

Shprehje mbrojtëse ‘hedges’ and *përforcues* ‘boosters’: A corpus analysis of student academic texts in Albanian (L1) and English (L2)

V. Dheskali

Chemnitz University of Technology (Germany);
vincenzo.dheskali@s2015.tu-chemnitz.de

Abstract. To use various realizations of modality (particularly modalization) and intensity and to keep the balance of the right amount of conviction with tentativeness is at the heart of effective writing within the academic context [Halliday, Matthiessen 2014; Hyland 1998a, 1998b]. It is essential for novice writers to know how to negotiate with their readers and express the due amount of probability, usuality and degree of the statements’ truthfulness while discussing established facts, personal opinions or referring to other sources. These can be achieved through the usage of the devices that Holmes [1990] and Hyland [1998b] termed **hedges** and **boosters**. These devices are important elements of well-written academic papers.

Despite the importance of the skills in academic writing, which includes a proper usage of hedges and boosters, very little research has been conducted on academic writing in Albania [Toska 2015]. Thus, initiating research on these phenomena, which I name *shprehje mbrojtëse* ‘hedges’ and *përforcues* ‘boosters’, in this field is crucial.

This paper explores the way Albanian BA, MA and PhD students use hedges and boosters in their academic writings in Albanian (L1) and English (L2). It compares propositional subcategories of hedges (approximators, e.g. APPROXIMATELY) and boosters (intensifiers, e.g. TOTALLY) with authorial hedges (shields, e.g. I THINK) and boosters (emphasizers, e.g. CERTAINLY). The paper borrows concepts from the widely known theory of Systemic Functional Grammar [Halliday 1985; Halliday, Matthiessen 2014] to prove its usefulness in explaining hedges and boosters. The theory is interwoven with Prince et al.’s [1980] categorization of hedges and Quirk et al.’s [1985] model of boosters.

For the purposes of the study, two corpora of writings by Albanian students in Albanian (around 2.2 million words) and in English (around 600,000 words) are compiled. Each corpus contains texts by both male and female writers. Disciplinary domains pertain to both “soft” and “hard” sciences (Social Sciences, Languages and Literature, Medicine, Chemistry, Physics, Mathematics and Informatics). Results of the quantitative analysis show that hedges are significantly favored by Albanians in L1. This indicates, among other interpretations, that they prefer to project a “humble” image about

themselves in their academic writings and conform to their “low” vertical power and institutional roles within the scientific community.

Keywords: corpus linguistics, academic writing, Systemic Functional Grammar, modalization, hedge, booster, Albanian student corpora, Albanian, English, L1, L2.

***Shprehje mbrojtëse* ‘хеджи’ и *përforcues* ‘бустеры’: корпусный анализ студенческих работ на албанском (L1) и английском (L2) языках**

В. Дескали

Хемницкий технический университет (Германия);
vincenzo.dheskali@s2015.tu-chemnitz.de

Аннотация. В настоящей статье изучается использование хеджей и бустеров албанскими студентами бакалавриата, магистрантами и аспирантами в академических работах, написанных на албанском (L1) и английском языках (L2). В исследовании пропозициональные **хеджи** (аппроксимативные выражения, например ПРИБЛИЗИТЕЛЬНО) и **бустеры** (интенсификаторы, например ПОЛНОСТЬЮ) сопоставляются с изменяющими истинностную ценность высказывания хеджами («щиты», англ. shields, например Я ПОЛАГАЮ) и бустерами (усилительные выражения, например НЕСОМНЕННО). Исследование опирается на теорию системной функциональной грамматики [Halliday 1985; Halliday, Matthiessen 2014]. Положения этой теории сочетаются в исследовании с категоризацией хеджей, предложенной в работе [Prince et al. 1980], и моделью использования бустеров согласно [Quirk et al. 1985].

Для целей исследования были созданы два корпуса письменных работ албанских студентов, на албанском (около 2,2 млн словоупотреблений) и английском языках (около 600 000 словоупотреблений). Результаты количественного анализа показывают, что в работах албанских студентов, выполненных на родном языке (L1), предпочтение чаще всего отдается маркерам хеджирования. Одним из возможных объяснений является то, что авторы предпочитают создавать «скромное» представление о себе в своих работах, в соответствии с занимаемым ими «невысоким» положением и незначительной ролью в научном сообществе.

Ключевые слова: корпусная лингвистика, академическое письмо, системная функциональная грамматика, модализация, хедж, бустер, корпуса письменных работ албанских студентов, албанский язык, английский язык, первый язык (L1), второй язык (L2).

1. Introduction

In this paper, I propose a new method of analyzing hedges and boosters that combines the theories and methods developed in the studies by Prince et al. [1980], Quirk et al. [1985], Bondi [2008], Halliday [1985] and Halliday and Matthiessen [2014]. I use the well-known theory of Systemic Functional Grammar (SFG) originated by Michael Halliday (see [Halliday 1985; Halliday, Matthiessen 2014]) to define the higher delicacy levels of modality. By combining Prince et al.'s [1980] theory on hedges and Quirk et al.'s [1985] theory on boosters I aim at proving that they can be jointly used to explain and analyze hedges and boosters. The paper is part of my larger PhD project on hedges and boosters across Albanian and Italian student academic writings in L1 (first language) and L2 (second language: English) [Dheskali n. d.].

The understanding of L2 writing and the relation between L1 and L2 are relevant for my corpus-based comparison of Albanian L1 and L2 (English) novice writers in the academic context. Though there are relevant similarities between L1 and L2 writers and their writings, academic institutions and empirical research demonstrate that there are significant differences as well. Differences between L1 and L2 writings may be related to different linguistic proficiencies, learning experiences and classroom expectations, perception of audience and writer, and preferences for ways of organizing texts [Hyland 2003: 31]. It is important to be aware of these socio-cultural, linguistic and mental differences when comparing L1 and L2 writings as in my paper. In summary, L2 writing is a complex process that involves L1 knowledge, previous L2 writing experiences and many other cognitive, social, cultural, ideological, personal, educational, and contextual factors (see last paragraph of this subsection). That is why it is essential to extend the knowledge on L1 and L2 writing by an underexplored group such as Albanian students.

The discussion on student L1 and L2 writing leads inevitably to the discussion on the status of academic writing in Albania, which remains a largely unexplored area of research, cf. [Toska 2015]. There have been only a very few papers on this topic, such as [Edusei 2015; Toska 2015; Panajoti 2015]. In the Albanian universities, such as the University of Durrës and the University of Vlora, there are no courses offered on academic writing for students of English. These aspects surely impact the way students write their academic theses or term papers.

It is necessary to briefly discuss here the main concepts of this paper, namely **hedges** and **boosters**, and present some reasons why Albanian students should consider them for their writing. These persuasive devices assist writers of academic texts in presenting acceptable claims and fostering agreement from their readers. Hