. Oxford: Blackwell Pub.

Küchler, J. (fc). *Usages of may and will in Chinese and German student writings: A corpus-based comparison.* (unpublished Master's thesis). Technische Universität Chemnitz, Chemnitz, Germany.

Lihua, L. (2010). *Interpersonal Rhetoric in the Editorials of China Daily: A Generic Perspective*. Bern: Peter Lang AG.

Matarese, V. (2013). *Supporting Research Writing: Roles and Challenges in Multilingual Settings*. Oxford: Chandos Publishing

Palmer, F. (2001). *Mood and modality* (2nd ed.). Cambridge: Cambridge University Press.

Portner, P. (2009). *Modality*. Oxford: Oxford University Press.

Rasinger, S. M. (2013). *Quantitative research in linguistics: An introduction*. London: Bloomsbury Academic.

Schlosser, T. (2012) *Ethnic variation of modal auxiliaries in Malaysia* (unpublished Magister thesis). Technische Universität Chemnitz, Chemnitz, Germany.

Schmied, J. (2011). Academic writing in Europe: A survey of approaches and problems. In: J. Schmied (ed.), *Academic writing in Europe: Empirical perspectives.* Göttingen: Cuvillier, 1-22.

Schmied, J. (2015). Graduate academic writing in Europe in comparison: A research-based approach to metalanguage and genre. This volume, 1-24.

Xu, Zhichang (2010): *Chinese English. Features and implications*. Hong Kong: Open University of Hong Kong Press.

Schmied, J. (2015). Graduate academic writing in Europe in comparison: A research-based approach to metalanguage and genre. This volume, 1-24.

Xu, Zhichang (2010): *Chinese English. Features and implications*. Hong Kong: Open University of Hong Kong Press.

Code Glosses in Student Writing: A Comparative Study of Albanian and German BA Theses

Jacinta Edusei (Technische Universität Chemnitz)

Abstract

Academic texts "seek to inform readers of activities, objects, or people in the world, to persuade them to some action or thought, or seek to promote the writer's scholarly claims and credentials" (Hyland &Tse 2004: 25). As a result, it is very important that the intended meaning of a text is conveyed appropriately to readers. To ensure this, writers employ certain linguistic features known as interactive metadiscourse markers to organize ideational information in a way that will be easily comprehensible to the reader. Code Glosses are one of such interactive metadiscourse markers. By providing elaborations in the form of examples and/or reformulations, academic writers anticipate readers' understanding ability and need for clarification of propositions. This paper investigates the use of code glosses with exemplifying and reformulating functions in student writing. The corpus comprises 14 Bachelor of Arts (BA) theses of non-native speakers of English in Albania (i.e. students of English at University of Ismail Qemali, Vlore, Albania) and Germany (i.e. students of English at Chemnitz University of Technology). The quantitative analysis shows significant differences in the usage of code glosses between German and Albanian students. The results also show a high frequency in reformulation markers than exemplification markers.

1. Introduction

Academic texts "seek to inform readers of activities, objects, or people in the world, to persuade them to some action or thought, or seek to promote the writer's scholarly claims and credentials" (Hyland & Tse 2004: 25). Academic persuasion is made possible with the help of certain rhetorical strategies. Research into academic discourse has shown generic, disciplinary and linguistic variations in rhetorical strategies. Moreover, different disciplines and genres have been known to draw on different linguistic resources for the creation of knowledge (Hyland 2007: 284). Linguistic devices that assist writers to organize propositions and present them in a way that will be easily understood by readers have been referred to as Metadiscourse markers. Literature on metadiscourse and specific metadiscourse aspects (transitions, endophoric markers, hedges, attitude markers, engagement markers) reveals research in the student writing genre: Markkanen et al. 1993; Crismore et al. 1993; Wagner 2011; Povolná 2013; Nkemleke 2011. There is however no research on the use of Code Glosses in the student writing genre neither has there been a comparison of Albanian and German students writing.

In academic writing, it is very important that the intended meaning of a text is appropriately conveyed to readers. To ensure this, writers elaborate by providing examples and/or rephrase what has already been said in other words. Linguistic devices that provide exemplification and reformulations have been collectively referred to as *Code Glosses* (Hyland 2005; 2007). In other words, code glosses can be defined as linguistic devices that "supply additional information by rephrasing, explaining or elaborating what has already been said to ensure the reader is able to recover the writer's intended meaning" (Hyland 2007: 268)

The rationale for the study of code glosses is that they contribute to the coherence and reader-friendliness of the text. BA students are at the progressive phase of academic writing and they aspire for acceptance in the academic community. Despite being 'academic novices', they still have to persuade readers and the entire academic community of the truth of their claims. The use of code glosses demonstrates the extent of the student writer's awareness of readers, and their perception of the message. It also shows the academic novice's negotiation ability in a bid to persuade by facilitating the understanding of the intended message through illustrations and reformulations. As Hyland (2007: 284) puts it, "these small acts of elaboration thus convey clear disciplinary meanings where what counts as convincing argument and appropriate tone is carefully managed for a particular audience''. He further observes (ibid.) that "the Humanities and social science texts are more explicitly interpretative, using reformulation to draw implications for a potentially more diverse readership and exemplification to shift between abstraction and concreteness, moving from conceptual to real world reference". Biber et al. (1999: 884) argue that exemplification in academic writing is often introduced by *for example* and *for instance*, whilst reformulation is often introduced by in other words.

In this paper, we investigate whether differences in exemplification and reformulation strategies exist among Albanian and German students writing. We investigate how true Biber et al. and Hyland's assertions are in the case of student writings of non-natives speakers of English from the same genre and discipline but different linguistic (L1) and cultural backgrounds. Furthermore, we follow Hyland's (2007) categorization of code glosses according to two broad subfunctions: reformulation and exemplification.

2. Definition of code glosses/ code glossing

The online Oxford Dictionary provides two main definitions of 'gloss' in English; the second of which is suitable for our study:

a) as a noun – a translation or explanation of a word or phrase
b) as a verb – provide an explanation, interpretation or paraphrase for (a text, word, etc.)

However, the linguistic definition of the term is more extensive: Crismore & Farnsworth (1990: 124) define 'code glossing' "as the act of defining or explaining words, phrases or idioms that may be judged to be problematic for readers". Code glossing therefore involves the anticipation of the reader's need for clarification. Code glosses fall under interactive Metadiscourse markers (Hyland & Tse 2004) amidst transition markers, frame markers, endophoric markers and evidential. According to Hyland (2007: 268), the term 'code gloss' "represents a number of basic communication strategies used in the negotiation of meaning many different contexts, occurring in both spoken and written language, to facilitate the reader's understanding".

From the definitions above, we could summarize the reason for code glossing as follows:

1. facilitate the understanding of the reader;
2. ensure that the intended message of a text is appropriately conveyed;
3. contribute to the cohesion of discourse;
4. buttress a point for emphasis.

Exemplification and reformulation in academic writing may be signaled by linking adverbials, abbreviations, punctuations, such as *for example*, *for instance*, *such as*, *that is/that is to say/which is to say*, *namely*, *specifically*, *e.g.*, *i.e.*, *parentheses* (Blakemore 1993; Hyland 2004, 2007). Biber et al. (1999: 876) refer to most of these as appositive linking adverbials, which can be used to show that the second unit is to be taken as a restatement of the first, reformulating the information it expresses in some way or stating it in more explicit terms.

Swales & Feak (2012: 37) list *in other words, that is* and *i.e., for instance*, *for example* as 'some of the more common linking words and phrases' used to connect sentences by introducing a complete sentence or independent clause. They further describe *in other words*, *that is* and *i.e.* as possessing clarification functions whilst *for example* and *for instance* have illustration functions. Hůlková (2011:132,134) uses the term 'conjunctive adverbials' for most of the lexical devices in this study (see Table 1). She also identifies exemplification and reformulation as the sub-categories of appositional conjunctive adverbials (one of the six semantic categories) (ibid. 138). Greenbaum (1969), Quirk et al. (1985: 417), and Greenbaum & Quirk (1990: 162) referred to them as 'conjuncts' whilst Huddleston & Pullum (2002: 775) and Leech & Svartvik (2002: 187) called them 'connective adjuncts' and 'sentence adverbials' respectively. In this study, we use Hyland's (2007) term 'exemplification and reformulation markers'.

Exemplification markers	***Reformulation markers***
such as	Parentheses ( )
for example	i.e.
e.g.	that is/that is to say/which is to say
for instance	in other words
an example of	namely
like	called
say	referred to as
	particularly, in particular
	specifically
	especially
	which /this means

Table 1: Code glosses with exemplification and reformulation functions

2.1. Reformulation

Reformulation implies the restating of propositions in different words. According to Hyland (2007: 274), the reformulations have two main discourse functions: expansion and reduction. A writer expands his original proposition by presenting an explanation (1), (2), (3), (4) or an implication (5), (6), (7), (8); as the examples from the Albanian (ALC) and the German (BA) corpora illustrate:

(1) Reading is often referred to as a complex cognitive process. (BA10Ft_KM)
(2) Online interaction, that is, text-based instantaneous communication, allows learners to share ideas and receive responses immediately in real-life chats. (ALC001)
(3) Even though, children make many grammatical mistakes while acquiring the native tongue, they are able to sort out the received input and produce own sentences and build coherent speech. The inherent brain structure making this possible is called 'Language Acquisition Device' (LAD). (BA10Ft_JM)
(4) The Standard English is also the English that is learn in schools of Europe […] In England, it also has an accent associated with it, known as Received Pronunciation: (RP). (ACL010)
(5) Moreover, the state-of-the-art technology in speech recognition requires the setting of parameters for each individual speaker, that is to say that a system trained to understand one voice may have difficulties to cope with another. (BA09Ft_CN)
(6) It is appropriate to mention, that English is studied mainly as the first foreign language. This means that most of the European pupils have to choose two compulsory foreign languages. (ALC010)
(7) The learner reexamines and edits these text-based forms to make the interaction more meaningful and comprehensible. In other words, learners are more aware of the language structures that they and their peers use to com- pose messages. (ALC001)
(8) Reading is also interactive in the sense that many skills work together simultaneously in the process [...] In other words, a reader must shift from one focus to another in order to understand the sense of the written text. (BA10Ft_KM)

In examples (1), (2), (3) and (4), *referred to*, *called*, *known as* and *that is* introduce reformulations (second unit) of the preceding proposition (first unit).These reformulations provide clarification for technical terms or a technical term for a concept already expressed. For instance, in (1) and (2), the markers introduce a definition for ‘Reading’ and ‘Online interaction’. In (3) and (4), the reformulations provide more accessible terms for concepts already expressed in the sentences preceding *called* and *known as*.

Despite the notion of equivalence between the first and second units, the reformulations in (5), (6), (7) and (8) seem to sum up the main ideas the writers want to convey to readers. The reformulation markers used here (*that is to say*, *this means*, *in other words*) introduce implications of the preceding statements.

The other discourse function of reformulation, reduction, implies a restriction of the meaning of the original proposition by a paraphrase: (9), (10) or specification (11), (12).

(9) Likewise, if the original Albanian word order can be retained in English and give acceptable translated sentences, there is no need for syntactic analysis. In other words, analysis of the source language is determined strictly by the requirements of the target language. (ALC004)

(10) The second phase of childhood (age 6 to 12), according to Montessori, is widely known as the sensitive period (BA10Ft_JM)

(11) For them, George represents the past (specifically, American ideals of the past); Nick, the threatening future (perhaps Communism); Martha, the primitive and pagan instinct; and Honey (the emotionally unstable daughter of a preacher), the failure of religion. (ALC008)

(12) The area comprises the northern and western parts of Scotland, particularly the Highlands and Islands. (BA09Ft_CN)

As illustrated in examples (9) and (10), reformulations also function as restatements of ideas in different words. The reformulations introduced by *specifically* and *particularly* in (11) and (12) present detailed features of the preceding ideas thereby, narrowing the scope and restricting the meaning of what has already been said.

2.2. Exemplification

Like reformulation, exemplification plays a key role in academic persuasion. Exemplification is where the meaning of a proposition (first unit) is clarified by another proposition (second unit) which illustrates or supports the original. Killingsworth & Gilbertson (1992: 153) explains that examples “lead readers to a correct version of the sense experience and save them the trouble of having to do their own ‘translating’ of abstract material into language more immediately accessible to the senses”. This means that examples assist a writer to present a more concrete reference that clarifies a generalized idea. According to Hyland (2007: 279), examples in soft knowledge fields like Humanities are brief and practical expressions that illustrate and support a prior locution in three main ways:

1. offering an instance of a general category: (13) (14),
2. providing a parallel or similar case: (15) (16),
3. giving a precept or rule: (17) (18).

(13) Those in a position of power such as teachers, ministers of religion, bankers and government officials were mainly from Ontario. (BA09Mt_SS)

(14) These radical changes left many writers like Albee confused and discontent as values and standards changed. (ALC008)

(15) The languages which belong to the family of Indo-European languages are related to each other. For instance, German, English and Italian are related: they all belong to Indo-European. (BA09Ft_UA)

(16) Phoneme [a] in Figure 10 shows a peculiarity. As F2 decreases in frequency, it rather resembles a diphthong (e.g. [ao] in Figure 9) than a monophthong, like, e.g., [a?] in spa (cf. Figure 15) (BA09Ft_CN)

(17) CMC is special, for example, in that it allows shy and less-well motivated learners to interact with others …. (ALC001)

(18) Utilizing a speech analysis program such as Praat, complex sound waves can be fragmented into their component frequencies and the exact position of F1 and F2 can be specified. (BA09Mt_SS)

In examples (13) and (14), *such as* and *like* introduce particular instances of general categories (i.e. ‘those in position of power’ and ‘writers’). *For instance* and *like* in (15) and (16) present similar cases of the previous statements or ideas that are more recognizable. The third type is an association of type one and two. In (17) and (18), readers are presented with an illustration a quality or an image which clarifies the concepts. The second and third types of example are known to be less frequent than the first type (Hyland, 2007). In other words, most examples offer instances of general categories.

3. Methodology

3.1. Sample and data collection

The corpus in this study comprises 14 BA theses: eight written by students of English at University of Ismail Qemali, Vlore, Albania and six written by students at Chemnitz University of Technology, Germany. The German corpus was taken from a larger corpus called the *ChemCorpus* (cf. Schmied, this volume) whilst the Albanian corpus was a sample of the BA theses collected from the participants of the Summer School 2014.

All the theses were written and submitted between 2009 and 2013. The disparity in the number of theses for the individual sub-corpora was to avoid bias in the total number of words for each sub-corpus. As a result, both sub-corpora had roughly the same size: German Corpus (*ChemCorpus*) had a size of 80,290 words and the Albanian corpus (*AlCorpus*) had a size of 71,342 words. The total number of words was 151,632.

Corpus	Number of theses	Number of words	Average number of words per thesis
ChemCorpus	6	80,290	13,382
AlCorpus	8	71,342	8,854
Total	14	151,632	10,831

Table 2: Size of the corpus

3.2. Data analysis

The BA theses, mostly word documents, were first converted to normal text files and then analyzed with AntConc 3.2.4w (the 2011 version of the concordance tool). The texts were parsed for the selected code glosses. To ensure that the data were comparable despite the above mentioned disparities, the absolute frequency of occurrences of the exemplification and reformulation markers were normalized per 10,000 words. All occurrences were examined in context to ensure that they really had exemplification and reformulation functions. The selected code glosses were *such as, for example, e.g., an example of, like, for instance, say, that is/that is to say, i.e., in other words, put another way, known* *as, called, referred to as, which/this means, namely, in particular, particularly, specifically, especially, and parentheses( ).*

4. Results

The overall results showed variation in the use of code glosses in the corpus. As displayed in Table 3, German students used a slightly higher number of code glosses with exemplification and reformulation functions than Albanian students.

Code gloss /sub-corpora	*AlCorpus*	*ChemCorpus*	Occurrences of makers in the corpus
Exemplification markers	13.7	23.1	36.8
Reformulation markers	25.2	38.7	63.9
Total	38.9	61.8	100.7

Table 3: Code gloss markers by sub-corpora (per 10,000 words)

Figure 1 presents the total occurrences of exemplification markers per 10,000 words. The results show that German student use almost twice as many exemplification markers as Albanian students.

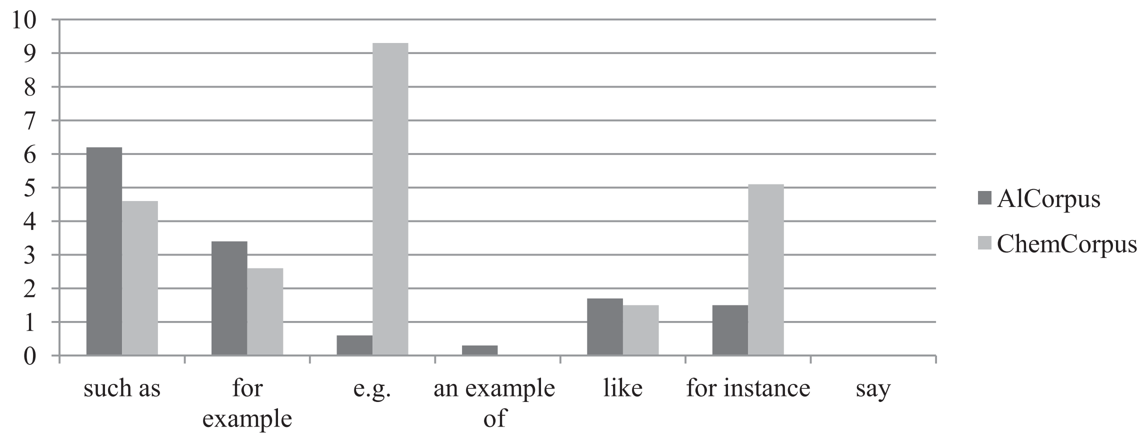


Figure 1: Total occurrences of exemplification markers in *AlCorpus* and *ChemCorpus* per 10,000 words

Both German and Albanian students used *such as*, *for example*, *e.g.*, *for instance* to signal exemplification. *Such as*, *for example*, *e.g.* and *for instance* illustrated the most significant differences between German and Albanian students. *An example of* merely occurred few times in the *AlCorpus* but not at all in the *ChemCorpus*. S*ay*, as an exemplification marker did not occur at all in the corpus. *E.g.* occurred more often in the whole corpus, most especially, in the *ChemCorpus*; followed by *such as* (specifically in the *AlCorpus*).

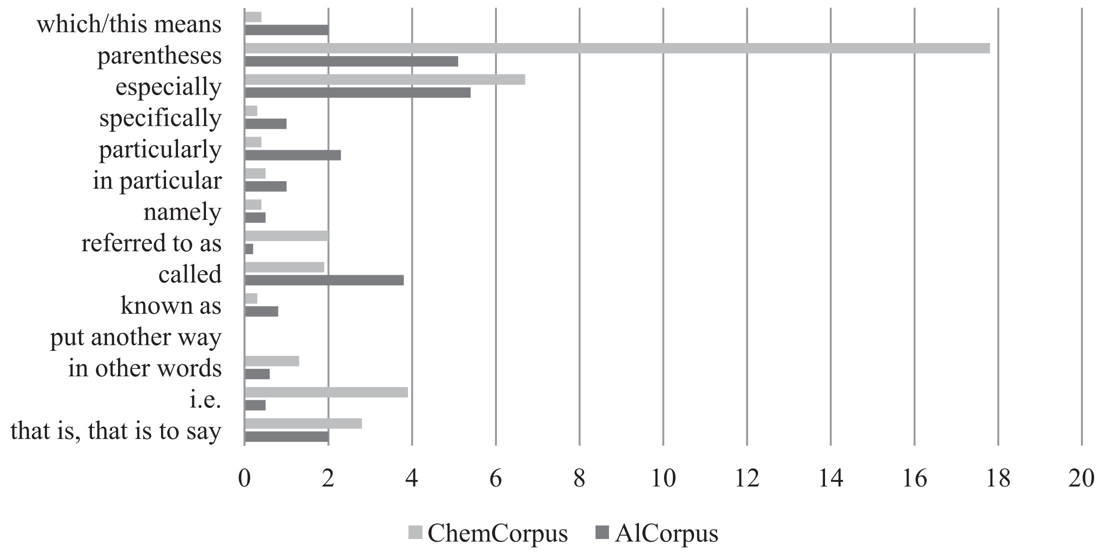


Figure 2: Total occurrences of reformulation markers in *AlCorpus* and *ChemCorpus* per 10,000 words

The total occurrence of reformulation markers was 63.9 per 10,000 words. As displayed in Figure 2, both groups of students used *that is/that is to say, i.e., in other words, called, referred to as, particularly, especially, (parentheses)* and *which/this means. Put another way* did not occur at all in the corpus. In general, German students used a higher number of reformulation markers than Albanian students (see Table 3).

5. Discussion

The results of this study have shown that novice academic students of English in Albanian and German exemplify and reformulate their claims differently. The findings confirm Hyland's claim that in the academic fields like humanities, reformulations are used "to draw implications for a potentially more diverse readership and exemplification are used "to shift between abstraction and concreteness, moving from conceptual to real world reference".

However, the hypothesis that exemplification in academic writing will be often introduced by *for example* and *for instance* (Biber et al. 1999:884) was not entirely true in the case of the student writings. Exemplification was introduced more often by *such as* and *for example* by Albanian students whilst German students used more *e.g.* and *for instance.* Thus, exemplification was signaled by different markers in the corpus.

Reformulation in the student writing genre of academic writing is not often introduced by *in other words* as suggested by Biber et al. (1999:884). The most frequent reformulation maker in the whole corpus is (*parentheses)* followed by *especially*, *i.e.*, *called, particularly, referred to as* and *which/this means. In other words* is in the less frequent category with *in particular*, *specifically*, *known as* and *namely*.

Parentheses (brackets) appear to be the most frequent code gloss used in *ChemCorpus* with an absolute frequency of 144 (17.8 per 10,000 words). The high frequency is not surprising since most of the theses sample of the *ChemCorpus* used in this study involved corpus linguistic analysis. Most of these occurrences were names of quoted texts from a selected corpus (19). Others were acronyms of already mentioned terms or words (20) and the rest, references (especially year of publication in the references or bibliography section of the theses)

(19) Mostly, that are the small events who provide the greatest joys, a word, a smile, a gesture. (ICLE-GESA5042)

(20) In Europe, a substantial part of research conducted with regard to the analysis of learner language is concerned with the International Corpus of Learner English (ICLE), which will be described in chapter 4.2. (BA09Ft_SK)

Interestingly, Albanian students seem to avoid the use of the abbreviation of *for example* (*e.g.*), where German students use it more often than the unabbreviated form (*for example*). Several internet discussions on academic writing and formal writing available on university websites (e.g. Monash University, University of South California) and networks (e.g. Western Vocational Lifelong Learning Network www.wvlln.ac.uk/uploads/academicwriting) advice students to avoid using abbreviations such *as e.g.* and *i.e*. *Perhaps,* this advice might be informing the Albanian students who avoided the abbreviations of *for example* and *that is.*

However, research into the use of code glosses by advanced users of English in academic writing (e.g. research articles, post graduate dissertations – MA and PhD theses) reveals frequent occurrences of abbreviations like *e.g. and i.e.* (Hyland & Tse 2004; Hyland 2007). A search of online corpora like Corpus of Contemporary American English (COCA), British Academic Written English (BAWE) equally presents evidence that *e.g.* and *i.e.* is used in academic writing. Furthermore, major style and publication manuals such as Modern Language Association (MLA), American Psychological Association (APA) and the Chicago Manual of Style agree that Latin abbreviations may be used in parentheses; as in examples (21), (22), and (23) below. In other words, they should not be used in the main body of a text. Thus, outside parentheses, the unabbreviated forms *(for example, that is/that is to say*) should be used (24) (25).

(21) These exposures to input, output, attention to feedback, and linguistic form are essential to SLA (e.g., Gass, 1997; Krashen, 1985; Long,1996; Pica, 1994; Swain, 1985) (ALC001)

(22) There are 2 more instance of this construction, but the preposition is missing. This is probably due to the fact that the Italian equivalents of look at are no phrasal verbs (e.g. guardare, visionare, observare, guardare, contemplare etc.) (BA09Ft_UA)

(23) Second language researchers are interested in two overlapping issues related to online communication: (1) how do the processes which occur in online communication assist language learning in a general sense (i.e., online communication for language learning); and (2) what kinds of language learning need to occur so that people can communicate effectively in the online realm (i.e., language learning for online communication). (ALC001)

(24) Additionally, the range of vowels selected for examination may have been too restricted. That is, vowels indicating the low-back merger and the third vowel involved in the

Canadian Shift should have been incorporated to obtain a more complete overview of the vowel qualities of each speaker. (BA09Mt_SS)

(25) The concept of RP is a particularly English one, having no equivalent in any other part of the English speaking world. For example, there is no US equivalent of RP. (ALC010)

From the above explanations, it appears that German students overuse or use *e.g.* and *i.e.* wrongly. Only 32.6% of the occurrences of e.g. in the *ChemCorpus* were used in parentheses. In the *AlCorpus* three out of the four occurrences (absolute frequency) of *e.g.* were used in parentheses and two out of the three occurrences (absolute frequency) of *i.e.* were in parentheses. Thus, suggesting more strongly that the decision of Albanian students to use, in general, might stem from the above mentioned manuals.

The question remains as to what could be the reason for the variation in code gloss usage in this corpus. One of the reasons may be the difference in experiences of language education. An interview of some Albanian students during the Summer School 2014 revealed that they had little or no contact with native speakers of English, unlike the German students. International exchange programmes afford German students of English the chance to garner experience in a native speaker environment such as the United Kingdom and the United States of America. As a result, German students of English are able to achieve close to native-like competences, even though English is officially a foreign language in both Albania and Germany.

Strikingly, *for example* and *for instance* translate as *zum Beispiel* in German and *përshembull* in Albanian. *Such as* and *like* translates as *so wie*/*wie* in German and *si*/*si përshembull* in Albanian. Despite the translations of these words (from L1 to L2), there seems to be no confusion in their use in English in the corpus. All four exemplification makers were found in the corpus. The students seem to use the markers interchangeably. Both Albanian and German students used *such as* more often than *like*. However, Albanian students seem to prefer *for example* whilst German students favour *for instance* (see Table 4 below).

Exemplification markers		*AlCorpus*	*ChemCorpus*
si/si përshembull	*such as*	6.2	4.6
so wie/wie	*like*	1.7	1.5
Total		7.9	6.1
përshembull	*for example*	3.4	2.6
zum Beispiel	*for instance*	1.5	5.1
Total		4.9	7.7

Table 4: Occurrences of such as/ like and for example /for instance in *AlCorpus* and *ChemCorpus* per 10,000 words

Furthermore, despite Toska's (this volume) observation of a frequent use of *përshembull* and *domethënë* (*that is*) in Albanian PhD dissertations, novice Albanian students of English in this study do not seem to follow this trend in their BA theses. For instance, *for example* and *for instance* (both translate *përshembull* in Albanian*)*, make up only 35 percent (4.9 per 10,000 words) of the exemplification markers in the *AlCorpus*. This finding tends to imply little or no L1 transfer in the use of exemplification markers. Perhaps, the large repertoire of code glosses in English makes it possible to vary markers depending on the meaning the writers intend to convey.

In considering mother tongue (L1) influences on their use of code glosses, we may also have to consider the influence of Greek and/or Italian on the linguistic strategies of Albanian students. However, in this case, the influence may only be oral and not necessarily written since all the students may have attended academic writing courses prior to the completion of their bachelor studies. With English as an international lingua franca, "academic writing is gradually establishing itself as a central element in the new BA and MA degree programmes in European universities" (Schmied 2011: 19).

More data may be needed to obtain more representative findings that confirm the outcomes of this study. A comparison of BA theses and MA theses written by students from both Albania and German and then a native speaker corpus of the same discipline and genre might enable us to ascertain how close German and Albanian students are to native patterns.

6. Conclusion

Code glosses are essential in academic writing. At the novice stage of academic writing, students also acknowledge the relevance of exemplification and reformulation.

The fact that these students were allowed to graduate seems to suggest that they were able to persuade their 'more academic expert' examiners of their claims. Even though, there might be differences (and also imperfections) in comparison to advanced writers from the same European states, the use of exemplification and reformulations signals effective writing. As Hyland & Tse (2007: 175) points out, "[w]riting effectively means anticipating the needs of readers, both to follow an exposition and to participate in a dialogue, and occasionally devices are used to perform both functions at once".

The findings of this study point to the fact that among novice non-native speakers of English in the same academic discipline, variation exists in the use of code glosses with exemplifying and reformulating functions.

Even though the decision to put something in other words may be a decision about style (Blakemore 1993: 119), code glosses, like other metadiscourse markers, should not be considered as stylistic devices that writers can vary at will. The decision to employ either exemplification or reformulation markers depends on the context. Though there exists a large repertoire of lexical devices which can signal exemplification and reformulation, academic writers must be cautious in order to select the appropriate ones that can help them negotiate meaning and facilitate the understanding of the reader.

References

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman grammar of spoken and written English*. London: Longman.

Blakemore, D. (1993). The relevance of reformulations. *Language and Literature, 2,* 101-20.

Crismore, A., Farnsworth, R. (1990). Metadiscourse in popularand professional science discourse. In W. Nash (ed.). *The writing scholar.* Newbury Park: Sage, 118-136.

Crismore, A., Markannen, R. & M. Steffensen. (1993). Metadiscourse in persuasive writing. A study of texts written by American and Finnish university students. *Written Communication, 10 (1),* 39–71.

Greenbaum, S. (1969). *Studies in English adverbial usage*. London: Longman.

Greenbaum, S. & R. Quirk (1990). *A student's grammar of the English language*. London: Longman.

Huddleston, R. & G. K. Pullum (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hůlková, I. (2011). Conjunctive adverbials in academic written discourse: Conjunctive adverbials in academic written discourse. In J. Schmied, *Academic Writing in Europe: Empirical Perspectives*, 129-142.

Hyland, K. (2007). Applying a gloss: Exemplification and reformulating in academic writing. *Applied Linguistics, 28, 2,* 266 – 285.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Killingsworth, M. & M. Gilbertson (1992). *Signs, genres and communities in technical communication.* Amityville, New York: Baywood.

Hyland, K. & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25 (2),* 156 – 177.

Leech, G. & J. Svartvik (2002). *A communicative grammar of English* (3rd ed.). London: Longman.

Markkanen, R., M.S. Steffensen & A. Crismore (1993). Quantitative contrastive study of metadiscourse.problems in design and analysis of data. *Papers and Studies in Contrastive Linguistics 28,* 137-157.

Nkemleke, D. (2011). *Exploring academic writing in Cameroon English: A corpus-based perspective*. Göttingen: Cuvillier,

Oxford Dictionary. (2014, November 11) Retrieved from http://www.oxforddictionaries.com

Povolná, R. (2013). On some variation in the use of discourse markers by Czech and German students. In *Discourse and Interaction, 6.2.,* Brno: Masaryk University. 41-60.

Schmied, J. (2015). English for academic purposes: Research-based approach to metalanguage and genre. This volume, 1-24.

Schmied, J. (2011). Academic writing in Europe: A survey of approaches and problems. In J. Schmied & C. Haase (eds.). *Academic writing in Europe: Empirical perspectives*. Göttingen: Cuvillier, 1–22.

Swales, J.M. & C.B. Feak (2012). *Academic writing for graduate students: Essential tasks and skills*. Michigan: The University of Michigan Press

Toska, B. (2015). Structural, textual and commentary linkers in Albanian PhD theses: A pilot study. This volume, 163-176.

Wagner, S. (2011). Concessives and contrastives in student writing: L1, L2 and genre differences. In J. Schmied (ed.) *Academic writing in Europe: Empirical perspectives*. REAL Studies 5. Göttingen: Cuvillier Verlag. pp. 23–49.

Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985). *A comprehensive grammar of the English language*. London: Longman.

Introductions by Serbian Native Speakers in Academic Writing

Jasmina Djordjevic (Union University Belgrade)

Abstract

The aim of this article is to investigate the influence of acquired standards of writing, dominant in the Serbian language, on academic writing in the English language. By applying rubrics, an assessment tool based on predefined criteria and expected levels of quality, a qualitative and quantitative analysis of a corpus of 50 different essays written by third-year English language students has been performed. The analysis was expected to establish that Serbian native speakers studying English at tertiary level tend to write incoherent essays in English presumably because of the negative transfer of the writing style developed in Serbian language classes at elementary and secondary education level. The inheritance that the students bring to their English writing is very dominant and unless the students are provided extensive instruction and guidance, they are unable to reduce the influence of the acquired writing style. The negative transfer is particularly obvious in the introduction and the general structure of the final written products submitted by the students.

1. Introduction: The issue of influence in language learning and acquisition

A common point of view found in the literature on language learning and acquisition is that when a person is in the process of learning or acquiring another language, an inevitable by-product will be that the native language and the new language being learned or acquired will influence each other in some way (Ellis 1994; Kellerman 1995; Larsen-Freeman and Long 1991; Odlin 1989; Selinker 1992). In some cases, the influence might be positive; in others, it might be negative. In either case, “those instances of deviation from the norms of either language which occur in speech of bilinguals as a result of their familiarity with more than one language” (Weinrich 1953: 1) support the belief that a person’s first language has an unavoidable effect on their second language and vice versa (Auer, Hinskens & Kerswill 2005; Heine & Kuteva 2005; Johanson 2002; Pavlenko & Jarvis 2001).

What most of the above-mentioned authors have been trying to point out is that learners and/ or speakers of more than one language are prone to providing a perfect soil for language contact. Consequently, this contact will induce some change. A closer look at the history of studies on contact-induced language change so far makes clear that to date researchers have proposed different points of view regarding language-induced change, referring to it as ‘interference’ (Weinreich 1953), ‘cross-linguistic influence’ (Pavlenko & Jarvis 2001), ‘code-copying’ (Johanson 2002), ‘transfer’ (Heine & Kuteva 2005) and ‘convergence’ (Auer, Hinskens & Kerswill 2005).

The difficulties that students of English as a foreign/ second language encounter may be identified as any of the above-mentioned. In line with the aim and scope of the research presented in this article, problems and difficulties related to the skill of writing might not only be the result of purely linguistic factors, but also of the conventions of writing acquired in the process of education in the native language. A general term that has been accepted regarding this area is 'discursive transfer', which shall be discussed in more detail in section 2. At this point, suffice it to say that the empirical research to be presented in this article is an attempt at identifying the most common causes of transfer recognised in the introductions and the general structure of final written products submitted by Serbian native speakers in academic writing in English.

2. The problem of discursive transfer

Pavlenko & Jarvis (2008: 102) refer to discursive transfer as "the ways thoughts are introduced, organized and contextualized within an oral or written discourse, and [it] also relates the conversational strategies that are used to maintain a conversation, as well as the concepts and notions that are conventionally expressed in a particular discursive context". The problem of discursive transfer has mainly been discussed in relationship to contrastive rhetoric where it has been established that "each language and culture has unique rhetorical conventions and that they negatively interfere with L2 writing" (Kubota 1998: 69). Based on this point, it can be concluded that rhetorical patterns reflect not just language ability but also writing conventions. The result is transfer related to rhetoric, which inevitably blurs the distinction between language transfer and writing system transfer (Cook & Bassetti 2005: 29).

In the section to follow, more details shall be provided on the problem of discursive transfer in writing and subsequently on academic writing.

2.1. Discursive transfer in writing

In-depth research into the problem of discursive transfer shows, among other things, that discursive transfer occurs during both production and reception (Connor 1984), that it occurs in both the forward and the reverse direction, i.e. from L1 to L2 and vice versa (Kubota & Lehner 2004) and that the transfer of writing skills is often positive in both directions (Kubota 1998).

Obviously, the accumulated evidence for discursive transfer is rather compelling so that a general conclusion may be drawn that an L2 user's discourse style and organization of information depend on various factors, such as age, L2 proficiency and experience in writing. The most important outcome of research in the area of discursive transfer is that the study of second language writing (Cook & Bassetti 2005) has been approached from the perspective of contrastive rhetoric (Connor 1996). As such, it has brought about considerable insight into the field of writing and literacy to the extent that "several scholars have begun to argue that biliteracy is a phenomenon distinct from bilingualism, and thus transfer effects pertinent to biliteracy should be studied on their own terms" (Pavlenko & Jarvis 2008: 15).

Therefore, the research presented in this article investigating the transfer of discourse style in the case of Serbian native speakers acquiring the skill of academic writing in English might provide valuable insight into the problem of poorly structured writing and a rather ineffective organization of information especially in the introduction of their essays.

2.2. Differences between the styles of Serbian and English writing

An accepted classification in the study of functional writing styles in Serbian has been provided by Tošović (2002) identifying five different styles: conversational, publishing, administrative, scientific and literary-artistic. The syllabi for Serbian language courses at the elementary and secondary level prescribed by the Serbian education system suggest the furthering of the literary-artistic style. The reasons for such tradition exceed the scope of this article but it is obvious that these preferences differ from the writing style preferred and furthered in English.

A closer comparison between these two styles yields the following main differences:

English	Serbian
The purpose of writing is to present a point of view clearly stated at the beginning of the writing in the form of a thesis statement.	The purpose of writing is to present the author's perception of reality, transform that perception into an image and to influence the reader with that created image.
The aim is to elaborate on the point of view by providing sufficient supporting arguments.	The aim of writing is first to influence the recipient and then to present the personal point of view.
The line of thoughts follows a logical structure of arguments.	The line of thoughts yields an aesthetic/ artistic transformation of reality.
The result is a coherent presentation of information and logical organization of firmly grounded ideas within paragraphs.	The result is a subjective transformation providing an individualized presentation of reality.
The objective is to provide a strong and effective conclusion based on the entire writing convincing the reader of the points presented.	The objective is not to provide conclusions but a discrete system of images.

Table 1: Main differences between writing in English and Serbian

Obviously, all the characteristics of the preferred form of Serbian writing outlined in Table 1 (Tošović 2002: 168-175) can be summarized as reflecting "either implicitly or explicitly, information about the author" and their personal points of view and attitudes (p. 175). In comparison to that, academic writing in English is expected to be objective while claims are to be supported by solid arguments and facts.

The research presented in this article is aimed at investigating the probable causes of a rather low level of performance in academic writing in English among Serbian native speakers studying the English language at tertiary level. A possible outcome of the relevant analysis of a corpus of 50 different essays written by third-year English language students should help answer the question why Serbian native speakers tend to write incoherent essays in English. Therefore, the research questions that the analysis presented here will be based on are as follows:

a) Do Serbian native speakers studying the English language transfer the conventions of the Serbian writing style to their academic writing in English?
b) How obvious is the negative transfer in the introduction and the general structure of the final written products submitted by the students?

The analysis of the corpus compiled for the purpose of this research should confirm the hypothesis that the performance of Serbian native speakers in academic writing in English is generally low because of the negative transfer of the writing style developed in Serbian language classes at elementary and secondary education level.

3. The research

As indicated above, the empirical research conducted for the purpose of this article was aimed at validating the hypothesis whether Serbian native speakers tend to transfer the dominant writing style acquired in Serbian language classes to their academic writing in English. The result of this transfer is a rather low performance in their written production in English, primarily evident in the structure and the organization of ideas presented in the final pieces they produce.

3.1 Corpus

The corpus compiled and subsequently analysed for the purpose of this research consists of 50 essays (E), each essay within the scope of the prescribed length of 350-500 words written by third-year English language students at the Faculty of Law and Business Studies Dr Lazar Vrkatic in Novi Sad during the summer semester of the school year 2013/ 2014 as part of their Legal English course.

The essay is an obligatory written product all students have to submit in order to qualify for the final oral exam. The overall assessment at the end of the semester is calculated on the basis of pre-examination activities comprising two tests (15% + 20%), attendance and activities (5%), essay writing (20%) and the final oral exam (40%). The students are free to decide on their essay topic but it has to be related to the syllabus covered in the textbook (Krois-Lindner 2011) used for the course during the semester.

The syllabus includes five topics from the area of law: Intellectual Property, Negotiable Instruments, Secured Transactions, Debtor-Creditor Relationships and Competition Law. As many as 36 students (72%) whose essays have been analysed for the purpose of this research chose a topic from the area of Intellectual Property as this area covers aspects, such as copyright infringement, internet piracy, trademarks, trade secrets, etc. Students seem to identify with this area more easily than with the remaining four topics, which must have been crucial for their choice of topic. Related to the theme of Debtor-Creditor Relationships, 8 students (16%) chose to write about insolvency practitioners, only 4 students (8%) wrote about cartels, i.e. Competition Law and 2 students (4%) chose to write about Secured Transactions. None of the students wrote about Negotiable Instruments.

In section 3.3. below, a detailed discussion of all the results obtained in this research shall be presented. At this point, it should be mentioned that all essays were coded by adding ordinal numbers to the code E yielding thus a corpus ranging from E01-E50. The shortest essay consists of 352 words and the longest one of 496 words.

3.2. Methodology

In line with the general conventions applied in discourse analysis (Herring 2004, 2010, Tannen and Trester 2013), several aspects had to be analysed in the compiled corpus. In addition, the criteria clearly indicating the negative transfer of the writing style developed in Serbian language classes at elementary and secondary education level had to be determined. A preliminary analysis of 20 essays indicated the protruding lack of a straightforward topic sentence introducing the reader into the problem discussed in the essay submitted by the student. Furthermore, it could be noticed that the structure of most of the analysed essays followed the pattern of an individualized presentation of ideas instead of an objective outline of arguments. Based on the preliminary findings, it was deemed insufficient to apply only one technique of analysis on the entire corpus of 50 essays so that it was decided to apply a combination of quantitative and qualitative analysis based on rubrics and content analysis.

Rubrics is an assessment tool usually applied for the evaluation and assessment of students' written language production. The assessment is based on a predefined set of clearly and precisely articulated expectations for writing assignments by listing criteria and describing levels of quality (Stephens & Randall 2011). For the purpose of this research, two criteria (C) were defined: C1 – a clear thesis statement provided in the introductory paragraph of the essay and C2 – a coherent organization of arguments reflected in the structure of the final essays. With respect to the levels of quality, all essays were rated on a scale of 1 to 5 (below standard, rather insufficient, approaching standards, meets standards and above standards). The descriptors for each of the defined levels of quality are presented in Table 3 and Table 4 respectively.

As stated above, a content analysis had to be applied on the final corpus along with the analysis based on rubrics so that the method of corpus-based content analysis (Kutter & Kanter 2012) was applied. This form of analysis is a "standardised hermeneutic procedure of text interpretation in the course of which the individual analyst assigns abstract categories to propositional contents" (p. 7). An important aspect of this approach is that the “categories do not correspond to the ‘observable’ (linguistic) characteristics of the text, but to hypotheses derived from social theory... [whereby]... the necessary validity and reliability of assignment (coding) is assured by a classification system” (p. 7).

The final analysis performed for the purpose of this research was based on the combination of the predefined criteria as well as the content identified in the thesis statement and the essay structure of each analysed essay. In other words, the conent of each essay from among the 50 essays was analysed with respect to a clear thesis statement provided in the introductory paragraph of the essay as a characteristic of C1 and a coherent organization of arguments reflected in the structure of the final essays as a characteristic of C2 (see section 3.3. for a detailed quantitative and qualitative analysis of the data obtained in this research).

3.3. Results and discussion

As has been indicated, a quantitative and qualitative analysis of the two established criteria has been performed on the compiled corpus. In the sections to follow, each criterion will be presented separately with respect to the determined characteristics. What needs to be pointed out in advance is that the overall results of the analysis performed yielded rather disappointing results (Table 2):

Criterion	5 – above standards		4 – meets standards		3 – approaching standards		2 – rather insufficient		1 – below standard	
	No.	%	No.	%	No.	%	No.	%	No.	%
C1	0	0	3	6.0	5	10.0	4	8.0	38	76.0
C2	0	0	2	4.0	3	6.0	4	8.0	41	82.0

Table 2: Overview of scores allocated to the analysed essays for both criteria

Another point to be made here is that the titles of the individual essays have not been analysed, either separately or in connection to the essays, because many students used titles suggested by the teacher, i.e. 14 students resorted to one of three different suggested titles:

- Stricter regulations should be introduced into the Employment Law in Serbia regarding nepotism in the work place;
- Women should be able to resume their careers after taking a maternity leave and
- The Function of the Insolvency Practitioner May Be Deemed both as Saviour and as Destroyer.

Another 10 students used very general titles for their essays, such as *Intellectual Property, Terms and Conditions of a Sale, Intellectual Property Rights, Copyright Protection and Competition Law*, etc. Since in both cases the titles could not yield data relevant to the topic of the research presented here, they have not been taken into consideration as an important variable.

3.3.1. *Clear thesis statement*

With respect to the first criterion, the content of the introductory paragraphs in all essays was examined on a scale of 1-5, whereby the score 1 was allocated to a thesis statement (TS) that was completely irrelevant to the topic and the score 5 to a thesis statement which names the topic of the essay and outlines the main points to be discussed (Table 3).

Criterion	5 – above standards	4 – meets standards	3 – approaching standards	2 – rather insufficient	1 – below standard
The thesis statement clearly introduces the topic	TS names the topic of the essay and outlines the main points to be discussed.	TS names the topic of the essay and poorly relates to the main body of the essay.	TS unclearly states the topic and lacks a reference to the main body.	TS vaguely states the topic and lacks a reference to the main body.	TS is irrelevant to the topic.

Table 3: The scores and quality descriptors allocated to the levels of performance for the first criterion

As indicated in Table 2, as many as 38 essays (76.0%) scored below standard with respect to the first criterion. In other words, most of the students failed to provide a thesis statement that would be relevant to the topic. A result like that indicates that Serbian students of English resort to lengthy, generalized and rather imprecise introductions, a characteristic common in writing in Serbian. The following excerpts should serve as an illustration of this problem identified in most of the analysed essays:

[E03] Title: Ethical issues and the impact of piracy on the author's Intellectual Property Rights
TS: Throughout the years, internet piracy or "file sharing" has constantly received global attention. Each year, the movie production businesses as well as application software and video game firms are faced with losing a great amount of money. This loss is a result of internet piracy. Many people would admit that file sharing is in a way unethical and consider it no different from a real theft. On the contrary, there are those that support file sharing and assert that their actions are righteous and ethical.

[E14] Title: Negative effects of counterfeiting
TS: The verb to "counterfeit" actually means to imitate something. In practice, counterfeit products refer to fake products or copies of the original products. Counterfeit products can include electronics, clothes, art, pharmaceuticals, documents, etc. Those products are usually of a bad quality. There are many kinds of counterfeiting.

[E32] Title: The introduction of insolvency practitioners in Serbia would save many companies from bankruptcy
TS: The profession of an insolvency practitioner includes, among other things, the attempt to save a business when it is in danger. That is someone who will do everything that is in his power to save a company so it can continue doing business. That is a huge problem we have in Serbia - the fact that there are many companies that are going bust on a monthly basis.

In comparison to that, the following are illustrations of introductions that have been evaluated as meeting the standard defined for this type of assignment.

[E05] Title: The most prominent impacts of cybersquatting on small businesses
Structure: Cybersquatting has become a major problem of the modern age. If you happen to run your own small business, you can encounter serious problems related to cybersquatting, such as the theft and use of your domain name for profit or personal gain thus jeopardizing your reputation. (Score 3)

[E27] Title: Individual actions as a way of protection against copyright infringement
Structure: Individual actions might serve to preserve intellectual property or creations of mind against copyright infringement. Two possible actions could be the use of password protection and defining threats in the form of a possible fine. (Score 4)

3.3.2. *Essay structure*

In regard to the second criterion, the content of the main body of all essays was examined on a scale of 1-5, whereby the score 1 was allocated to a an essay which neither includes the aspects mentioned in the thesis statement nor refers to them and the score 5 was allocated to an essay which includes all the aspects mentioned in the thesis statement and refers to them clearly (Table 4).

Criterion	5 – above standards	4 – meets standards	3 – approaching standards	2 – rather insufficient	1 – below standard
The <u>structure of the essay</u> refers to the relevant aspects mentioned in the thesis statement	Includes all the aspects mentioned in the TS and refers to them clearly	Includes all aspects mentioned in the TS but does not refer to all of them clearly.	Includes few aspects mentioned in the TS and refers to them vaguely.	Includes only one or some of the aspects mentioned in the TS and does not refer to them at all.	Neither includes the aspects mentioned in the TS nor refers to them.

Table 4: The scores and quality descriptors allocated to the levels of performance for the second criterion

As indicated in Table 2, as many as 41 essays (82.0%) scored below standard with respect to the second criterion. In other words, almost all the students failed to produce an essay, which would include all the aspects mentioned in the TS and refer to them clearly. A result like that indicates that Serbian students of English resort to providing an individualized presentation of what they believe is relevant instead of providing solid arguments supporting the topic they are supposed to pursue. The topic sentences presented in section 3.3.1. were developed as follows:

[E03] Title: Ethical issues and the impact of piracy on the author's Intellectual Property Rights

Structure: The student ignored the 'impact of piracy' and 'intellectual property' altogether. The entire essay is an elaboration on file sharing, a topic not announced either in the title or in the TS.

[E14] Title: Negative effects of counterfeiting

Structure: The student devoted the first half of the essay to different types of counterfeiting. The second paragraph of the main body was about the reasons for counterfeiting. Only in one sentence in the conclusion did the student refer to the negative effects by stating that they are harmful to a business.

[E32] Title: The introduction of insolvency practitioners in Serbia would save many companies from bankruptcy

Structure: The essay is rather disorganized. The student tried to start the essay by providing examples of companies in Serbia indicating they could have been saved if an insolvency practitioner had been available. However, instead of providing objective arguments supporting the idea of introducing insolvency practitioners in Serbia, the student drifted into feeling sorry for all those employers and employees who work at companies facing bankruptcy because of inadequate management.

The introductions presented in 3.3.1. considered to meet the standard defined for this type of assignment were developed as follows:

[E05] Title: The most prominent impacts of cybersquatting on small businesses
Structure: Though the student does refer to the arguments introduced in the TS, an obvious failure is evident in the presentation of the arguments. The essay lacks coherence and an overall straightforward development of the main topic. The score allocated to this structure is 2.

[E27] Title: Individual actions as a way of protection against copyright infringement
Structure: The student pursued the two announced actions in a straightforward manner. The first action was described in the first paragraph and the second action in the second paragraph of the main body of the essay. However, the arguments were presented in a fairly disorganized way ao that the score allocated to this structure is 3.

4. Conclusion

The research presented in this article was aimed at investigating the probable causes of a rather low level of performance in academic writing among Serbian native speakers studying the English language at tertiary level. The analysis performed on a corpus of 50 different essays written by third-year English language students was expected to answer the question why Serbian native speakers are unable to write coherent essays in English.

The analysis of the corpus compiled for the purpose of this research confirmed the hypothesis that the performance of Serbian native speakers in academic writing in English is generally low because of the negative transfer of the writing style developed in Serbian language classes at elementary and secondary education level. The research questions posted at the beginning of the research can be answered as follows:

a) Serbian native speakers studying English transfer the conventions of the Serbian writing style to their academic writing in English. In other words, Serbian students of English resort to lengthy, generalized and rather imprecise introductions. Apart from that, Serbian students resort to providing an individualized presentation of what they believe is relevant instead of providing solid arguments supporting the topic they are supposed to pursue
b) The negative transfer of Serbian writing conventions is obvious in both the introduction and the general structure of the final written products submitted by the students. This could be confirmed by the fact that 38 essays (76.0%) scored below standard as most of the students failed to provide a thesis statement that would be relevant to the topic. Furthermore, 41 essays (82.0%) scored below standard as the students failed to produce an essay, which would include all the aspects mentioned in the TS and refer to them clearly.

It is obvious that a more elaborate research should be conducted including a larger corpus, probably across several years of study and following the development of the students as they progress from one level to another. Yet, the research presented here indicates that the negative transfer of Serbian writing conventions is present in the students' written products submitted as part of their academic writing training. Instead of a conclusion, it should be pointed out that the most valuable inference from this research can be referred to the pedagogical implications for an academic writing course – that of helping Serbian native speakers studying English at tertiary level get rid of everything they learnt about writing in their Serbian language classes at elementary and secondary school.

References

Auer, P., Hinskens, P. & P. Kerswill (2005). *Dialect change: Convergence and divergence in European languages*. Cambridge: Cambridge University Press.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Connor, U. (1984). Recall of Text: Differences between first and second language readers. *TESOL Quarterly, 18,* 239-256.

Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second language writing*. Cambridge University Press.

Cook, V. & B. Bassetti (eds. 2005). *Second language writing systems.* Clevedon, Buffalo, Toronto: Multilingual Matters Ltd.

Heine, B. & T. Kuteva (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Herring, S. C. (2010). Computer-mediated conversation: Introduction and overview. *Language@Internet* 7, article 2. Retrieved from http://www.languageatinternet.org/articles/2010/2801

Johanson, L. (2002*). Structural factors in Turkic language contacts*. Richmond, Surrey: Cruzon Press.

Kellerman, E. (1995). Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics, 15,* 125-150.

Krois-Lindner, A. (2011). *International legal English.* Cambridge: Cambridge University Press.

Kubota, R. (1998). An investigation of L1-L2 transfer in writing among Japanese university students: Implications for contrastive rhetoric. *Journal of Second Language Writing* 7*,* 69-100.

Kubota, R., & A. Lehner (2004). Toward critical contrastive rhetoric. *Journal of Second Language Writing, 13,* 7-27.

Kutter, A. & C. Kanter (2012). Corpus-based content analysis: A method for investigating news coverage on war and intervention. International Relations Online Working Paper, Stuttgart: Stuttgart University. Retrieved from http://www.uni-stuttgart.de/soz/ib/forschung/IRWorkingPapers/IROWP_Series_2012_1_Kutter_Kantner_Corpus-Based_Content_Analysis.pdf [15 July 2014]

Larsen-Freeman, D. & M. Long (1991). *An introduction to second language research*. London: Longman.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge, UK: Cambridge University Press.

Pavlenko, A. & Jarvis, S. (2001) Conceptual transfer: New perspectives on the study of cross-linguistic influence. In E. Nemeth (ed.), *Cognition in language use.* International Pragmatics Conference, Vol. 1. Antwerp: International Pragmatics Association, 288-301.

Pavlenko, A. & S. Jarvis (2008). *Crosslinguistic influence in language and cognition.* New York: Routledge.

Schmied, J. (2011). Academic writing in Europe: A survey of approaches and problems. *Research in English Language and Linguistics, 5,* 1-23.

Selinker, L. (1992). *Rediscovering Interlanguage*. London: Longman.
Stephens, D. L. & B. L. Randall (2011). *Culturally proficient collaboration: Use and misuse of school counselors*. Thousand Oaks, CA: Corwin.
Tannen, D. & A. M. Trester (2013). *Discourse 2.0. language and new media.* Washington, DC: Georgetown University Press.
Tošović, B. (2002). *Funkcionalni stilovi* [Functional styles]. Graz: Institut für Slawistik der Karl-Franzens-Universität Graz.
Weinreich, U. (1953). *Languages in contact*. London and Paris: Mouton.

Expressing Commitment and Detachment in Non-Native English BA Graduate Papers

Zorica Trajkova (Ss Cyril and Methodius University Skopje)

Abstract

This paper seeks to explore how Macedonian and German university students express commitment and detachment in their BA graduate papers through the use and pragmatic function of two interpersonal metadiscourse markers: *hedges* and *boosters*. The corpus consists of eleven BA papers written by German students of English linguistics and eleven BA papers written by Macedonian students of English linguistics. The findings suggest that these two interpersonal metadiscourse markers are used by both German and Macedonian students, but that there are variations as to their distribution and pragmatic function. This is most probably a result of the perception of their persuasive power by the writer based mostly on the two cultures' different value systems concerning this particular genre, i.e. the format of this written text in the two societies.

1. Introduction

This article explores the form and pragmatic function of hedges and boosters in BA papers. The aim is to see in which manner students modify their arguments to commit fully or partially to their truth value and appear more persuasive to the readership.

2. Academic discourse

2.1. BA thesis as a written genre

BA graduate papers are a type of argumentative academic writing. It is the final piece of writing that students submit before they graduate. It usually focuses on a small scale research in the student's area of study and sometimes it needs to be defended before a committee of (usually three) professors.

There has not been extensive research on the use of metadiscourse markers in this particular genre (see for instance Pungă & Borchin, 2014) compared to the research carried out on other types of academic writing (Nivales 2011; Crismore, Markkanen & Steffensen 1993; Hyland 2004 etc.).

The appropriate use of metadiscourse markers, especially hedges and boosters, in BA papers is very important because they help the writers establish a suitable relationship with the readers as well as express their thoughts, beliefs and research results appropriately. The BA paper offers them the possibility to deepen their research by using all the knowledge of linguistics they have gained throughout their studies. This kind of writing is the first more serious type of academic text they are required to work on. Prior to this, they mostly have had experience of writing argumentative essays, seminar papers or short research articles. However, since students usually do smaller scale research in BA papers than in master theses or doctoral dissertations, it is assumed that their knowledge in the specific language area is somewhat limited. Therefore, it is essential that they are careful when making their statements, i.e. when modifying their assertions. In order to be more persuasive for the readership, it is highly recommended that they tone down potentially “risky” claims and intensify those which they are certain about or have evidence for.

2.2. What is metadiscourse?

Metadiscourse, a term coined by Zellig Harris in 1959, is used to offer a way of understanding language in use, representing a writer’s attempt to guide the readers’ perception of a text (Hyland 2005: 3). It is also known as metatext or metalanguage, and is generally defined as ‘text about text’ or ‘discourse about discourse’, i.e. it is the text the writer uses to talk about their own text. However, as Hyland (2005: 39) states: *“metadiscourse does not simply support propositional content. It is the means by which propositional content is made coherent, intelligible and persuasive to a particular audience”.* Therefore, it is very important that one distinguishes well between the metadiscourse and non-metadiscourse (propositional) function of words or phrases in a text (cf. Schmied 2015: 3ff.).

There are a number of different ways in which metadiscourse markers have been categorized (Halliday 1974, Williams 1981, Kopple 1985, Crismore 1989, Crismore et.al 1993, Hyland 1998, 2000, 2005). However, generally there are two types identified: *textual1* and *interpersonal metadiscourse markers*. Hedges and boosters are considered to be interpersonal metadiscourse markers because they help the writer establish a relationship with the readers and present their stance towards the text.

2.3. Interpersonal discourse: defining hedges and boosters

As mentioned before, the use of interpersonal metadiscourse markers shows that the writer is trying to establish a certain relationship with the readers and is not simply stating unmediated facts or rudely imposing their beliefs on them. Linguistic literature on hedges and boosters often regards the two concepts as closely related because they help writers to mitigate the claims they are not fully confident about and strengthen the ones they strongly believe are true. The appropriate use of hedges and boosters in BA theses helps the writer to adequately express their thoughts and views on a certain topic without being too imposing or sometimes vague. Therefore, they help the writers to easily persuade the readers in the truth value of their arguments.

---

[1] They are used to organize the discourse better and are in charge of the cohesion of the text.

2.3.1. *Hedges*

Hedges[2] are interpersonal metadiscourse markers which are employed by writers to mitigate the strength of their propositions and present them as opinions rather than facts thus avoiding their possible opposition by the audience. Through the use of hedges the writer gives directions to the readers on how they should read and interpret the text.

Hedging has been widely researched by many linguists (Crismore & Farnsworth 1990, Crismore et al.1993, Hyland 1998, Vande Kopple 1985 etc.). Crismore & Vande Kopple (1988: 184-185) draw a link between hedging and metadiscourse by saying that hedges function interpersonally and tend to mark modality, which is why they are often considered part of metadiscourse. Hedging is associated with epistemic modality since they 'both express degree of writer's confidence or belief about the likelihood of a proposition' (Trajkova, 2011: 72). Brown and Levinson (1987) treated hedging as a sign of politeness, both negative[3] and positive[4].

2.3.2. *Boosters*

Boosters[5] are metadiscourse markers employed by writers to emphasize certainty in the truth value of their propositions and establish a good rapport with the readers. Their usage in graduate papers is very important because they help the students to make strong, confident claims when they are pretty sure in their truth value. Hence, it is expected that they would appear more confident and knowledgeable to the readers.

2.3.3. *Classification of hedges and boosters*

Several categories of words (phrases and sentences, too) have been identified so far to have the function of hedges and boosters. For the purposes of this research, we focus only on word forms and follow Hyland's (1998, 2005) classification:

---

[2] The linguistic term hedge/hedging was introduced by Lakoff (1972) to describe "words whose job is to make things more or less fuzzy" and it is generally defined as a technique of using tentative language by writers/ speakers in order to express certain degree of doubt and uncertainty in the truth of their propositions. Hedges appear with many different terms in the linguistic literature: *compromisers, downtowners, weakeners, downgraders, softeners*

[3] Negative politeness (1987: 129) refers to 'addressee's want to have his freedom of action unhindered and his attention unimpeded. It performs the function of minimizing the particular imposition that the face-threatening act unavoidably affects'.

[4] According to Brown and Levinson (1987: 101), some hedges, such as: *sort of, kind of, like, in a way* can have positive-politeness function. Positive politeness refers to 'redress directed to the addressee's positive face, his perennial desire that his wants (or the actions/acquisitions/values resulting from them) should be thought as desirable'.

[5] Boosters can be found in the literature under different terms: *emphatics, intensifiers* or *certainty markers*

	HEDGES	BOOSTERS
Epistemic modal verbs[6]	*may, might, could, would, should*	*will, must*
Epistemic lexical verbs[7]	*suggest, assume, believe, think, guess, estimate, feel* (judgmental verbs); *appear, seem, indicate, imply* (evidential verbs)	*demonstrate, know, prove*
Epistemic adverbs	*perhaps, possibly, probably, likely*	*certain(ly), obvious(ly), undoubtedly, clear(ly)*
Epistemic adjectives	*possible, probable, un/likely*	*certain, obvious, clear*
Epistemic nouns	*assumption, possibility, suggestion*	*the fact, claim*

Table 1: Classification of hedges and boosters

3. Corpus and research methodology

The corpus consisted of 11 BA graduate papers written by German students majoring in English at the Technical University Chemnitz, Germany and 11 BA graduate papers written by Macedonian students of English from Ss. Cyril and Methodius University in Skopje, Macedonia. All the papers were written in the area of linguistics. It is important to note down that the German papers were about three times longer than the Macedonian ones (14,223 versus 5,242 word tokens), because the English department at the Macedonian university had brought an internal rule to restrict the number of pages to 25-40.

The analysis was carried out both manually and with the help of computer software. Firstly, the programme *concordance* AntConc 3.2.3w (as well as the *'Wordsmith tools':* software used to list most frequent word uses in a selected corpus, *Key Words* by Mike Scott 1998) was used to obtain a list of all the uses of certain words/ phrases in the corpus. Then, they were analysed manually to distinguish those which have metadiscourse functions of hedges and boosters in the text. After that, the selected ones were analysed semantically. Finally, they were analysed formally, i.e. they were classified in an appropriate word group.
Two initial hypotheses were made prior to the analysis. The expectations were that:

1. there will be differences in the use of the various subtypes of hedges and boosters across the two corpora (GerBA and MacBA) based mostly on cultural differences and the perception of this genre by the two societies;

[6] We followed Coates's (1983: 23) framework of modality in deciding which epistemic modal verbs have the pragmatic function of hedges in the spoken texts.
[7] These verbs have modal meaning because they express the level of speaker's confidence in the truth of the proposition. The same goes for the epistemic adjectives, adverbs and nouns.

2. German students will use hedges and boosters more frequently than Macedonian students because these students have been explicitly taught about metadiscourse markers throughout their studies.

4. Distribution and pragmatic function of hedges and boosters

The analysis showed that there was a difference in the distribution of the hedges and boosters in both corpora. Figure 1 features the distribution of hedges in the corpus. It can be seen that German BA graduate papers contained more hedges in general. The most frequently used ones by both groups of writers (German and Macedonian) were the modal verbs. German writers used more verbs than adverbs, while Macedonian writers used adverbs more than they used verbs. The epistemic adjectives and nouns were used least frequently in both corpora.

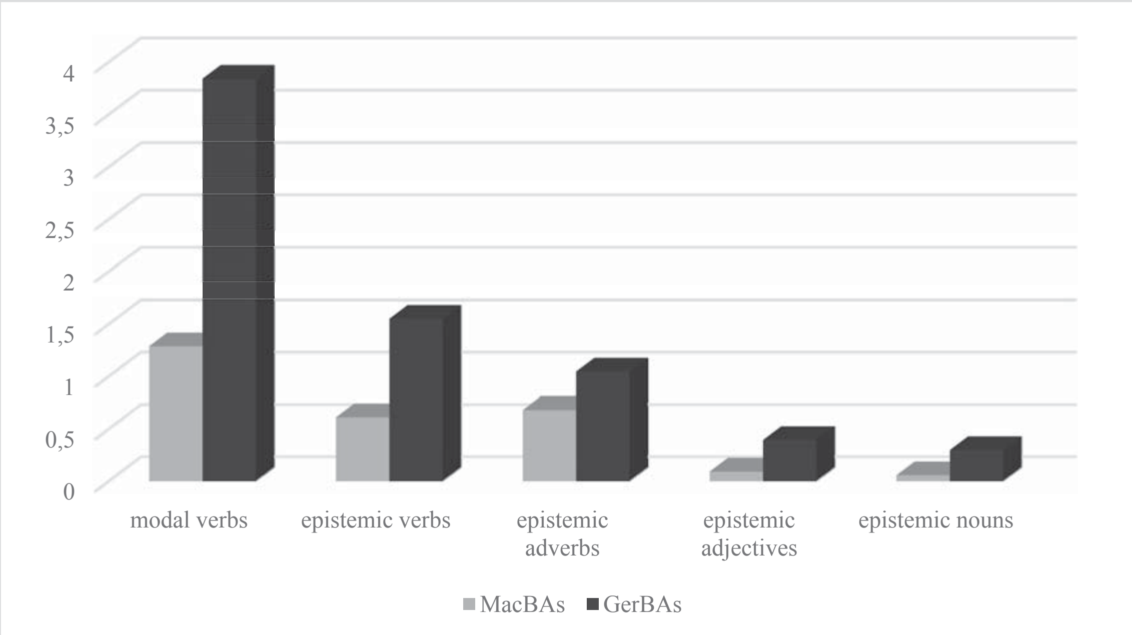


Figure 1: Overall distribution of hedges in the corpus (per 1,000 words)

As for the analysis of the use of boosters, Figure 2 shows that there was a difference in the preferred form of boosters between the German and the Macedonian students. It can be seen that the German students opted mostly for epistemic adverbs and verbs to modify their statements and intensify their persuasive effect. Although not that frequently, they also used epistemic nouns and adjectives. Modal verbs were the least preferred form of boosting. Macedonian writers, on the other hand, used mostly epistemic verbs to intensify the illocutionary force of statements. The adverbs were the second preferred choice. The epistemic nouns were the third, the modal verbs their fourth choice. The epistemic adjectives were the least preferred form.

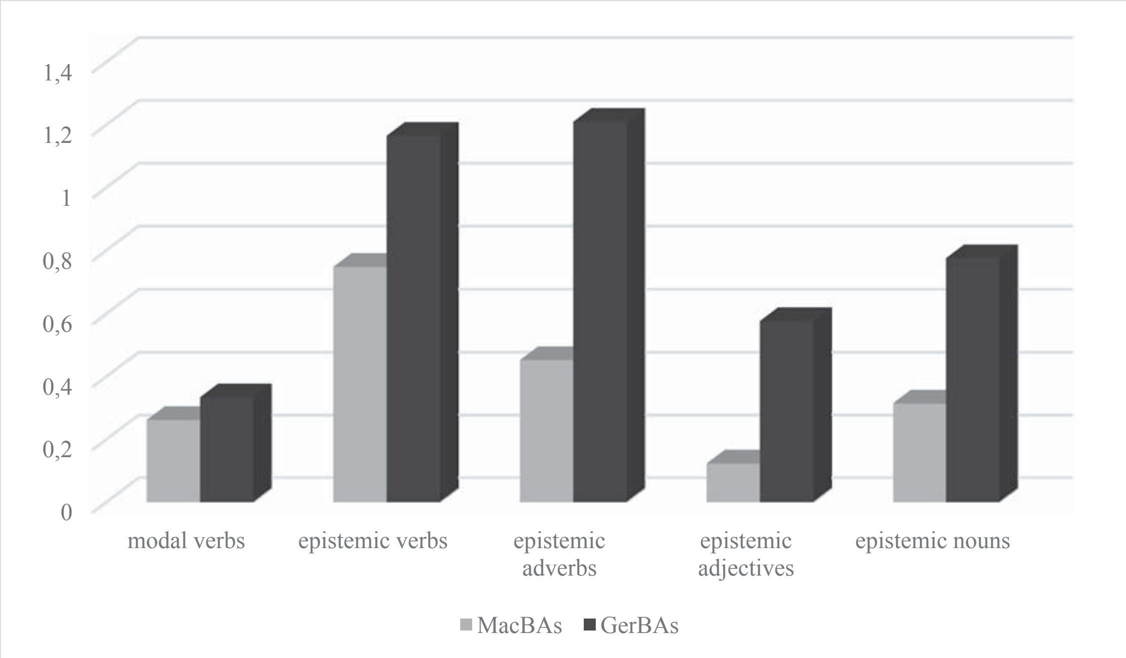


Figure 2: Overall distribution of boosters in the corpus (per 1,000 words)

The analysis, therefore, proves our first hypothesis correct because it shows that there are differences in the preferred subtypes of hedges and boosters across the two corpora (GerBA and MacBA). We believe that this is mostly result of the cultural differences or the different perception of this genre by the two societies. It might be that the students' cultural background or accepted norms at the university about the role of the writer in such types of written texts has influence on the preferred choice of language items for the presentation of their arguments. The instruction they get at their mother institutions i.e. language departments might play a role in this, too.

4.1. Modal verbs

The lexical category mostly associated with epistemic modality is modal auxiliaries. They "provide the least marked and thus the most straightforward means of expressing modality in English" with *would, will, could, may,* and *might* occurring most frequently in written discourse (Coates 1983: 23). Since, sometimes, it is really hard to distinguish among different meanings of modals, Palmer (1986) and Coates (1983) make a clear distinction between epistemic and deontic modality. We followed Coates's (1983: 23) framework of modality in deciding the epistemic function of modal verbs and, therefore, their pragmatic function of hedges and boosters in the texts.

4.1.1. *Modal verbs with the function of hedges*

Figure 3 below shows the overall distribution of modal verbs with the function of hedges in both corpora. The analysis shows that *might, may, could* and *would* were the most frequently used modal verbs as hedges in both the Macedonian and German corpus. *Should* and *ought to* were used very rarely. It is interesting to note down the significantly great use of the modal verb *could* (1.75 per 1,000 words) by the German students. It is not really clear why this might be the case, but it is obvious that writers find this modal verb most appropriate for expressing tentative possibility. They also use *would* more frequently than the Macedonian students (0.91 vs. 0.55 per 1,000 words) who, on the other hand, use *may* more often than the German students (0.43 vs. 0.32 per 1,000 words). What is also notable is the equal use of the modal verb *may* by both groups of students.

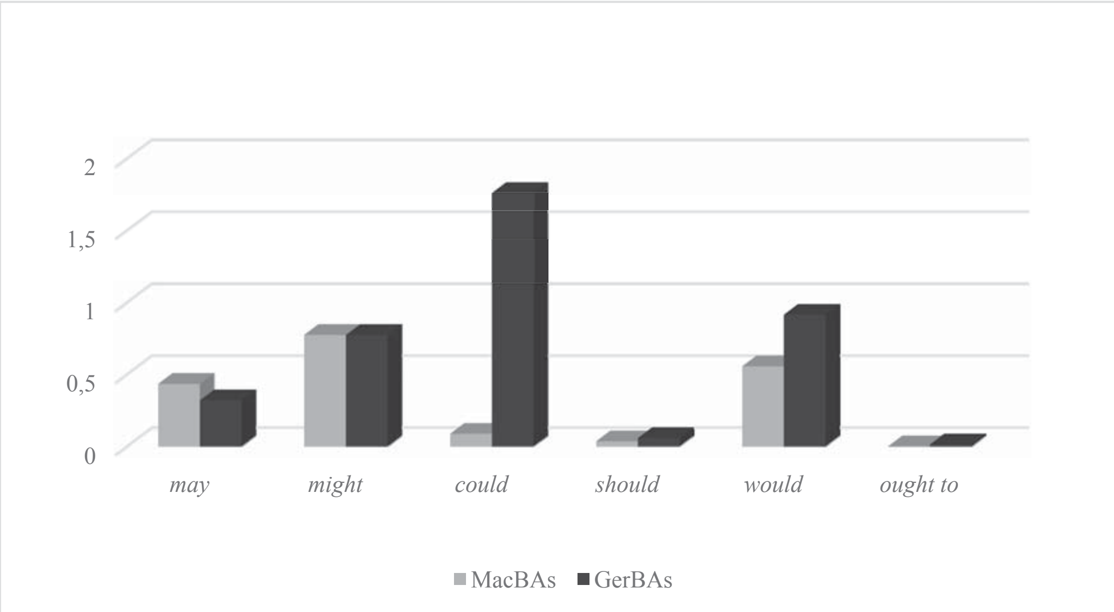


Figure 3: Distribution of modal verbs per 1,000 words

MAY /MIGHT

When used with their epistemic function, *may* and *might* express possibility (see examples 1-4). For instance, in (1) the writer states their opinion that there is a possibility of distinguishing between nationalities in future investigations.

(1) For further investigations it may be helpful to distinguish between distinct nationalities, which would also involve a larger sample size. (GerBA_01)
(2) However, the reluctant use of these phrases may contribute the "non-nativeness" of a text. (GerBA_03)
(3) It might be helpful to mention in some cases the item "еден" is used for emphatic purposes. (MacBA_8)
(4) Therefore, I would suggest that a larger sample size might have resulted in more reliable data. (GerBA_1)

There was one strange case (see example 5 below) of the use of *may* (as a hedge) with *I am convinced* (as a booster). It might be that the writer felt they have made too strong a claim so they decided to lower its effect with the use of a modal verb.

(5) I am aware that this definition of dank has a very strong connection to connotation, but I am convinced that it may be a denotation of its own too. (GerBA_9)

SHOULD

When *should* is used as hedge, it expresses the writer's tentative assumption based on inference. For instance, in (6), the writer expresses their assumption that something they stated before is probably true because 'the data shows a significant rise in population' (an inference they make).

(6) This should be true because the data shows a significant rise in population. (GerBA_1)

(7) This should come as no surprise, since a large number of the words that contain the prefix a- belong to the field of science, naming terms that are internationally accepted. (MacBA_4)

COULD

The epistemic function of *could* is the same with the one of *may/might*. When *could* is used as a hedge, it expresses tentative possibility (see examples (8) and (9)).

(8) The relative distribution of the different phrasal verbs varies significantly. This could be another reason why texts from L2 learners sound 'non-native'. (GerBA_3)

(9) One could talk about grammatical gender in terms of the Macedonian language, where all words related to the head noun change their form to conform to the grammatical category gender of the noun.(MacBA_9)

WOULD

Would in its epistemic function is used to express a hypothetical prediction *(I confidently expected/ I expect given unlikely conditions*). For instance, in (10), the writer states their expectation that both cases they mention before would have a 'significant linguistic and pragmatic impact on the participant's language behaviour'.

(10) Both cases would probably have a significant linguistic and pragmatic impact on the language behavior of the participant, which would falsify the results of this study. (GerBA_7)

(11) In fact, those English passive sentences are translated with active ones in Macedonian because passive sentences would be incorrect or literal. (MacBA_7)

4.1.2. *Modal verbs with the function of boosters*

When used in their epistemic sense, both *will* and *must* seem to be stronger modal verbs in illocutionary force than the previous ones. In addition, they are usually followed by strong epistemic verbs which generally have the function of intensifiers. Therefore, we classified them as boosters. They were very rarely used in the corpus. Both German and Macedonian students used *will* mostly, whereas *must* was used only by the German students.

WILL

In its epistemic sense *will* is used by the writers to make a prediction about the present based on repeated experience (*I confidently expect*). For instance, in (12), the student makes a prediction stating that most probably their paper will show that a certain linguistic choice is preferred by the speakers. In example (13), however, the distinction between the epistemic and non-epistemic use of the modal verb *will* is clearly presented. The first use 'the lexemes will be assessed' is deontic because it refers to a physical activity, while the second use 'the examples will illustrate' is what the writer confidently expects after the assessment.

(12) Probably, the paper **will be** able to show that when 'be like' is used it tends to be the preferred choice by the speakers. (GerBA_10)

(13) Each of the 8 lexemes will be individually assessed (*NON EPISTEMIC) and examples **will** illustrate (EPISTEMIC) the changes. (GerBA_9)

MUST

When used as a booster, *must* signifies the confident inference made by the writer based on deduction (*I am sure*). As a modal verb, it is much stronger in force than *will* and *would*. Examples (14) and (15) feature the use of *must.* As it can be seen, in both examples the writer makes an inference *(the influence must be secondary* (e.g. 14), *the data must be biased* (e.g.15)*)* based on some deduction *(the feature is spread in other varieties* (e.g.14), *the questionnaires were also biased* (e.g.15)*).*

(14) The greater frequency of contact clauses in Irish English suggest, […], however, this influence **must be** secondary since the feature is widely spread in other varieties. (GerBA_2)

(15) If the questionnaires were, as the students said, biased, then the data collected **must be** biased as well. (GerBA_4)

4.2. Epistemic verbs

Epistemic lexical verbs are very common means of expressing modality in written discourse. They are often referred to as 'speech act' verbs as they are used to perform acts such as doubting and evaluating rather than merely to describe acts.

4.2.1. *Epistemic verbs with the function of hedges*

In this paper we follow Hyland's (1998: 120) classification of two types of epistemic verbs with the function of hedges:

1. *judgemental verbs,* employed by writers to mitigate their claims by indicating that they are presenting information as a subjective opinion or a deductive conclusion. The most frequently used judgemental verbs, according to Hyland, are: (speculative) *indicate, suggest, propose, predict, assume, speculate, suspect, believe, imply,* (deductive) *imply, estimate, calculate* etc.

2. *evidential verbs*, which are used by writers to refer to evidentiary justification based either on reports of others, the evidence of the writer's senses or the feasibility of matching evidence to goals. The most frequent evidential verbs, according to Hyland, are: (quotative) *report, note,* (sensory) *appear, seem,* (narrators) *attempt, seek.*

As it can be seen from Figure 4 below, both Macedonian and German students used epistemic evidential verbs much more frequently than judgmental verbs. This is probably expected in this type of writing because writers are supposed to make deductions based on facts, (which in this case are probably the research results) and are not expected to make some vague judgements. The figure shows that the German writers use epistemic verbs more frequently than the Macedonian in most cases. The most frequently used verbs by them are: *seem* and *tend to* (evidential) and *argue, suggest* and *think* (judgemental). Macedonian students also opt for *seem* and *tend to* (from the evidential verbs) and *suggest, believe* and *guess* (from the judgmental verbs) more frequently. It is interesting to note that we have not come across the verb *think* in the Macedonian corpus.

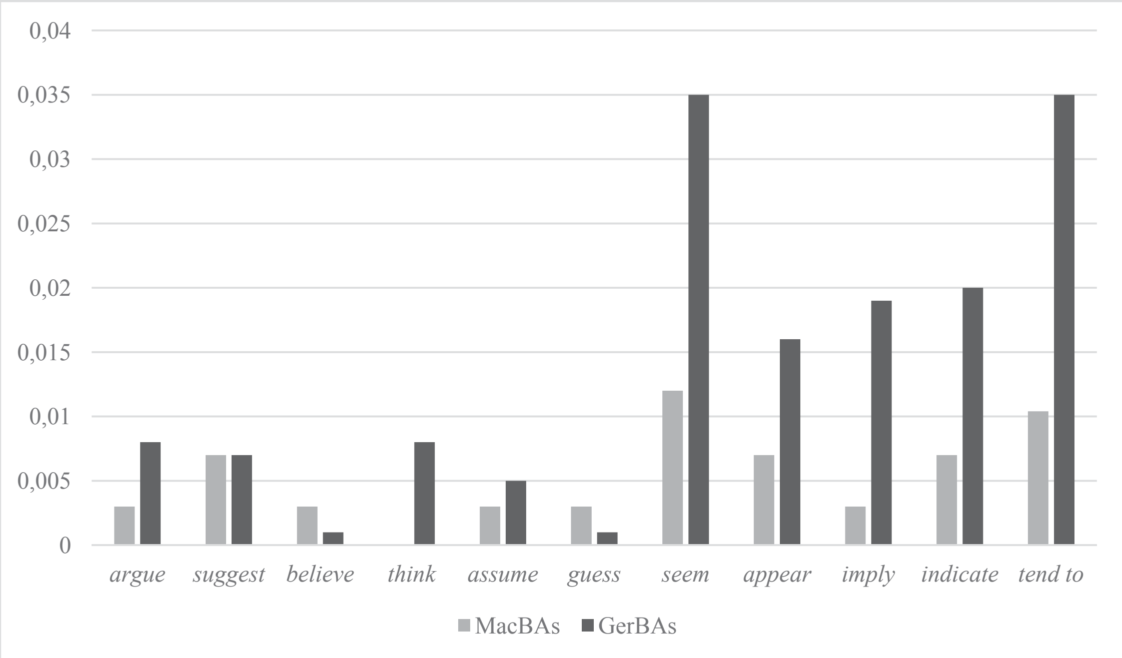


Figure 4: Distribution of epistemic verbs with the function of hedges (per 1,000 words)

4.2.2. *Epistemic judgmental verbs*

The examples (16) – (21) feature the use of the epistemic judgmental verbs: *think, suggest* and *assume.* In all of them the writers express their opinion about a certain situation based on some inference they make about it. However, it is clear that they hedge the statements because they fear that they might not have enough evidence to support them.

(16) I **think** that this small difference contributes to the fact that whenever is sometimes used in cases in which StE would prefer the use of when. (GerBA_10)

(17) The results of the analysis **suggest** that the English preposition "in" in most of the Macedonian cases is translated as "“во”. (MacBA_3)

(18) So far, these results **suggest** that ‘be like’ is the preferred choice by those people using it. (GerBA_11)

(19) I think it is safe to **assume** that the noun is not undergoing a change of meaning; neither a semantic shift nor semantic broadening. (GerBA_9)

(20) Therefore, it is likely to **assume** that 'be like' outnumbers the other quotatives and substitutes them in many functions (GerBA_11)
(21) While in the second one, we **assume** that the speaker is already familiar with the subject. (MacBA_8)

4.2.3. *Epistemic evidential verbs*

In comparison to the examples given above, in the examples (22) – (26), the use of the evidential verbs is presented. The writer here seems to have more evidence to make the inferences. However he/she still decides to make a tentative assumption and not a strong claim.

(22) It is difficult to assign a single translation equivalent to dis- in Macedonian. While it **seems** that the most frequent translation of words containing dis- is the Macedonian prefix не-… (Mac BA_4)
(23) Especially language differences and the resulting miscommunication between the genders **seem** likely to be discussed in public. (GerBA_1)
(24) This **implies** that the gender differences in hedging are not as obvious as they were expected to be. (GerBA_1)
(25) "Slightly", on the other hand, **appears** to be less uncertain because the adverb mostly compares one fact to another one. (GerBA_1)
(26) Some of the explanations mentioned above may **appear** to be too simple. (MacBA_4)

When doing an analysis of this kind, it is very important to be able to clearly distinguish between the verbs' metadiscourse function and their non-metadiscourse use. For instance, in example (27), *appear* does not have a metadiscourse function, because, compared to (25) and (26) it has the meaning of *show* and is not used by the writer to make any inferences.

(27) *…whole corpus directories downloaded with HTTrack are included and it **appeared** quite often that the same concordance hit was listed several times, which increases the total percentage of occurrences. (GerBA_6)

The can see the same situation with the use of the verb *indicate* in (29). In comparison to its usage in (28), where it has an epistemic function, in (29) it has the meaning of *show*, and is not used by the writer to make any inference based on evidence.

(28) This would **indicate** that significant findings are less hedged than minor important results**.** (GerBA_1)
(29) *The diagram above **indicates** the percentage of white clothing mentions in the descriptions.(GerBA_4)

4.2.4. *Epistemic evidential verbs with the function of boosters*

Figure 5 below shows that the most frequently used epistemic verb with the function of a booster in both corpora was *show*. German writers used it twice as frequently compared to Macedonian writers (0.86 vs. 0.42 per 1,000 words). Other verbs which were used with the function to boost statements were *prove, demonstrate* and *emphatic do*. However, their usage was much lower compared to the verb *show*. *Find, emphasise* and *point out* were used least frequently in both corpora.

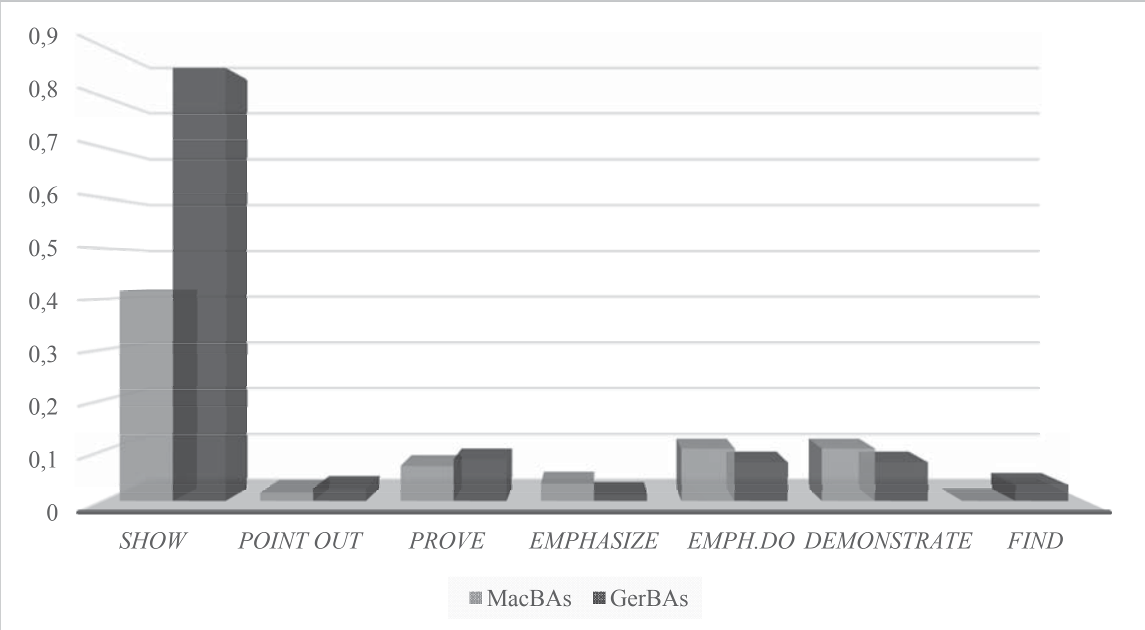


Figure 5: Distribution of epistemic verbs with the function of boosters (per 1,000 words)

The reason why writers used the verb *show* so often to intensify the force of their statements can be traced in its semantics. It means 'to demonstrate something by reasoning or procedure'. It is an evidential verb and the students used it mostly to strengthen their statements by referring to some evidence, as in (30). In (31), however, the writer might not have had very strong evidence so he/she decided to intensify the force of *show* by using the adverb clearly.

(30) To sum up, the examples **show** that the copula go is combined mainly with an adjective in the complement position. (MacBA_5)

(31) Suggest and believe clearly **show** the author's tentativeness and speculation on the subject being discussed, whereas conclude, in this example, is identified as a plausible deduction from propositions made before in the text. (GerBA_1)

The epistemic verb *demonstrate* (see examples (32) and (33) below) is semantically very close in meaning to the verb *show*, because it means 'to show to be true by reasoning or evidence; to prove'.

(32) Both examples above **demonstrate** that could is used in its epistemic sense because it can be replaced by the phrase it is possible that. (GerBA_1)

(33) On the other hand, it is **demonstrated** that -ess is the most common form among the feminine suffixes in English, but a reduction in the use of -ess forms over time is apparent (MacBA_11)

The verb *prove* is another epistemic lexical verb with a metadiscourse function of a booster. It is used 'to establish the truth or validity of something by presentation of argument or evidence'. Hence, when the writer in (34) uses it, they do not leave any doubt on the part of the reader that the results are not correct or that they don't show what they were supposed to show.

(34) As assumed before, the results **prove** that that is used more often in spoken language than in written, 100 times compared to only 38 times. In the North even more often than in the South. (GerBA_2)

4.3. Epistemic adverbs

Epistemic adverbs are the third category of markers which can have the function of both hedges and boosters. They are also used by the writers to show their confidence or lack of confidence in the truth value of their statements.

4.3.1. *Epistemic adverbs with the function of hedges*

The analysis of the use of adverbs with the metadiscourse function of hedges showed that they were more frequently used by the German students (Figure 6).

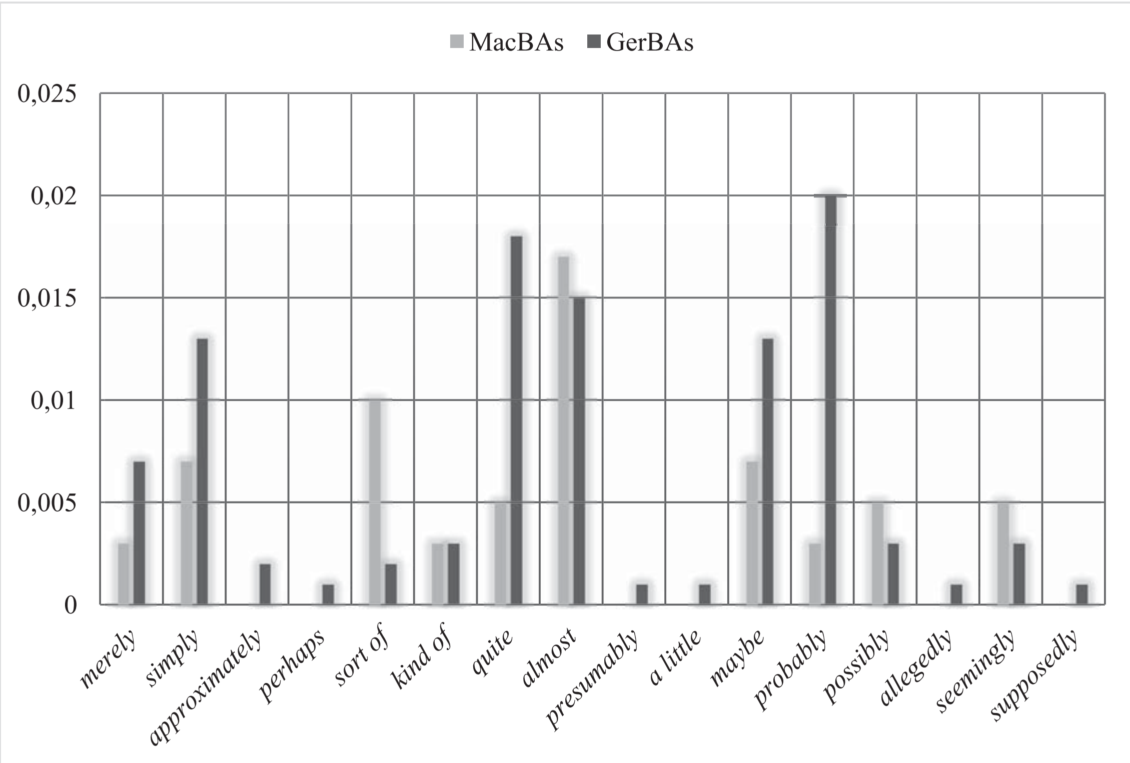


Figure 6: Distribution of epistemic adverbs with the function of hedges (per 1,000 words)

Only few were used more often by Macedonian students (like: *sort of, possibly, seemingly*). The most frequently used adverbs in the BA papers by both groups of students were: *almost, probably, quite, maybe, simply, merely*. The adverbs *possibly, probably, maybe, perhaps* and *presumably* are very similar in meaning. All of them were used by the writers to express their assumption about the likelihood of the proposition in a form of an opinion, rather than fact. They were used as sentence adverbs (see examples (35), (38), (39)) when they modified the whole sentence, or as modifiers of the verb (see examples (36), (37), (40), (41), (42)).

(35) **Possibly,** the program is constructed for a different type of corpus analysis. (GerBA_6)
(36) … we cannot **possibly** guess whether the referred subject is male or fem (MacBA_10)
(37) The main problem **probably** lies in their various uses or maybe the different …(MacBA_3)
(38) **Probably,** the paper will be able to show that when be like is used it tends to be the preferred choice by the speakers. (GerBA_11)
(39) **Maybe** the unfamiliar spelling could have an influence on that. (GerBA_9)
(40) The main problem probably lies in their various uses or **maybe** the different translations in different languages. (MacBA_3)
(41) Since the chosen skills are usually allocated to the gender role of men, status and hierarchy will **presumably** have an effect on the answer patterns. (GerBA_7)
(42) The lexeme came first into use in 1958 and **perhaps** it has the potential to trigger a full semantic shift concerning the denotation of clout. (GerBA_9)

The following three adverbs: *almost, approximately* and *quite* are somewhat different in meaning from the previous ones. They are adverbs of degree and when used as hedges (see examples (43), (45) and (47)) they show the writer's belief conerning the intensity or degree of the action expressed in the utterance. However, one should be very careful when trying to determine their metadiscourse function because they can easily be mistaken for the adverbs which do not have metadiscourse function. In such cases (see examples (44), (46) and (48)), they usually precede numbers or adverbs of quantity and are not used to serve the writer to lower the force of their claims (compare the use of *almost* in (44) and (43), approximately in (46) and (45), and quite in (48) and (47)).

(43) One might try to distinguish between the privative and reversal meaning, but such distinction is **almost** impossible to make, because the meanings are connected and overlap (MacBA_4)
(44) *Should is **almost** twice as frequent in spoken English as it is in written English, which is not surprising since should is a modal verb that is used in everyday language and is more common in colloquial style. (GerBA_8)
(45) Additionally, one participant provided the traditional dialect meaning and one described it as something less than human, which **approximately** fits the majority's definition. (GerBA_9)
(46) *In the German and American Corpus it was **approximately** as often at the beginning as in the middle of a sentence. (GerBA_5)
(47) Consequently, it is **quite** certain that the lexeme will disappear in the near future. (GerBA_9)
(48) *That is due to the fact that with antconc, whole corpus directories downloaded with HTTrack are included and it appeared **quite** often that the same concordance hit was listed several times, which increases the total percentage of occurrences. (GerBA_5)

4.3.2. Epistemic adverbs with the function of boosters

Epistemic adverbs with the function of boosters were also more frequently used by the German students (see Figure 7). *Especially* appeared to be used most frequently. Other adverbs which were also commonly found in both corpora were: *clearly, in fact, of course* and *definitely*. There was a number of more adverbs with the function of boosters which appeared in both corpora but their frequency was pretty low.

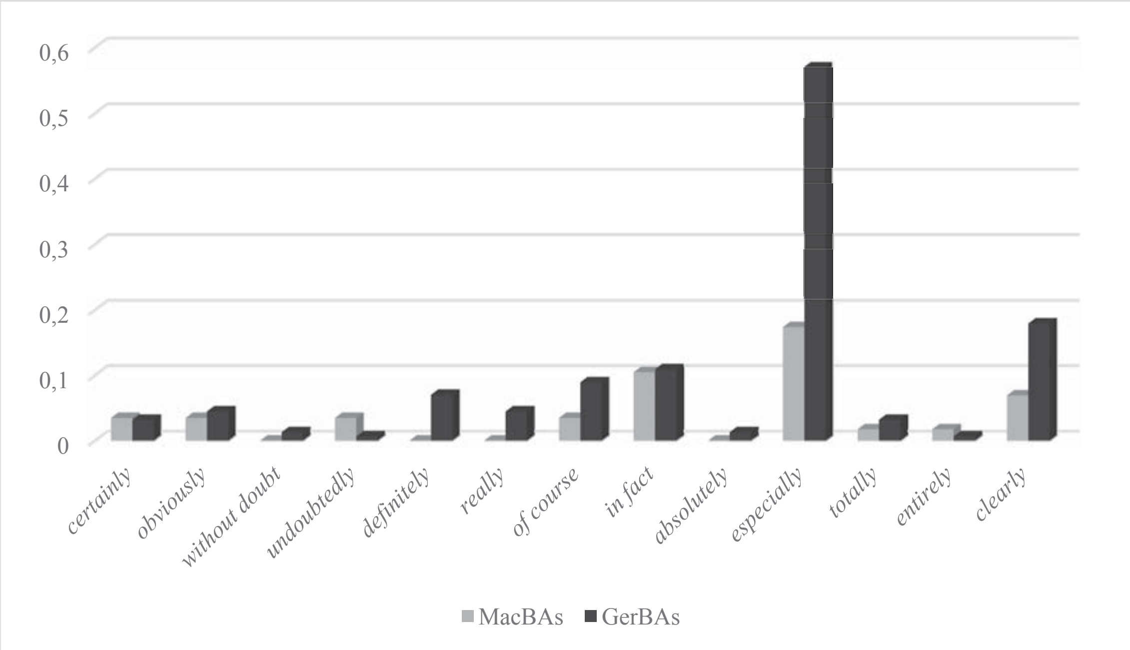


Figure 7: Distribution of epistemic adverbs with the function of boosters (per 1,000 words)

The examples (49) – (51) feature the use of the two most frequently used epistemic adverbs: *especially* and *clearly*, which were obviously used to strengthen the effect of the adjectives (as in (49) and (51)) or the verbs (as in (50)).

(49) It can be noticed from the examples above that it is **especially** difficult to create a female job titles that end in the suffix –… … (MacBA_10)
(50) Also, the comparison is crucial when it comes to proper translating because it **clearly** points out the different circumstances when English passive sentences can be translated with passive in Macedonian and when they cannot.(MacBA_7)
(51) However, differences are **clearly** recognizable among the groups. (GerBA_7)

4.4. Epistemic adjectives

Hedges and boosters can have the form of epistemic adjectives too. According to Hyland (1998: 130) they can serve to reduce or intensify the writer's categorical commitment to their statements in a written text.

4.4.1. *Epistemic adjectives with the function of hedges*

The most frequently used adjectives in both corpora were *likely* and *possible*. *Probable* was used only by German students and its usage was very low. It is important to mention that almost all these epistemic adjectives were used as subject complements in sentences with expletive 'it' (see examples below). They help the writer to express lack of commitment to their statements and present them as their tentative observation or assumption.

(52) It is **probable** that mistakes made during the design of the questionnaire were not spotted early enough. (GerBA_9)
(53) It is **possible** that the results would have been different when using another categorisation system. (GerBA_5)
(54) *It is **possible** in sentences with existential there and verbs like meet and have, but it is not possible after definite noun phrases. (GerBA_2)
(55) It is **likely** to be regarded as somewhat 'clumsy' if repeated frequently, especially where reflexive forms are involved: (MacBA_9)
(56) Therefore, it is very **likely** that those coming from rural areas might have … (GerBA_11)
(57) *Male students were very **likely** to start their descriptions with the couples' hair and its colour, whereas in the female group adjectives and adverbs were added to give a clearer picture. (GerBA_4)

When distinguishing their function, attention should be paid not to confuse their metadiscourse function with the non-metadiscourse one. For instance, if one compares the use of possible in example (53) to example (54), they would easily notice that in the latter the writer does not give their opinion on the possibility of the action expressed in the sentence, but just state that a certain item is possible in certain sentences. In other words, the writer does not talk about their own text here but rather about the content. The same goes for example (57). If one compares the use of *likely* in this example, with the one in sentences (55) and (56), they would notice that the writer again talks about the content here (*what male students were expected to do*) and not about their own text.

4.4.2. Epistemic adjectives with the function of boosters

The analysis of the use of epistemic adjectives with the function of boosters showed that the most frequently used ones in both corpora were *obvious, clear*, *certain* and *convince* (the last one was mostly used by German students).
When the writer uses '*I am convinced that*' as in example (58) below, they don't leave any doubt whatsoever to the reader that the information they present might not be true. Of course, the responsibility they take is also very high.

(58) I am aware that the study may not be representative for the whole population of Sheffield, but I am **convinced** that it is a first step to explain and assess semantic erosion in its lexis. (GerBA_9)

When the writer uses *certain* in phrases of the type: *I am certain* or *it is certain*, they state that their argument is indisputably correct and the readers should not hesitate to accept it (see for example (59) and (60)). However, one should be cautious not to mistaken the use of *certain* as a booster with its non-metadiscourse function as a pronoun, as in (61). In example (61), it has the meaning of '*some but not all'*.

(59) Additionally, I am **certain** that the influence of Standard English meanings is evident in this change and that traditional dialect meanings are replaced by Standard English ones. (GerBA_9)
(60) These societies are not quite socialized and it is for **certain** that unbending rules and rigorous norms exist there. (MacBA_10)

(61) *Phonological and grammatical dialect variation is given extensive research, but research in the field of lexical and semantic variation is sparse, and this field of study experienced **certain** neglect in the past. (GerBA_9)

4.5. Epistemic nouns

The use of epistemic nouns with the function of hedges and boosters was also analysed in the corpora.

4.5.1. Epistemic nouns with the function of hedges

The analysis showed that the most frequently used nouns with the function of hedges were *assumption, probability* and *possibility* in the German corpus, and only *possibility* in the Macedonian corpus. Similar to other types of words analysed, epistemic nouns were used by the writers to express the possibility about the truth value of the statement. For instance, in (62) and (63) the writer expresses their assumption about the probability of the action expressed in the statement which follows.

(62) Another **possibility** is that females had in fact imagined two prototypical associations for this exercise, which by some means points at their multitasking ability, but this issue shall be discussed in the following chapter. (GerBA_4)

(63) In general this leads to the **assumption** that in addition to technical details or shape, experiences with the object are stored and might even have a higher priority. (GerBA_4)

4.5.2. Epistemic nouns with the function of boosters

The only epistemic noun in the corpus which was used to intensify the action in the main clause was *fact*. Example (64) features its usage. The writer here states that the statement they made in the previous sentence is a fact, and it should be accepted as such with no hesitation.

(64) Also, it is not considered as a copula for formal usage and therefore it is predominately used in common daily communication. Due to this **fact**, it can be found in some idiomatic expressions. (MacBA_5)

4.6. Overall distribution of the hedges and boosters in the corpus

The overall distribution of hedges and boosters in the two corpora is presented in Figure 8 below. As it can be seen, German students used both of these metadiscourse markers more frequently than Macedonian students. Hedges were used more often by both groups of students and their usage is very significant in the German corpus (more than twice as frequently, compared to Macedonian corpus). It is obvious that German students try to be more cautious when presenting their arguments in the theses.

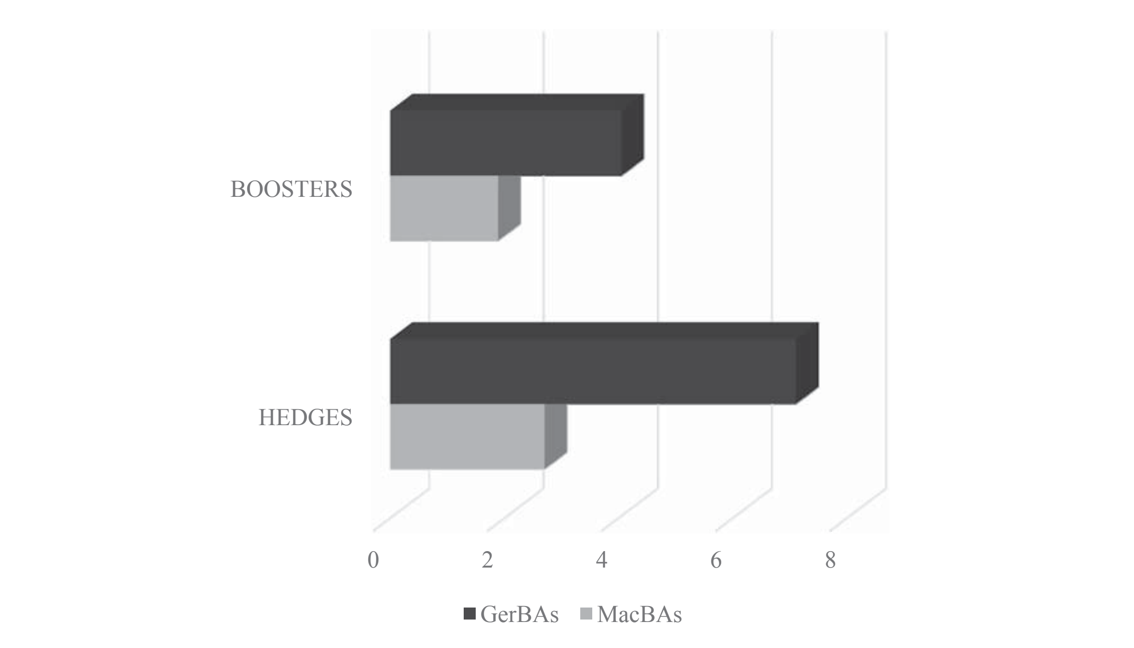


Figure 8: Overall distribution of hedges and boosters in the corpus (per 1,000 words)

These findings prove our second hypothesis true. We had these expectations due to our prior knowledge that throughout their studies, German students had been explicitly taught about the use of hedges and boosters as metadiscourse markers with persuasive function. It is obvious that instruction plays a great role in making students aware of the usage of the different categories and subcategories of metadiscourse markers.

5. Conclusion

This paper aimed to explore how Macedonian and German university students express commitment and detachment in their BA graduate papers through the use and pragmatic function of two interpersonal metadiscourse markers: *hedges* and *boosters*. It focused on identifying the different forms with which these markers are expressed and distinguishing between their metadiscourse and non-metadiscourse function. The findings proved both our initial hypotheses correct. It showed that both German and Macedonian writers used hedges and boosters in their theses, but that there were variations in their distribution and pragmatic function, based mostly on the two cultures' different value systems concerning the format of this genre in the two societies. Furthermore, as expected, the findings showed that German writers used these markers more frequently than Macedonian writers. It is assumed that the reason for this is that they had been explicitly taught about their form and usage as part of their academic writing curriculum.

We believe that in order to confirm these findings from the analysis, that a further analysis on a larger corpus and different disciplines should be done. Finally, we also recommend that students should be explicitly instructed throughout their studies about the form and function of metadiscourse markers because we believe that it would help them to improve their writing style and be more persuasive for the readership.

References

Brown, P., Levinson, S. C. (1978). *Politeness. Some universals in language usage.* Cambridge: Cambridge University Press.

Coates, J. (1983). *The semantics of the modal auxiliaries*. Beckenham: Croom Helm.

Crismore, A., Markkanen, R. & M. Steffensen. (1993). Metadiscourse in persuasive writing. A study of texts written by American and Finnish university students. *Written Communication*, *10 (1),* 39–71.

Crismore, A., Farnsworth, R. (1990). Metadiscourse in popularand professional science discourse. In W. Nash (ed.). *The writing scholar.* Newbury Park: Sage, 118-136.

Crismore, A., Vande Kopple, W. J. (1988). Readers' learning from prose: The effects of hedges. *Written Communication,* 5, 184-202.

Halliday, M. A. K. (1974). The place of functional sentence perspective in the system of linguistic description. In F. Danes (ed.). *Papers on functional sentence perspective.* Prague: Academia, 43-53.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing.* London: Continuum.

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

Hyland, K. (2000). Hedges, boosters and lexical invisibility: Noticing modifiers in academic texts. *Language Awareness, 9 (4),* 179-197.

Hyland, K. (1998). *Hedging in scientific research articles.* Amsterdam & Philadelphia: John Benjamins Publishing Company.

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2,* 458-508.

Lakoff, G. (1972). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Chicago Linguistic Society Papers 8,* 183–228.

Nivales, L. M. (2011). Hedging in college research papers: Implications for language instruction. *Asian EFL Journal, 35-45.*

Palmer, F. (1986). *Mood and modality*. Cambridge: Cambridge University Press.

Pungă, L. & Borchin, M. (2014). The metatextual role of introductions to BA papers. *Romanian Journal of English Studies. Volume, 11 (1),* 36–47.

Schmied, J. (2015). Graduate academic writing in Europe: A research-based approach to metalanguage and genre. This volume, 1-24.

Trajkova, Z. (2011). Toning down statements in newspaper editorials. *The Journal for Languages and Literatures of the Faculty of Philosophy in Novi Sad, Vol I,* 71-84.

Vande Kopple, W. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication, 36,* 82-93.

Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes, 20,* 83-102.

Williams, E. (1981). On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry, 12,* 245-247.

Structural, Textual and Commentary Linkers in Albanian PhD Dissertations: A Pilot Study

Bledar Toska (University of Vlora "Ismail Qemali")

Abstract

This short pilot study investigates the use of structural, textual and commentary linkers in Albanian PhD dissertations. Based on the systemic functional approach, it aims at exploring how these linkers are employed in these dissertations to establish logico-semantic relations between parts of texts in expansion through elaboration and enhancement as well as in projection. It is also discussed how they operate at the local and global levels of text to enable its internal structural organization and the interpersonal. The Albanian Dissertation Corpus (ADC) built for this article contains works from four major disciplines: economics, humanities, natural sciences and social sciences. The quantitative and qualita tive analysis of the linkers in dissertations across these disciplines shows some frequency differences as well as some use preferences, particularly in humanities and natural sciences. Although not an in-depth study, it attempts to concentrate on relevant use aspects of various linkers in this genre of academic writing in Albania.

1. Introduction

Works on academic writing are not new to the academic circles and they have had a major impact on applied linguistics, particularly in the last decades. As Schmied observes, "academic writing has established itself almost as an independent discipline in applied linguistics" (2011: 1), a discipline which is interested not only in linguistic aspects of language use but also in social, cultural or cognitive features frequently influencing some of the choices scholars make in academic writing (Hyland 2009). The focus in this study is on PhD dissertations, as central to the written academic genres and to the academic enterprise. In this regard, doctoral students are expected to "develop their own ideas, to pursue their own research agenda and to write up everything in a major contribution, which is the result of some sophisticated Ph.D. project" (Schmied 2011: 2), involving a high degree of specialist language competence (Hyland 2009).

It would be more appropriate to consider academic writing to be a linguistic and extralinguistic process, a discourse referring "to the ways of thinking and using language which exist in the academy" (Hyland 2009: 1) to interact with the scholarly community. Hůlková states that its chief function is referential, "which means that the overall impact of any piece of discourse produced within the register in question should be explicit, unambiguous and logical" (2011: 129) as well as relevant to the conventions other scholars expect to be followed. Connectivity is essential when scholars attempt to provide a well-structurally organized academic prose. It enables explicit and logical links among parts of a text facilitating its comprehension and interpretation (Toska 2012).

The resources that I am interested in in this study are those linguistic units that are traditionally called conjunctions and adverbials. For the sake of simplicity, I will use the term *linkers* here and I will categorize them in three major groups, according to their functions: *structural linkers* including items such as *because*, *although*, or *while*, *textual linkers* including *however*, *anyway* or *thus* and *commentary linkers* including *certainly*, *naturally* or *fortunately*. In terms of the connectivity that they establish in text and discourse, structural linkers enable relations between clauses in a sentence, textual linkers enable intersentential relations and commentary linkers enable interactive relations between the writer and the reader. There will be some more discussion about some of their other aspects, particularly in PhD dissertations that will be analyzed in the third and salient part of this article.

The purpose of this small-scale pilot study is to investigate in general terms the use of these three types of linkers in dissertations across some major disciplines and to discuss relevant contrastive aspects of use in relation to the choices made in Albanian PhD dissertation. In addition to the comparative statistical analysis, the study also intends to somehow highlight particular use tendencies exhibited by certain linkers in relation not only to paradigmatic but also syntagmatic elements in a text. Only some of the most important differences will be included in the analysis part, but recommendations will be made for further investigation in future studies.

Unfortunately, academic writing, academic discourse and particular theoretical and practical issues related to them have received very little attention in Albanian scholarly circles. The principal purpose of this short study is to investigate perhaps one of the most important aspects in dissertations, namely, that of connectivity by means of linkers. Although these resources have been recently explored in some very good studies from textual and discoursal perspectives in Albania (cf. Dibra & Varfi 1999; Topalli 2011; Toska 2012; Mëniku 2013; Dauti 2014), they have hardly been investigated in academic genres. My main motivation for conducting this study is to initiate a field of study which hopefully can be extremely valuable for future research in applied linguistics, particularly for novice scholars interested in textual connectivity issues.

The theoretical framework adapted in this paper is the systemic functional approach (Halliday & Matthiessen 2004), which sees language as a network of interrelated systems, a set of options enabling language users to express themselves and meet their communicative needs. "The lexicogrammatical system as a whole, operates as the realization of the semantic system, which is what the speakers *can mean*" (Halliday 1978: 39), what is referred to as meaning potential. From this perspective, grammar is interpreted functionally as a powerful resource for construing meaning. Linkers also form an independent lexicogrammatical system operating both at the local and global level of texts to produce intended meanings out of all the meanings the other potential paradigmatic choices within the same system can produce. The notion of choice is very important here, since, as Halliday claims, "meaning-making power of language is activated by the choices we make" (2013: 36).

The next part of this article provides a detailed description of the data collection process and the methodology followed for the analyses. The results obtained are discussed quantitatively and qualitatively in the third section instantiated with several examples taken from ADC. Some conclusive remarks and possible directions for further future research are included in the last part of the article.

2. Data and Methodology

To analyse the use of structural, textual and commentary linkers in dissertations, I built the ADC, which includes dissertations from four main disciplines: economics, humanities, natural sciences and social sciences, written in Albanian during the last two years (2012-2014). They were downloaded from the University of Tirana website (http://www.unitir.edu.al/), originally in pdf files, and were later converted to texts to enable the corpus analysis. Four sub-corpora were also built, each of which consisting of 40 dissertations, to facilitate the comparative analysis of linkers across these disciplines. ADCE includes dissertations in finance, management and marketing, ADCH in linguistics and literature, ADCN in chemistry, biology, physics and mathematics and ADCS[1] in political sciences, philosophy, sociology, psychology and pedagogy. The size of each sub-corpus is indicated in Table 1 below.

disciplines	dissertations	words
Economics	40	2,264,596
Humanities	40	3,464,509
Natural Sciences	40	1,704,504
Social Sciences	40	2,234,393
total	**160**	**9,668, 002**

Table 1: Academic disciplines, number of dissertations and words in ADC

The four corpora were analyzed separately in the versatile commercial software WordSmith Tools 6. The normalized frequencies for the three types of linkers are shown in Table 2 and Table 3 in the next section of the paper. The overall number of linkers analyzed is 108 (66 structural and textual and 42 interpersonal), listed and categorized in the appendix along with the English equivalents. The choice of the linkers was made based on some personal intuition as to their use frequency in Albanian and dissertation genre. The obtained quantitative results will be compared and qualitative analyses will be discussed in

[1] ADCE stands for Albanian Dissertation Corpus of Economics, ADCH for Albanian Dissertation Corpus of Humanities, ADCN for Albanian Dissertation Corpus of Natural Sciences and ADCS for Albanian Dissertation Corpus for Social Sciences.

order to obtain some useful insight into the use of linkers in the above mentioned disciplines.

3. Results and Discussion

This salient section of the work attempts to define and highlight some of the main functions and characteristics of the three types of linkers considered here. The first subsection compares and contrasts the use of structural and textual linkers in PhD dissertations across the four disciplines. The second subsection considers textual and discoursal uses of commentary linkers in the four sub-corpora instantiated with some relevant examples.

3.1. Structural and textual linkers

Following the Systemic Functional Grammar approach, Halliday and Matthiessen (2004: 373-383) state that there are two basic systems determining the way clauses are related to each other: (i) the degree of interdependency and (ii) the logico-semantic relation. The former includes parataxis and hypotaxis, what are commonly known as coordination and subordination in traditional grammar. The latter includes two general types of relations: expansion when the secondary clause expands the primary clause by elaborating, extending or enhancing it and projection when the secondary clause is projected through the first clause.

Structural linkers enable both paratactic and hypotactic relations within complex sentences, as illustrated in the following examples:

(1) Kryesore për autorët modernistë është që bota gjendet e ndarë në dy pjesë, domethënë ekziston hierarkia dyshtresore e realitetit. (ADCH_03)
[It is important for the modernist writers that the world is divided into two parts, that is, there exists the double-layered hierarchical reality.]

(2) Nuk mund të përgjithësojmë kaq shumë një fenomen vetëm sepse adoleshentët apo një grup i caktuar shoqëror preferon të ndjekë atë lloj modeli. (ADCH_22)
[We cannot generalize a phenomenon so much only because teenagers or another social group prefers to follow that kind of model.]

Paratactic linkers such as *domethënë* (*that is*), as employed in the first instance, “have a purely logical connective function” (Morley 2000: 44) and express “the logical relationship between two clauses of equal structural status” (Eggins 2004: 264). On the other hand, as it is illustrated in the next example, hypotactic linkers such as *sepse* (*because*) “in addition to their binding function, … introduce … the adverbial subordinate clause and mark the content nature of its circumstantial relationship”, such as reason, condition, time etc. (Morley 2000: 44). Unlike textual linkers, structural ones establish relations between clauses in a complex sentence, that is, they have an intrasentential function. We will see, however, that they not only create structural links within complexes, but also assist the interpretation and comprehension of that part of text.

Textual linkers have intersentential functions and establish cohesive links between the parts of a text, as exemplified below:

(3) Në vendet e rajonit aplikohen dy sisteme të penzioneve, sistemi i investimeve dhe sistemi i transfereve. *Prandaj* edhe sa i për kete vendeve të rajonit që aplikojnë skemës së pensioneve, ekzistojnë dallime në mes tyre. *Kështu*, disa nga vendet aplikojnë skemën pensional të investimeve do të thotë paguajnë kontribute të detyrueshme në shtyllën e dytë, siq janë: Bullgaria filloi të aplikoj këtë sistem nga viti 2000, Kroacia nga viti 2002, dhe Maqedonia nga viti 2005. (ADCE_31)
[In the countries of the region two pension systems apply, the investment system and the transfer system. *Therefore*, there exist differences even as regards the countries in the region that implement the pension system. *So*, in some of the countries the investment pension system applies, which means they pay obligatory contributions in the second column, for instance: Bulgaria began to apply it in 2000, Croatia in 2002, and Macedonia in 2005.]

The notion of function in these cases is extremely important, since textual linkers are considered cohesive devices not because of the semantic relations they establish, but because of "the function they have of relating to each other linguistic elements that occur in succession but are not related by other, structural means" (Halliday & Hasan 1976: 227). Having said that, *prandaj* (*therefore*) and *kështu* (*so*) in the example above show the link between the sentences they precede and the previous one, by making it clear for the reader how the propositional content of the second sentence "fits into the text" (Bloor & Bloor 2004: 54).

What is more, textual linkers as cohesive means, "clearly contribute to the perception of a text as coherent on condition that they are used appropriately by the writer and interpreted correspondingly by the reader(s), i.e. in agreement with the writer's communicative intentions" (Povolná 2012: 132). In the given example these linkers enhance the meanings of the sentences by reference to cause and allow the writer to express explicitly the intended meaning out of a potential linguistic system of choices, which is also "the basic tenet of" systemic functional approach to texts (Fontaine 2013: 1).

Both structural and textual linkers can be employed to enable logico-semantic relations; the former between clauses in sentences and the latter between sentences in text. As pointed out, this is known as expansion, when the second linguistic structures (i.e. clauses or sentences) expand the first ones, as exemplified in examples (1), (2) and (3). Two of the three types of expansion, elaboration and enhancement, will be part of the comparative analysis, while extension has not been included in this study. Structural and textual linkers which operate within the extension type are very frequent in language and include important items such as *and*, *but*, *or*, *moreover*, *however*, etc. Since some of them not only link clauses to clauses and sentences to sentences but also words to words (as instantiated in the following example), it was impossible to distinguish in the data results their linking function between clauses/sentences or words. The latter are not units which fall within the scope of this study.

(4) Kam kryer dy vizita shkencore *dhe* kontakte të herëpashershme që u finalizuan me botime të përbashkëta. (ADCN_39)

[I have had two scientific visits *and* continuous contacts which have been finalized with joint publications.]

The types of linkers establishing logico-remantic relations in expansion by means of elaboration which are found in ACD are almost all textual, signifying that the description or the specification of the meaning of a sentence by another is done at the textual level rather than at the local or structural one. I present the normalized frequencies of such occurrences for 10,000 words in Table 2 and the distribution of them in elaboration in Figure 1.

disciplines	**elaboration**	**enhancement**
Economics	4.3	137.9
Humanities	6.5	162.6
Natural Sciences	2.1	87.8
Social Sciences	3.7	155.8

Table 2: Normalized frequencies of structural and textual linkers in expansion (tokens per 10,000 words)

Figure. 1: Distribution of structural and textual linkers in elaboration and enhancement in ADC sub-corpora

It can be noted that there is clearly a strong preference for employing far more linkers in dissertations in Humanities. Statistically speaking, their use is three times more frequent in ADCH than in ADCN and obviously much higher than in ADCS and ADCE. This indicates that there is a common tendency in ADCH to elaborate on the meaning of sentences, which should not be surprising, since it consists of dissertations in linguistics and literature which need to elaborate statements further, as it is illustrated below.

(5) Konteksti i universitetit shqiptar është ai ku një orë mësimi ndiqet nga studentë që kanë të njëjtin formim arsimor. *Me fjalë të tjera*, edhe pse synohet për të arritur një diversitet prejardhjeje arsimore në degën e përkthimit, në nivel Master, ora e lëndës së përkthimi, aktualisht frekuentohet nga studentë që kanë përfunduar ciklin e parë të studimeve pranë Fakutetit të Gjuhëve të Huaja. (ADCH_13)

[The Albanian university context is such that a class is attended by students of the same educational formation. *In other words*, even though it aimed to attain a diversity of educational background in the translation degree programme, at the master's level translation classes are actually attended by students who have already finished their BA degree at the Faculty of Foreign Languages.]

(6) Ndikimi i provimeve mund të zgjasë për një periudhë kohore të shkurtër, apo të gjatë. *Për shembull*, nëse ndikimi i një provimi që kryhet për efekt pranimi në universitet, apo në një punë të caktuar, është i pranishëm vetëm gjatë periudhës kur nxënësit përgatiten për testim. (ADCH_16)
[The influence of exams can last for a short or a long period of time. *For instance*, if the influence of an exam taken for admission purposes at a university or for a vacancy, is present only during the period when students prepare for the exam.]

It also seems that textual linkers in elaboration which restate or exemplify such as *domethënë* (*that is to say*) or *për shembull* (*for instance / for example*) are the most frequent ones in all sub-corpora. For instance, in ADCH there are 41 occurrences/per million words, in ADCE 23, in ADCN 17 and in ADCS 24, which, coincidentally, are not very far from the distribution percentage of these linkers in elaboration found in the entire ADC (Figure 1 above). Although this might not be very significant, the fact that textual linkers in ADCH point more commonly to the re-presentation, restatement and exemplification of meanings in parts of texts indicates the importance of more elaborative global structures in humanities dissertations. As Ventola and Mauranen state, "the generic organization in terms of global structures plays an important role in text formation, and specifically in the way such structures are realized through discourse choices" (1991: 462), which enable writers to choose what to say and "also choose what to mean" (Halliday 2013: 17).

Structural linkers are more frequently used in CAD to establish logico-semantic connectivity in expansion by means of enhancement, particularly in hypotactic relations, which include some very common ones such as *nëse* (*if*), *kur* (*when*), *sepse* (*because*) and *ndërsa* (*while*). Unlike in the case of elaboration, use differences found in the four sub-corpora are less significant. The chart in Figure 1 shows that linkers in enhancement are more evenly distributed in ADC. Linkers used in ADCH, ADCS and ACDE have almost the same percentage frequency, that is 30 %, 29 % and 25 % respectively. Also, these linkers are less preferred in ADCN, even though they are employed more frequently here than in elaboration. The sub-corpus exhibits around 16 % of all linkers in ADC, which is expected, since, as in the case of elaboration, hard sciences require fewer linkers to establish explicit structural and textual connectivity.

However, a closer look at the sub-corpora examples shows that the differences of degree in enhancement are less important than the types of linkers that are preferred in each sub-corpus. I will analyze only two frequent categories of linkers: conditional structural linkers and causal textual linkers.

It appears that the conditional hypotactic relations between clauses in hard sciences do not play an important role in the interdependency structure of sentences, since there are only 4.5 uses of linkers (per million words) such as *nëse*, *në qoftë se* and *po qe se* (*if/whether*) in ADCN. This can also show that clauses that complement the meaning of each other and that of the text at the global level have little reference to conditional circumstances. Most of the occurrences are purely of the formal logical argumentation or similar natural phenomena, as instantiated below:

(7) Vlera e D luhatet midis vlerës 0 dhe vlerës 1 (0 = D = 1). *Nëse* vlera e D afrohet shumë afër vlerës zero, kjo i korrespondon një ekosistemi heterogjen ose me diversitet llojor të lartë; e kundërta, *nëse* vlera e D afrohet më shumë vlerës 1, atëherë kjo i korrespondon një ekosistemi më shumë homogjen dhe me diversitet llojor të ulët. (ADCN_09)
[The D value varies between 0 and 1 values (0=D=1). *If* the D value approaches value 0, this corresponds with a heterogeneous ecosystem or high diversity of kinds; alternatively, *if* the D value approaches value 1, then this corresponds with a more heterogeneous ecosystem and low diversity of kinds.]

Conditional hypotactic relations established by such linkers in the rest of the sub-corpora make up almost 90 % of the use frequency in ADC.

The enhancement process is also common intersententially, namely at the global level of texts, in which case textual linkers are effective means to signal logical and structural relationship between parts in academic prose (Hůlková 2011: 130). Causal textual linkers in ADC are general and specific "discourse choices which account for the cohesive semantic organization of a text" (Ventola & Mauranen 1991: 462), which is also proved by their relatively high frequency in the corpus, particularly in ADCH and ADCE. Examples (8) and (9) provide two instances of such occurrences.

(8) Këto parafjalë dallohen për larminë që shfaqin në përdorim, si dhe për ndërfutjen e plotësuesve të foljes. *Pra*, këto parafjalë kanë një përdorim mjaft të gjerë si në gjuhën e folur, ashtu edhe në atë të shkruar. (ADCH_35)
[These prepositions are distinguished for their variety of use as well as for complementing the verb. *Thus*, these prepositions are widely used in the spoken and the written language.]

(9) Në këtë artikull përshkruhen portofolët e aseteve me anë të mesatareve të rendimenteve, variancës së tyre dhe korrelacionit ekzistues. *Për këtë arsye*, kjo metodë njihet gjithashtu me emrin 'analiza mesatare-variancë'. (ADCE_01)
[In this article the portfolios of assets are described by means of the efficiency means, their variance and the existing correlation. For this reason, this method is also known as 'means-variance analysis.']

Yet, they are not frequently employed in ADCN (only 18% of all occurrences in ADC), which I believe, it is due to similar reason which I mentioned above in the case of conditional structural linkers.

3.2. Commentary linkers

Various terms have been used in the literature to refer to linguistic devices such as *unfortunately*, *probably*, or *supposedly*. For instance, Halliday and Matthiessen (2004) call them *modal adjuncts*, Eggins (2004) *comment adjuncts*, Biber et al. (1999) *stance* or *attitude adverbials*, Fraser (1997) *commentary pragmatic markers* and Morley (2000) *interpersonal adjuncts*. I use the term *commentary linkers* in this article because, as we will see, their main function is

to link the speaker's comment or attitude with the sentence's content. Because they can also serve as units which enable interaction between the writer and the reader, they are also "considered interpersonal elements in the clause, since they add an expression of attitude and evaluation" (Eggins 2004:161).
I also follow Halliday and Matthiessen's categorization of commentary linkers into two broad types: propositional and speech-functional. The first type expresses the speaker's comment on the proposition and the second his or her attitude towards the speech function (2004: 129-132). Examples (10) and (11) illustrate these functions of commentary linkers.

(10) Një pozicion i tillë etik, që lidhet siç thamë edhe me interesat, bën që të vërtetat të deformohen duke ndikuar kështu ndjeshëm mbi rezultatet: nga 27 deri në 57%. *Pa dyshim* që problemi ka zgjidhje. (ADCS_06)
[Such an ethical position, related as mentioned earlier also with the interests, distorts the truths thus significantly influencing the results: from 27 to 57%. *Obviously*, there is a solution to the problem.]

(11) Nëse shkolla e praktikon diversitetin si vlerë nën modelin një shembull të shkëlqyer që vjen nga SHBA, zgjedhjen e Barak Obamës, një emigranti me ngjyrë si President. *Personalisht* një modul të tillë e shoh me shumë vlerë. (ADCS_17)
[If school practices diversity as value under the model a good example comes from the USA, the election of Barack Obama, a black emigrant, as President. *Personally*, I regard this module as noteworthy.]

In (10) *pa dyshim* (*obviously*) asserts the propositional content of the sentence it precedes and in (11) *personalisht* (*personally*) engages the writer personally in expressing an individual comment on the sentence content. In other words, the comment signaled by the propositional commentary linker is that the writer believes that the problem can be solved, while the speech functional commentary linker is employed to foreground the individuality nature of the content. Indeed in most contexts, commentary linkers are intentionally employed to "specify the speaker's own assessment of the probability of the central proposition" and to "express his/her personal attitude towards or comment on the proposition" (Morley 2000: 105).

And finally, commentary linkers can operate within the semantic domain of projection, manifested interpersonally. For instance *pa dyshim* (*no doubt*) in (10) has the same function as *Nuk kam dyshim që* (*I don't doubt that*) and *personalisht* (*personally*) in (11) can be replaced with *për mua* (*to me*). Apart from highlighting the fact that this network system of linkers establishes lexico-semantic relations differently in a text, namely through projection and interpersonal connectivity, it is not my intention to explore additional issues related to it in this pilot study (cf. Halliday & Matthiessen 2004: 603-613).

The results in Table 3 and Figure 2 show that commentary linkers are not very common in ADC. This is due to the fact that they tend to operate on the realm of the interpersonal, and dissertations should offer "an objective description of what the natural and human worlds are actually like" (Hyland 2005: 65), although this varies from one discipline to the other. For instance, propositional commentary linkers in ADCH are three times more frequent than in ADCN, presumably because writers are more inclined to employ them in dissertations in literature or linguistics rather than in mathematics or physics. What is more, since the former are more subjective in nature and the latter more objective, the writer bears in mind that the choice of employing or not these linkers should focus “more directly on the participants of the interaction, with the writer adopting an acceptable persona and a tenor consistent with the norms of the community” (Hyland 2005: 54). As Halliday wisely puts it, whenever something is chosen, something else is not chosen; that is the meaning of choice. What is not chosen, but could have been, is the meaning of that particular instance of choice” (2013: 25-26).

disciplines	**propositional**	**speech-functional**
Economics	3.4	6.1
Humanities	6.5	5.8
Natural Sciences	2	3.7
Social Sciences	4.2	4.2

Table 3: Normalized frequencies of commentary linkers in projection (tokens per 10,000 words)

Figure 2: Distribution of commentary linkers in ADC sub-corpora

A closer look at some frequent propositional commentary linkers in ADC shows that some of them are far more preferred in one sub-corpora rather than the rest. For instance, *ndoshta* (*maybe/perhaps*) is one of them in ACDH (41 % of all occurrences in ACD versus 19 % in ACDE, 18 % in ACDN and 22 % in ACDS). Indeed, works in literature and linguistics are likely to use more hedges focusing on an aspect of the proposition by playing it down (Morley 2000: 106). The following two examples demonstrate this.

(12) *Ndoshta*, një faktor shumë i rëndësishëm në mungesën e krijimit të kësaj lëvizjeje vendase, ishte edhe mungesa e ekzistencës së teksteve dhe librave teorikë të përkthyer ose jo që mund të hidhnin dritë mbi fenomenin e surrealizmit. (ADCH_05)
[*Maybe* a very important factor in the lack of the creation of this local movement was even the lack of texts and theoretical books, translated or not, which could shed light on the surrealism phenomenon.]

(13) Një idiomë në anglishte mund ketë strukturim të rregullt a të parregullt dhe *ndoshta* edhe një strukturë gramatikore jo korrekte. (ADCH_08)
[An idiom in English can have a regular or irregular structuring and *perhaps* even an incorrect grammatical structure.]

On the other hand, *natyrisht* (*naturally*), an assertive propositional commentary linker is very common in ADCS compared to the other sub-corpora (42 % out of all occurrences in ADCS versus 17% in ACDE, 18 % in ACDN and 23 % in ACDH). As Morley observes, such linkers are employed to “state the terms of the speaker's referential stance” (Morley 2000: 106), as instantiated in the following extract.

(14) *Natyrisht*, përkujdesja për fëmijët ka karakteristikat e veta, të cilat shkojnë përtej qëllimeve të këtij studimi. (ADCS_27)
[*Naturally*, child care has its own characteristics, which go beyond the scope of this study.]

There are also frequency differences in the use of speech-functional commentary linkers, as shown in Figure 2 above. However, the types of linkers employed and their functions in texts are perhaps more interesting to be analyzed here. Commentary linkers such as *në të vërtet*, *në fakt*, *realisht* and *faktikisht* (*actually*, *in fact* and *as a matter of fact*), which “check on the factuality of the content of the proposition” (Morley 2000: 106), occur more often in ADCN. Obviously, in hard sciences it is expected that arguments are presented on real factual basis, as exemplified in example (15):

(15) Ndryshimet e vrojtuara në klimë, në sistemet fizike dhe në sistemet biologjike për periudhën 1970-2004 janë ato të sipërpërmendura. Dhe *në fakt*, ndryshimet klimatike kanë filluar të japin shenjat negative të ndikimit në rendimentin e grurit. (ADCN_30)
[The observed changes in the climate, in biological and physical systems for the period 1970-2004 are those which were mentioned above. And *in fact*, climatic changes have begun to show negative signs on their influence on the grain efficiency.]

So, the personal engagement is kept to a minimum. This can also be shown by the fact that speech-functional commentary linkers such as *sinqerisht*, *ndershmërisht*, *personalisht* and *për mendimin tim* (*frankly*, *honestly*, *personally* and *in my opinion*) are not used at all in ADCN, but occur in the rest of the sub-corpora in contexts which clearly show that they are employed for persuasive ends, particularly in ADCH.

And finally, there is almost an even frequency distribution of linkers employed to express the degree of validity related to a statement in terms of generalization. They include very common devices such as *përgjithësisht*, *në përgjithësi* and *në tërësi* (*generally*, *in general* and *on the whole*). Interestingly, the statistical data showed that these linkers were the only ones through which discipline differences were less significant (27 % in ADCE, 24 % in ADCH, 24 % in ADCN and in ADCS 25 %). After all, linguistic choices create meaning, “and that is how it enables us to construe our reality and enact our social processes” (Halliday 2013: 35).

4. Conclusive Remarks

I reported some preliminary findings in this short pilot study. I discussed to some extent issues related to the use of structural, textual and commentary linkers in Albanian PhD dissertations in the four major disciplines of economics, humanities, natural sciences and social sciences, which are briefly summarized below.

The analyzed data shows that expansion by elaboration is usually enabled with textual linkers, notably in ADCH where there is a preference for restatement and exemplification. In enhancement, however, hypotactic relations between clauses are very common and structural linkers are indispensable for establishing them. Although there is an almost even distribution in ADCE, ADCH and ADCS, in ADCN there seems to be no necessity to employ many of them, because of the semantic content of statements. Commentary linkers are not very common either, since dissertations are expected to provide objective descriptions of research. There are, however, some interesting distinctions. For instance, hedging is preferred in ADCH, writer's referential stance is particularly frequent in ADCS, factuality appears to characterize dissertations in ADCN, personal engagement is a common phenomenon in ADCH, but not at all in ADCN. Despite these discipline distinctions, we can note that all the disciplines opt for some kind of validity degree related to statements in dissertations by means of common devices such as *përgjithësisht* (*generally*) or *në përgjithësi* (*in general*).

There were basically also two limitations which could have been avoided at the time this paper was written. First, although the reference parts have been removed from each of the dissertations that make up ADC, it was not possible to do so with in-text citations. So, it is obvious that uses of the three types of linkers included in citations are reflected in the statistical figures obtained. It is unlikely, however, that the use frequency of these linkers has "distorted" data to that extent that misinterpretations may have been made. Nevertheless, this claim remains a matter of intuition. Second, it was also not possible to analyze structural and textual linkers in extension across disciplines because of the distinction between uses that enable clause and sentence relations and those between single words (for example, between nouns or adjectives). This requires some manual annotation so as the software can distinguish the type of relations we are interested in. This was not possible, since this task proves to be time-consuming.

Obviously, this study aimed at providing a very general use overview of linkers in some disciplines. Further research can and should be more focused with particular concentration on specific text relations. Also, this comparative investigation can be narrowed down to particular relations in specific fields of study, as for instance, in literature and linguistics, mathematics and biology or finance and economics. It would also be more efficient to follow a more qualitative approach and examine some more significant aspects of use. There are certainly a good number of issues which can be addressed in relation to textual, structural and commentary linkers not only in the disciplines included here, but also in others such as law or medicine, which unfortunately this study failed to investigate.

References

Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman grammar of spoken and written English*. Essex: Pearson Education Limited.

Bloor, T. & M. Bloor (2004). *The functional analysis of English.* London: Edward Arnold.

Dauti, L. (2014). *Çështje të kohesionit gjuhësor në veprën e Bilal Xhaferit* (unpublished doctoral dissertation). University of Tirana, Tirana, Albania.

Dibra, K. & N. Varfi. (1999). *Gjuhësi teksti*. Tiranë: SHBLU.

Eggins, S. (2004). *An introduction to systemic functional linguistics*. London: Continuum.

Halliday, M.A.K. (2013). Meaning as choice. In: L. Fontaine, T. Bartlett & G. O'Grady (eds.), *Systemic Functional Linguistics: Exploring Choice*. Cambridge: Cambridge University Press, pp. 15-36.

Halliday, M.A.K & C. Matthiessen (2004). *An introduction to functional grammar* (3rd ed.). London: Edward Arnold.

Halliday, M.A.K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Baltimore, MD: University Park Press.

Halliday, M.A.K. & R. Hasan (1976). *Cohesion in English*. London: Longman.

Hyland, K. (2005). *Metadiscourse*. London: Continuum.

Hůlková, I. (2011). Conjunctive adverbials in academic written discourse: A corpus analysis based on a sociolinguistic approach. In: J. Schmied (ed.), *Academic rriting in Europe: Empirical perspectives*. Göttingen: Cuvillier, pp. 129-141.

Mëniku, L. (2013). *Konektorët e tekstit në gjuhën shqipe* (unpublished doctoral dissertation). University of Tirana, Tirana, Albania.

Morley, G.D. (2000). *Syntax in functional grammar: An introduction to lexicogrammar in systemic linguistics.* London: Continuum.

Povolná, R. (2012). Causal and contrastive discourse markers in novice academic writing. *Brno Studies in English, 38 (2),* 131-148.

Scott, M. (2014). *WordSmith Tools 6*. Liverpool: Lexical analysis software.

Schmied, J. (2011). Academic writing in Europe: A survey of approaches and problems. In: J. Schmied (ed.), *Academic writing in Europe: Empirical perspectives*. Gottingen: Cuvillier, pp. 1-22.

Topalli, T. (2011). *Gjuhësi teksti*. Tiranë: Gjergj Fishta.

Toska, B. (2012). *Studim përqasës i konektorëve në tekstet argumentues të gjuhës angleze dhe asaj shqipe* (unpublished doctoral dissertation). University of Tirana, Tirana, Albania.

Ventola, E. & A. Mauranen (1991). Non-native writing and native revising of scientific articles. In: E. Ventola (ed.), *Functional and systemic linguistics: Approaches and uses*. Berlin: Mouton de Gruyter, pp. 457-492.

Appendix – The list of linkers analysed in ADC and the English equivalents (partially adapted from *Halliday & Matthiessen 2004*)

expansion	elaboration	sidoqoftë, gjithsesi	anyway, anyhow, in any case
		të paktën	at least
		për shembull	for example, for instance
		me fjalë të tjera, domethënë	in other words, that is (to say)
	enhancement	në, nëse	whether
		kudo (që), ngado (që), tekdo (që)	wherever
		kur	when
		para se, përpara se	before
		pasi, mbasi	after
		që kur, qysh	since
		derisa, gjersa	until, till
		teksa, ndërsa, tek, si	while, whilst, as
		kurdo që, kurdoherë që, sa herë që	whenever, every time, each time
		porsa, posa, sapo (që), qëkurse, teksa	as soon as, immediately, the minute/second/moment/instant
		se, sepse, pasi, meqenëse, meqë	because, since, as, for
		sikur	as if, as though, like
		në, po, nëse, sikur, në qoftë se, po qe se	if
		vetëm në qoftë se	unless
		kështu që	so
		megjithëse, megjithëqë, ndonëse	although, though, even though
		më parë, më përpara, përpara	before, earlier, previously
		ndërkohë, ndërkaq, njëkohësisht	meanwhile, in the meantime, simultaneously
		prandaj, për këtë arsye, pra, kështu (që), në këtë mënyrë, rrjedhimisht	therefore, thus, as a result, so, consequently
		atëherë	then
		si përfundim, përfundimisht, së fundi	finally, lastly
projection	comment: propositional	natyrisht, sigurisht, patjetër	naturally, of course
		pa dyshim, patjetër	undoubtedly, certainly, surely
		çuditërisht, për çudi, papritur, papritmas	surprisingly, unexpectedly
		me sa duket	evidently, supposedly
		ka të ngjarë, ndoshta, ka mundësi, mundet	presumably, probably, possibly, maybe, perhaps
		fatmirësisht, për fat të mirë, fatkeqësisht, për fat të keq	luckily, fortunately, sadly, unfortunately
		me shpresë se, të shpresojmë se	hopefully
		çuditërisht, për çudi	strangely, oddly
		me të drejtë	rightly, correctly
		zyrtarisht	officially
		teorikisht, praktikisht	theoretically, practically
	comment: speech-functional	sinqerisht, ndershmërisht, shkurtimisht	honestly, seriously, briefly
		në të vërtet, në fakt, realisht, faktikisht	actually, in fact, as a matter of fact
		përgjithësisht, në përgjithësi, në tërësi	generally, in general, on the whole
		sinqerisht, ndershmërisht	frankly, honestly, to be honest
		midis nesh	between you and me
		personalisht, për mendimin tim, nga pikëpamja ime	personally, in my opinion, from my point of view

Authorial Identity in PhD Theses in Albania(n)

Armela Panajoti (University of Vlora "Ismail Qemali")

Abstract

That academic writing is not detached from issues of identity, most notable in the authorial presence in any research work, has already been acknowledged. The most conspicuous element of this presence is the use of first person pronouns. Despite this, signaling authorial presence by favouring one form (*I*, *we*, or impersonal forms of addressing) over another has often cultural implications. This study investigates the presence of authorial identity in PhD theses written by Albanian PhD students in Albanian in the period 2012-2014. The main reason for undertaking this study is that Albanian students have for long been encouraged to use either impersonal forms of addressing or the plural *we* to approach their own study. In this study, I aim to look into if and how this tendency has changed in the last years given that more exposition to work from scholars worldwide might have affected the way Albanian students and professors approach research studies. Data will be drawn from a corpus built for this purpose with PhD dissertations from several fields of study. Gender will be another variable in this study to point out expected differences in favouring one linguistic resource over another.

1. Introduction

Identity is an issue that has been given very much attention in various areas of research. It has been widely discussed in literary studies, cultural studies, social sciences, philosophy, discourse studies and so on. The list could be endless. Academic writing, as a particular genre of writing, is not immune to identity. Williams maintains:

> "[…] identity is always present in writing. The idea that any writing can be disconnected from identity is absurd. There is no writing, not a scholarly article, a newspaper editorial, or a technical manual, that does not carry with it an identity of the author. The difference with these forms of writing is that the identity of the author is implicit and assumed by the reader. Most often in academic writing it is an assumed identity of a particular kind." (2006: 712; emphasis in original)

I will draw on this quote to account for my interest in this study to investigate the presence of authorial identity in PhD theses written by Albanian PhD students in Albanian in the last two years. The aim is to see how authorial identity is represented and expressed in these PhD theses and what linguistic resources Albanian PhD students employ to express it and alert its presence as well as to see what meanings are conveyed through them.

The main reason for taking an interest in this issue is that for long, Albanian students have been encouraged to use either impersonal forms of addressing or the plural we to approach their own study. In this study, I aim to look into if and how this tendency has changed in the last two years given that more exposure to work from scholars worldwide might have affected the way Albanian students and professors/supervisors approach research studies.

What is more, unlike research articles, PhD theses are written by novice authors under the supervision of a more experienced or established scholar and as such the process of writing can be seen as two-sided. The writing of a PhD thesis is among other things a long consulting process with a supervisor. In the case of Albanian students, and I suppose of others as well, supervisors will often set the norms of writing, especially when in the absence of a clear-cut stylesheet or reference style PhD students are left with the doubt as to what choices would work best in their writing. In this respect, the supervisor is the main source of orientation for them and his/her suggestions, comments and ideas will often influence the decisions PhD students will make in their work. The supervisor would certainly be an interesting variable to look at, but I have my ethical concerns and technical doubts as to how research involving it can be carried out.

For this analysis, data will be drawn from a corpus built for this purpose with dissertations from several fields of study. Gender will be another variable in this study to point out expected differences in favouring one linguistic resource over another.

My hypotheses are:

(1) Prevalence of *we* over *I* especially in non-humanities.
(2) Gender differences as regards especially the use/favouring of *I*.
(3) Presence of impersonal forms, mostly in non-humanities.

2. Theoretical considerations

That academic writing is not detached from issues of identity, connected primarily with authorial presence in the research work, has already been acknowledged. The interest in authorial identity dates back in the early 90s. Although these studies would not make mention of authorial identity, the relation to the author is one of their most conspicuous characteristics. At the core of these studies have been mainly the roles of first personal pronouns in academic discourse and reader-author interaction. The two have eventually given way to different models or taxonomies for resolving this issue.

Drawing on Halliday & Hasan (1976: 48), Fairclough (1989: 128) and Bakhtin (1986: 301), Vassileva (1998) would be among the first to use quantitative data collected from a corpus of articles in five different languages: English, French, German, Russian and Bulgarian to account for the various qualitative characteristics of favouring I or we within a particular context and to emphasize in her study the cultural component for the choices authors make.

Ivanic's book (1998) is among the first studies to outspokenly relate writer identity with academic writing. By adopting a social constructionist view of identity, she comes up with a tripartite model of self, namely, "autobiographical self" ("the identity which people bring with them to any act of writing, shaped as it is by their prior social and discoursal history" (24),), "discoursal self" ("the writer's 'voice' in the sense of the way they want to sound, rather than in the sense of the stance they are taking" (25)), "self as author," i.e. "a writer's relative authoritativeness"(24) to account for the interaction between language and identity.

By using the concept of “creating identities,” as proposed by the Systemic Functional Linguistics tradition, Tang and John (1999) take up Ivanic’s work further by resorting to her idea of a continuum of ‘I’s and provide an elaborate taxonomy for it: 1. I as representative; 2. I as the guide; 3. I as the architect; 4. I as the recounter of the research processes; 5. I as the opinion holder and 6. I as the originator.

Among the most groundbreaking contributions are Hyland’s who would call the academic writing process “an act of identity” (2002: 1092) and the writing itself a “representation of self” (1091). In all of these studies, the key concept is interaction. In this respect, Hyland (2005a) would create one of the most comprehensive models of interaction—metadiscourse, which is divided into two main categories, interactive and interactional. The model has proven useful to other studies (Hyland 2005b; 2010).

Harwood (2005) adds to this discussion the self-promotional effect of personal pronouns by viewing it in the context of a growing research market.

Following studies would elaborate on Tang and John’s taxonomy by adding or omitting ‘I’s in the continuum. So, Starfield and Ravelli (2006) omit the first role, i.e. Representative, merge those of the Guide and the Architect, and add a new authorial role, the Reflexive I. Whereas, Sheldon (2009) adds narrativity and reflexivity in the role of the Reflexive I. Muñoz (2013) identifies a new role in the continuum of ‘I’s: I as the Interpreter, which she regards as one of the most powerful among the roles and which occurs immediately below I as the Originator on the scale.

All of these studies discuss authorial presence and identity mainly in research articles (Raymond 1993; Vassileva 1998; Kuo 1999; Harwood 2005; Matsuda &Tardy 2007) and few focus on student writing (Tang and John 1999) or MA theses or PhD theses (Hyland 2010; Thompson 2012). As I mentioned earlier, PhD writers as novice writers are not utterly independent in the process of writing and of adopting a style of writing. Most of their choices depend on the advice they receive from their supervisors, the guidelines they follow and the training received in their study courses. Unlike research article authors who want to promote themselves and make themselves visible in the research market, PhD candidates do not necessarily aim towards the same end. For them, getting to the end of their PhD degree is their final goal. Once they are done with it, they can of course think about launching themselves in the research market and get some promotion.

3. The study

3.1. Methodology

The aim of this study is as I outlined earlier to investigate authorial identity in PhD theses written by Albanian students in Albanian. By authorial identity, I will not refer to authorial presence merely, namely to how an author signals his/her presence in the paper, but also to the representation of the author, that is to how he/she projects and manifests his/her identity as an author. For the analysis, I will focus on those linguistic resources authors employ not only to alert their presence, but also to project themselves as authors, to interact with their readers, show confidence and so on.

For this purpose, I have built a small corpus of PhD theses, the Albanian Corpus of Dissertations (ACD), which contains 60 dissertations from several fields of study. These were accessed online on the website of the University of Tirana, Albania. The PDF files were converted into text files for analysis through the commercial software WordSmith Tools 6.0. Since two variables are taken into consideration in this study, gender and research area, several subcorpora were created for this purpose, containing ten texts each, namely ACD-Econ(M) – 10 texts, ACD-Econ(F) – 10 texts, ACD-LL(M) – 10 texts, ACD-LL(F) – 10 texts, ACD-NSc(M) – 10 texts, ACD-NSc(F) – 10 texts[1]. In order to create a well-balanced corpus, dissertations making part of the corpus were randomly selected. These are all available online and part of a larger corpus I am building and trying to enrich for future research. Abstracts, acknowledgements, examples, quotes and references were removed from the texts in the corpus. The table below provides some details about the corpus.

Albanian Corpus of Dissertations (ACD)		
	Texts	**Tokens**
ACD-Econ(M)	10	558,186
ACD-Econ(F)	10	467,064
ACD-LL(M)	10	635,308
ACD-LL(F)	10	567,954
ACD-NSc(M)	10	266,206
ACD-NSc(F)	10	278,168

Table 1: The Albanian Corpus of Dissertations (ACD)

The study investigates, both quantitatively and qualitatively, first personal pronouns (unë - I, ne - we) as well as personal (mendoj - I think, mendojmë - we think, studimi im - my study, punimi im - my paper, studimi ynë - our study, punimi ynë - our paper) and impersonal linguistic resources and structures (ky studim/punim - this study/paper; në këtë studim/punim - in this study/paper) and others in the corpus. Their study is far more complex than in English because Albanian is an inflective language, that is, the doer or agent is included in the inflection of the verb. For this reason, besides I and we, I have also looked into a commonly used verb in research papers mendoj, mendojmë (English - think) to see whether there is a consistent ratio of the use of these two verb forms with

[1] ACD-Econ(M) - Albanian Corpus of Dissertations, Economic Sciences (Male); ACD-Econ(F) - Albanian Corpus of Dissertations, Economic Sciences (Female); ACD-LL(M) - Albanian Corpus of Dissertations, Language and Literature (Male); ACD-LL(F) - Albanian Corpus of Dissertations, Language and Literature (Female); ACD-NSc(M) - Albanian Corpus of Dissertations, Natural Sciences (Male); ACD-NSc(F) - Albanian Corpus of Dissertations, Natural Sciences (Female).

those of I and we. The choice of this verb for analysis is based on its high frequency of use in PhD theses rather than on the meaning(s) it conveys.

Frequency analysis was conducted to provide quantitative data for the interpretation of the occurrences of these linguistic resources in the dissertations. Whereas, the functions of first-person plural pronouns were analysed qualitatively on the basis of the actual occurrences in the sampled texts. For the functions, I have not stuck to any of the taxonomies mentioned above, but have followed a rather loose categorization. The results have been presented in the tables below. In Table 2, findings for the first-person pronouns in the six sub-corpora have been presented. As it is indicated in the footnote, despite the hits received for these pronouns, they do not alert the author's presence in all the cases. The larger number of hits for *ne* (*we*) often results because of a misspelling of the preposition *në* (English- *on*/*in*/*at*) to *ne* in the written form.

	Unë (I)				*Ne (we)*			
Discipline	Male		Female		Male		Female	
	Freq.	%[2]	Freq.	%	Freq.	%	Freq.	%
Economic Sciences	0	0	6	0.06	98[3]	0.36	242[4]	1.04
Language and Literature	44	0.13	42	0.18	370[5]	1.16	572[6]	2.02
Natural Sciences	0	0	0	0	120[7]	1.24	500[8]	3.6

Table 2: Occurrences of *unë* and *ne* in the sub-corpora

In Table 3, findings for the uses of *think* in the first-person, singular and plural, in the six sub-corpora have been presented, whereas Table 4 presents the use of personal versus impersonal forms in the six sub-corpora.

2 The percentage is calculated per 1, 000 words in the subcorpus.
3 Of these, 40 are misspellings of *në*.
4 136 are misspellings of *në*.
5 28 misspellings of *në*.
6 164 misspellings of *në*.
7 34 misspellings of *në*.
8 50 misspellings of *në*.

	Mendoj (I think)				*Mendojmë (We think)*			
Discipline	Male		Female		Male		Female	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Economic Sciences	0	0	28	0.18	26	0.12	8	0.1
Language/ literature	56	0.30	28	0.12	34	0.1	40	0.14
Natural Sciences	6	0.18	34	0.44	0	0	16	0.14

Table 3: Occurrences of *mendoj* and *mendojmë* in the subcorpora

	Punim[9] (paper)				*Punimi im (my paper)*				*Punimi ynë (our paper)*			
Discipline	Male		Female		Male		Female		Male		Female	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Economic Sciences	130	0.46	58	0.32	0	0	0	0	0	0	0	0
Language/ Literature	100	0.32	48	0.18	0	0	0	0	10		6	
Natural Sciences	8	0.1	16	0.14	0	0	0	0	6		0	
	Studim (study)				*Studimi im (my study)*				*Studimi ynë (our study)*			
Economic Sciences	112[10]	0.40	82[11]	0.36	0	0	0	0	2		0	0
Language/ Literature	226[12]	0.72	84[13]	0.3	4		0	0	12		0	0
Natural Sciences	160[14]	1.2	122[15]	0.88	0	0	0	0	44		10	
	Mund të thuhet (It can be said)				*Mund të them (I can say)*				*Mund të themi (We can say)*			
Economic Sciences	0		22		8		0		116		52	
Language/ Literature	54		34		0		0		154		112	
Natural Sciences	0		18		0		0		32		42	

Table 4: Personal vs. impersonal forms in the subcorpora

9 This word appears mainly in collocations like ky punim (English - this paper), në punim/në këtë punim (English - in this paper).
10 The greater number of occurrences for this word is explained by the fact that the word studim (English – study) refers not only to the study in question, but also to studies carried out before by other scholars (50 hits), or means “under study” (12 hits).
11 To refer to other studies (32 hits), to mean “under study” (8 hits).
12 To refer to other studies (44 hits), to mean “under study” (16 hits).
13 Other studies (10 hits), “under study” (8 hits).
14 Other studies (10 hits), “under study” (88 hits).
15 Other studies (32 hits), “under study” (74 hits).

3.2. Interpretation of results

3.2.1. *Quantitative analysis*

As it can be observed from the Tables 2 and 3, the use of *unë* (*we*) is prevalent in all disciplines, with some slight differences in gender in its use in all the three disciplines. The use of *I* is present in PhD theses in Language and Literature, with practically no significant gender differences in use (See Table 2), and it is not encouraged by the Economic Sciences and Natural Sciences. The results in Table 3 display consistency with the results in Table 2. They confirm that the use of the first person is again more dominant in Language and Literature, although we can note some gender differences, with the male candidates favouring the use of *mendoj* (*I think*) more than female candidates, despite the almost equal number of hits obtained for *unë* (*I*). In Economic Sciences, neither *unë* nor *mendoj* were used by the male candidates, whereas female candidates used both although not to a particularly significant degree. Although no occurrences of *unë* were obtained in Natural Sciences for both genders, the use of *mendoj* is notable in both, but what is more interesting to note here is the great number of hits for this word in the theses written by female candidates. The results of *mendojmë* (*we think*) are also consistent with those for *ne* (*we*). It is again the female candidates who favour the use of the first-person plural forms in all disciplines, although this is more significant in Language and Literature and Natural Sciences.

The fourth table shows a clear preference for impersonal forms by the candidates of both genders in all disciplines when approaching one's own study, which should suggest hesitation to "appropriate" one's study. The only observable difference is in the choice of the word *studim* (*study*) or *punim* (*paper/thesis*). Apparently, female candidates in Economic and Natural Sciences display a tendency to use the word *studim* more often. Again, the tendency to prefer the first-person plural forms (*ynë* - *our*; *ne* - *we*) rather than the singular forms (*unë* - *I*; *im* - *my*) is to be noted. What is more significant here is the tendency to use more often the first person *ne* (*we*) instead of the impersonal forms, as in *Mund të themi* (*We can say*). As it can be seen, these are very much on the rise even in Natural Sciences.

3.2.2. *Qualitative analysis*

The qualitative analysis is based on the discussion of some examples taken from the corpus and a table, which identifies the functions of use for all the linguistic resources analysed above.

Qualitative research shows that in the case of PhD theses in Language and Literature written by female candidates, the use of *unë* (*I*) does not always project the author's identity, but most often refers to the author's experience as researcher or teacher especially in the case of theses in the field of language teaching. Below are some examples for comparison:

(1) Vëzhgimet e drejtpërdrejta bëhen kur kolegët marrin përsipër të vëzhgojnë mësimdhënien e njëri-tjetrit dhe më pas vazhdojnë me diskutim konstruktiv përreth asaj

që është vëzhguar ashtu siç praktikova **unë** me koleget e mia, të cilat japin mësim lëndën e dramës në FGJH të UT. **Unë** zgjodha E.V.j dhe P. T. sepse të dyja japin mësim dramën si **unë**, ju pëlqen të diskutojnë pas orëve të mësimit përreth metodave krijuese të prezantimit në orët tona të mësimit dhe shpesh ndajmë eksperiencat tona të lëndës së drames.

[Direct observations are made when colleagues decide to obverse each-other's teaching and then carry on with constructive discussion about what they have observed, which is what **I** did with my colleagues, who teach drama at the Faculty of Foreign Languages, University of Tirana. **I** chose E.V.j and P. T. because both teach drama like **me**, love discussing after class about creative ways of introducing] ACD-LL(F)

(2) Sigurisht **unë** nuk pretendoj të jap një zgjidhje për problemet që ngrihen, por për t'u përqendruar në pika të tilla të rëndësishme që duhet të ndryshojnë.
[Obviously, I do not pretend to provide a solution to the arising problems, but to focus on important issues that demand change.] ACD-LL(F)

(3) **Unë mendoj** se kjo reklamë ngre çështje politike që kanë të bëjnë me ngacmimin seksual në vendin e punës.
[I think this commercial rasies political issues which have to do with sexual harassment in the workplace.] ACD-LL(F)

As it can be noted, in example (1), the *I* is not an authorial *I* since the author of this paper is making clear reference to her research and teaching experience rather than expressing her own voice as author. Whereas in examples (2) and (3), the *I* is an authorial *I* since the author is displaying her own stance.

Within the Language and Literature division, there appears to be more commitment towards an *I*- identity of the author especially in language teaching theses. Nevertheless, the number of these included in the corpus is not enough to sustain such a claim. What can though be noticed is the tendency in common for both genders.

The case of *mendojmë* (English - *we think*) points to another aspect of authorial identity, mainly in female candidates' theses. It does indicate the presence of a plural author in most cases, but in many others it is used for supposition (example (5) below) or to involve the reader in the supposition (examples (4) and (6) below):

(4) **ne** si individë **mendojmë** dhe funksionojmë si të tillë
[…**we** as individuals **think** and function as such…] ACD-LL(F)

(5) Nëse **mendojmë** për natyrën e krijimtarisë në shkolla dhe kulturën që përçohet në arsim, është e rëndësishme të shkojmë përtej steriotipave të mësimdhënies tradicionale dhe të shohim mundësinë e ndërthurrjes së tyre me metodologjitë e reja të mësimdhënies.
[When **we think** about the nature of creativity in schools and the culture taught in education, it is important to move beyond traditional teaching stereotypes and see the possibility to entwine them with the new teaching methodologies.] ACD-LL(F)

(6) **nuk duhet ta mendojmë** globalizimin si zhvillimin më revolucionar të dekadave të fundit, por **duhet të pranojmë** faktin që po **hyjm***ë* në një epokë globale, ndryshe nga modelet e mëparshme.
[**we should not see** globalization as the most revolutionary development of the recent decades, but **we should accept** the fact that unlike the previous models **we are entering** a global epoch.] ACD-LL(F)

The last table, although presenting the data qualitatively reveals differences in the use of first personals. It definitely shows a tendency to prefer the plural over the singular in all disciplines. Only in the ACD-LL (F) and ACD-Econ(F) can we note a tendency for a more pronounced singular *I*. The functions of *unë* and *ne* have been identified and listed following a careful observation of the examples in context in the theses from all the disciplines under study.

	Economic Sciences				**Language and Literature**				**Natural Sciences**			
	Male		Female		Male		Female		Male		Female	
Functions	*Unë (I)*	*Ne (We)*	*Unë (I)*	*Ne (We)*	*Unë (I)*	*Ne (We)*	*Unë (I)*	*Ne (We)*	*Unë (I)*	*Ne (We)*	*Unë (I)*	*Ne (We)*
Misspelling of *në*		✓		✓		✓		✓		✓		✓
Making a point		✓		✓		✓						
Supposing		✓				✓				✓		✓
Presenting data/reporting findings		✓								✓		✓
Concluding			✓	✓				✓		✓		✓
Giving opinion		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Recommending			✓		✓							✓
Agreement			✓			✓						
Hypothesizing			✓	✓								✓
Summarizing					✓							
Reflecting					✓							
Suggesting					✓							
Providing reasons/Arguing					✓	✓		✓	✓	✓	✓	
Involving the reader						✓						
Affirming						✓						
Proposing/outlining							✓	✓		✓		✓
Personal experience							✓		✓		✓	
Approaching one's own study							✓		✓		✓	
Aim										✓		✓

Table 5: The functions of first personal pronouns

4. Concluding remarks

No final conclusions are claimed in this study. However, some general tentative remarks could be made at this stage.

The prevalence of personal linguistic resources not only in humanities but also in non-humanities, especially in Natural Sciences shows a tendency on the part of the authors to project and represent themselves more confidently to their audience. Nevertheless, it is still the plural first personals that dominate, especially in PhD theses written by female candidates in all the three disciplines. However, the presence of impersonal forms to address or approach one's study should suggest hesitation to commitment. Although it was expected the impersonal forms be present in Natural Sciences, it is surprising that they are significantly present in Language and Literature and interesting at the same time that the use of the first person, especially the plural, is on the rise in the Natural Sciences. Seen from a cultural perspective, this should suggest that Albanian PhD candidates are encouraged to mainly use plural first personals and that the candidates themselves feel more confident when they use them.

A more balanced gender ratio is displayed in language and literature in the use of *I*. The difference in the number of hits obtained for both is not so big as to sustain significant gender differences in the choice of *I*.

Despite these remarks, a larger corpus, more data and further research are needed not only to see whether these claims can be fully sustained, but mainly to benefit academic writing itself by highlighting areas for improvement and thus making writers aware of their own writing.

References

Bakhtin, M. (1986). *Speech genres and other late essays.* Austin: University of Texas Press.

Fairclough, N. (1989). *Language and power.* London: Longman.

Halliday, M.A.K. & R. Hasan (1976). *Cohesion in English.* London: Longman.

Harwood, N. (2005). 'Nowhere has anyone attempted . . . In this article I aim to do just that' A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics, 37,* 1207–1231.

Hyland, K. (2010). Metadiscourse: Mapping interactions in academic writing. *Nordic Journal of English Studies, 9 (2),* 125–143.

Hyland, K. (2005a). *Metadiscourse*. London: Continuum.

Hyland, K. (2005b). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, *7 (2),* 173–192.

Hyland, K. (2002). Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics, 34,* 1091–1112.

Ivanic, R. (1998). *Writing and identity: The discoursal construction of identity in academic writing*. Amsterdam: John Benjamins Publishing Company.

Kuo, C. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes, 18 (2),* 121–138.

Matsuda, P. K. & C. M. Tardy. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes, 26,* 235–249.

Muñoz, F. S. (2013), Bildhaftigkeit und Metaphorisierung in der Grammatikvermittlung am Beispiel der Passivkonstruktion. *Zeitschrift für Interkulturellen Fremdsprachenunterricht 18: 1,* 4-20.

Raymond, J. C. (1993). I-Dropping and androgyny: The authoriaI *l* in scholarly writing. *College Composition and Communication, 44 (4),* 478-483.

Sheldon, E. (2009). From one *I* to another: Discursive construction of self-representation in English and Castilian Spanish research articles. *English for Specific Purposes*, *28 (4),* 251-261.

Starfield, S. and Ravelli, L.J. (2006). 'The writing of this thesis was a process that I could not explore with the positivistic detachment of the classical sociologist': Self and structure in new humanities research theses. *Journal of English for Academic Purposes, 5 (3),* 222-243.

Scott, M. (2012). *WordSmith Tools version 6*, Liverpool: Lexical analysis software.

Tang, R. and S. John. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes, 18,* 23-39.

Thompson, P. (2012). Achieving a voice of authority in PhD theses. In: K. Hyland and C. S. Guinda (eds.), *Stance and Voice in Written Academic Genres*. New York: Palgrave Macmillan, pp. 119-133.

Vassileva, I. (1998). Who am I/who are we in academic writing? A contrastive analysis of authorial presence in English, German, French, Russian and Bulgarian. *International Journal of Applied Linguistics, 8 (2),* 163-190.

Williams, B. T. (2006). Pay attention to the man behind the curtain: The importance of identity in academic writing. *Journal of Adolescent and Adult Literacy, 49 (8),* 710–715.

Language Transfer from L2 to L1 in Academic Writing: The Case of Macedonian ESL Teachers

Silvana Neshkovska, Jovanka Lazarevska Stanchevska (Ss Cyril and Methodius University Skopje)

Abstract

Language interference, i.e. transfer from one's native language (L1) to one's second language (L2) and vice versa is unavoidable in the case of bilingual individuals. Although language transfer from L1 to L2 is considered to be the predominant form of transfer and it has received the lion's share of research attention, yet the influence of L2 on L1 is also very important and should not be disregarded or underestimated. This paper focuses on the latter type of language transfer (from L2 to L1) in the context of a specific type of written discourse, i.e. academic writing. Namely, it investigates the borrowing transfer from the English language as a second language which causes Macedonian ESL teachers to make language mistakes while writing their Ph. D thesis in their native language, Macedonian. In that respect, the analysis included three PH.D thesis written in Macedonian by Macedonian ESL teachers and it discloses that, despite their proficient mastery of both languages and the meticulousness with which they approach using both languages, yet their usage of L1 (the Macedonian language), in some respects, is evidently under the influence of their L2 (the English language). More precisely, having grown so accustomed to the rules governing the usage of the English language, for instance, they unconsciously disregard the existing Macedonian terms and replace them with borrowings from English; or they resort to using syntactical structures of the sentences which are typical of the English language but atypical of the Macedonian language.

1. Introduction

Language interference or transfer is a much debated issue in applied linguistics, second language acquisition, and language teaching. Berthold et al. (1997) define language interference as "the transference of elements of one language to another at various levels including phonological, grammatical, lexical and orthographical". Phonological interference comprises items such as stress, rhyme, intonation and speech sounds from the first language influencing the second. Grammatical interference is defined as the first language influencing the second in terms of word order, use of pronouns and determinants, tense and mood. Interference at a lexical level provides for the borrowing of words from one language and converting them to sound more natural in another and orthographic interference includes the spelling of one language altering another (Berthold et al., 1997 in Skiba, 1997).

Weinreich (1953/1968) distinguishes between two types of transfer: borrowing transfer and substratum transfer. Substratum transfer is the type of cross-linguistic influence investigated in most studies of second language acquisition; such transfer involves the influence of a source language (typically the native language of a learner) on the acquisition of a target language, whereas borrowing transfer refers to the influence a second language has on a previously acquired language (which is typically one's native language) (in Odlin 1989: 12).

Another thing which distinguishes between borrowing and substratum transfer is that their results are often very different. Namely, borrowing transfer normally begins at the lexical level, since the attrition of the language absorbing the foreign vocabulary normally begins with the onset of strong cultural influences from speakers of another language. The group exerting the influence, is often, though not always, a speech community with larger numbers, greater prestige and more political power. In such cases, words associated with the government, the legal system, the schools, the technology, and the commercial products of the dominant majority are among the first to make their way into the minority language, but massive lexical borrowing may also supplant much of the vocabulary of everyday living. The effects of substratum transfer, on the other hand, are more evident in pronunciation (and also in syntax) than in the lexicon (Odlin, 1989: 13).

This particular study aims to tackle issues with reference to borrowing transfer. More precisely, it focuses on the borrowing transfer from English as L2 to Macedonian as L1. Although this type of language transfer could be analyzed in a variety of contexts, this study prioritizes written discourse, or more accurately, the academic writing of Macedonian lectors of English. This particular target group has been singled out, primarily, because Macedonian lectors of English are required to write their MA and Ph. D theses in their native language, the Macedonian language, despite the facts that English is the language which they specialize in and that the host of the relevant literature they consult for their research is in English. Hence, not surprisingly, the enormous exposure to the English language inevitably leads them to unconsciously make certain linguistic miscues in their academic writing which are not in accordance with the standard rules of their native language.

This small scale research, in fact, aims to shed some light on the types of miscues which prevail in the academic writing of Macedonian lectors of English and which could, undoubtedly, be attributed to the negative language transfer from their second language, the English language.

The findings of this study, hopefully, will be very beneficial for other Macedonian lectors of English, and students who experience similar problems in their academic writing.

2. The corpus and methodology

The corpus analyzed for this particular study comprises three Ph. D theses1 written by three Macedonian ESP teachers who teach English at university (two

1 Two of the selected Ph. D theses belong to the field of pragmatics, whereas the third one belongs to the field of literary translation.

of them at the Faculty of Philology in Skopje and one of them at the Faculty of Education in Bitola).

More precisely, 50 pages of each of these three PH.D theses, i.e. 150 pages in total, were selected and subjected to analysis in order to ascertain the types of language errors made by Macedonian ESP teachers due to direct influence from the English language. The analyzed texts were the final versions of the Ph. D theses which have been proofread by chartered lectors of the Macedonian language. These texts, in fact, contained the comments and the corrections made by the lectors by means of *Track* Changes and New Comment within the Review Option of Microsoft Word Program.

Having pinpointed and singled out all the errors which were manifestly a result of the language transfer from L2, our subsequent task was to identify the different types of errors and to classify them in distinct groups.

3. Discussion of the results

The thorough analysis of the selected pieces of academic writing has yielded several important insights. Namely, 7 different types of mistakes which could be attributed to borrowing language transfer from the English language were identified in the analyzed linguistic corpus (Table 1):

(1) replacement of Macedonian collocations with English collocations;
(2) omission of the short forms of the dual object;
(3) the usage of redundant prepositions;
(4) omission of prepositions after conjunctions;
(5) usage of wrong prepositions;
(6) wrong word order and
(7) borrowing words (nouns, verbs, conjunctions etc.).

Types of language mistakes						
replacement of Mac. with Eng. collocations	omission of the dual object	redundant prepositions	omission of prepositions after conjunctions	usage of wrong prepositions	wrong word order	borrowing words (nouns, conjunctions etc.)
4	**6**	**12**	**3**	**1**	**3**	**9**

Table 1: Language mistakes made as a result of the borrowing language transfer from L2 to L1 in academic writing

As it is shown in Table 1 the most frequent mistake detected in the analyzed academic writing of the three Macedonian ESL teachers was the redundant usage of prepositions in expressions which do not normally require the presence of those prepositions. For instance, in the Macedonian expression '*izobilstvoto na teorii',* whose English translation equivalent is '*the abundance of theories',* the preposition *'na'* is completely unnecessary and it undoubtedly appears due to language transfer from L2. The same is the case with the expressions: '*koristenje na vozvishen stil'* ('*the usage of an elevated style*') and '*ostvaruvanje na vazni* *komunikaciski celi'* (*'the achievement of important communication goals'*), whose correct versions in Macedonian do not require the presence of the preposition '*na'*.

The second very frequent mistake refers to the direct and unnecessary borrowing of words from English, which greatly overshadows the usage of the corresponding Macedonian terms. In other words, Macedonian ESL teachers in their academic writing in Macedonian very frequently resort to unnecessary borrowing of words from the English language instead of using their Macedonian counterparts. In that respect, the analysis discloses that different types of lexical items (nouns, adjectives, conjunctions etc.) from the English language after having undergone slight pronunciation modifications are introduced into the Macedonian lexical system evicting the corresponding Macedonian terms as a consequence. Thus, for instance, the adjective '*involvirana'* ('*involved'*) in the Macedonian expression '*drugata involvirana strana'* ('*the other party involved'),* is obviously borrowed from the English language. Namely, the adjective '*involved'* has undergone changes to become more compatible with the Macedonian lexical system, and more importantly, has completely overtaken the position of the existing Macedonian equivalent '*vkluchena*. The noun '*interakcija'* ('*interaction'*) is also borrowed from the English language*,* substituting or overshadowing the usage of the existing Macedonian noun '*opshtenje'*. Another instance, which exemplifies the unnecessary introduction of borrowed words from the English language into the Macedonian language, is the conjunction '*so cel da'* ('*in order to'*) in expressions such as '*uslovite shto eden iskaz treba da gi ispolni so cel da bide dobro oformen*' ('*the conditions which an utterance should meet in order to be a well-formed utterance*'). The Macedonian counterpart of the conjunction '*in order to'* in this and similar expressions should be '*za da'*.

Another quite frequent and interesting miscue present in the analyzed academic writing referred to the omission or, more precisely, the complete disregard for the existence of the short forms of the dual indirect objects in the Macedonian language, which does not exist in the English language. Thus, in the Macedonian expression '*bi sakala da se zablagodaram na mojata mentorka*' ('*I would like to thank my mentor'*), Macedonian ESL teachers tend to omit the usage of the short form of the dual indirect object '*i*'('*bi sakala da i se zablagodaram na mojata mentorka*'). Also, in '*Nekoi od nedostatocite bi mozele da se pripishat na faktot…*' ('*Some of the shortcomings could be attributed to the fact….'*), the short form of the indirect object, '*mu',* is missing. The correct version should be '*Nekoi od nedostatocite bi mozele da mu se pripishat na faktot…*'.

The replacement of Macedonian collocations with English collocations, i.e. the usage of literal translation of some English collocations in Macedonian, was also one of the common miscues made as a result of the borrowing transfer of L2. To illustrate this, let's consider the collocation '*pokrenuva prashanja*' ('*to raise issues'*) which in the analyzed corpus replaces the adequate Macedonian counterpart of this collocation '*postavuva prashanja*'. Another example would be the collocation '*cvrsta garancija'* which replaces the corresponding Macedonian collocation '*celosna garancija'* and which is direct translation of the English collocation '*a firm guarantee'*.

Although less frequently, the usage of wrong prepositions was also detected in the analyzed academic writing due to L2 negative language transfer. For example, the correct preposition in Macedonian in the expression '*sprotivna na*'('*opposite of/to'*) should be '*so*'.

Then, the usage of wrong word order which is directly influenced by the English language could be illustrated with the expression '*ironijata ne sekogash ja zabelezuvaat site'* ('*the irony is not always noticed by everybody*'). The correct word order of this sentence in Macedonian should be '*ironijata ne ja zabelezuvaat sekogash site*". More precisely, the position of the adverb '*always'* in English is typically after the form of the verb '*to be'*, so obviously, Macedonian ESL teachers in their academic writings in Macedonian follow the same tendency, regardless of the fact that they end up composing awkward syntactical constructions which are atypical of their native language.

Finally, in the analyzed corpus we came across another linguistic miscue due to the borrowing transfer from English into Macedonian, i.e. the omission of a preposition after conjunctions. For instance, in one of the three pieces of analyzed academic writing the following sentence was found: '*Razlikite ne se rezultat na nemoznosta da se prevede stilot ili subjectivnite karakteristiki na preveduvachot'* ('*The differences are not a result of the inability to translate or the subjective traits of the translator'*). In the Macedonian version of the sentence there should be another preposition '*na*' ('*of*') immediately after the conjunction '*ili*' ('*or*'), but, here, that is not the case due to the influence of the English language which does not require the repetition of the preposition '*of*' after the conjunction '*or*' in such syntactic structures.

4. Conclusion

This paper dealt with the so-called borrowing language transfer occurring in L1 under the influence of L2, which in this case refer to the Macedonian and the English language, respectively.

Although, this type of transfer in scientific literature has not normally been allocated the same amount of attention as substratum language transfer (from L1 to L2), yet this small scale study proves that its prominence should also be taken into consideration very seriously by language researchers.

The proof that we provide in favor of this type of transfer derives from the academic writing of Ph. D candidates whose native language is Macedonian but who at the same time are proficient users of English as a second language. Namely, their academic writing in Macedonian bears witness to the fact that their native language is in some respects affected by their second language.

Namely, even though the number and the variety of the language mistakes detected in their academic writing, is definitely not too large, yet the occurrence of these occasional mistakes verifies the fact that language transfer from L2 to L1 should not be disregarded and should receive an adequate amount of attention.

References

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.

Skiba, P. (1997). Code switching as a countenance of language interference. *The Internet TESL Journal*, *Vol. III, 10,* Retrieved from http://iteslj.org/Articles/Skiba-CodeSwitching

Neshkovska, S. (2014). *Expressing verbal irony in formal and informal speech in Macedonian and English: A contrastive analysis* (Doctoral dissertation). Ss. Cyril and Methodius University, Skopje, Macedonia.

Kitanovska-Kimovska, S. (2013). *Translating style: Based on examples from the English literature in Macedonian* (Doctoral dissertation). Ss. Cyril and Methodius University, Skopje, Macedonia.

Trajkova, Z. (2012). *The pragmatic role of metadiscourse markers in building and achieving persuasion in written and oral discourse in the Macedonian and the English language* (Doctoral dissertation). Ss. Cyril and Methodius University, Skopje, Macedonia.

Appendix

types of mistakes	The first lector of English	The second lector of English	The third lector of English 3
omission of the dual object	1.Би сакала да (ѝ) се заблагодарам на мојата менторка … 2.Имено, уште во втората половина на минатиот век на преводот (му) се посветувало особено внимание … 3. не (им) било во интерес на театарските трупи …. 4. некои од недоследностите може да(му) се припишат на фактот … 5.Според Женет, кој најмногу внимание (ѝ) посветува на хипертекстуалноста …	1.… дека В пред извесно време (им) издал важна деловна тајна на А на конкурентите на А.	
redundant prepositions	1.во еден многу потесен круг на творци 2. зборник на трудови 3. некои од прашањата 4. различни видови на претстави 5. различен број на актери 6. групата на можни корисници на придобивките од овој труд 7. таков вид на погрешна и лажлива јасна комуникација.	1.упатено на некој (некому) што не го исполнил ветувањето 2.мноштво на механизми 3.изобилството на теории 4. намерно кршење на некои од прагматичките принципи 5. упатено на некој (некому) na nekoj што изгледа прекрасно	
omission of prepositions	1.Во таков случај разликите меѓу стилот на нашиот оригинал и стилот на преводот не се резултат на неможноста стилот да се преведе или (на) субјективните карактеристики на преведувачот		
wrong preposition	1.Понекогаш идеологјата на преведувачот не е иста, или дури е и спротивна на (со) идеологијата на оригиналниот автор		

replacing Mac. with Eng. collo-cation	1.Сите прашања кои ги **покренавме** (поставивме) овде ги оставаме отворени.		1.пишувачот ја искажува својата **сигурност** (убеденост) во тоа што го зборува 2.точно тоа што **му е** (го има) **на ум.** 3.**цврста** (целосна) гаранција
word order	1....туку е **секогаш** поврзан (секогаш) со мноштво други текстови ...	иронијата не **секогаш** ја забележуваат (секогаш) сите	има предвид дека не е **секогаш** изнесен (секогаш) доказот во завршниот збор
borrowing words (nouns, conjunctions	**екстралингвистичко** (надворешнојазично)	1 **со цел да** (за да) се нагласи разминувањето 2. **инволвирана** (вклучена) 3.услови коишто еден исказ треба да ги исполни **со цел да** (за да) биде добро оформен 4. **со цел да** (за да) не ја расипат забавата 5. **со цел да** (за да) предизвика емоции кај публиката 6. **екстралингвистичко** (надворешнојазично) 7. **општење** (интеракција)	**интензивира** (засилува)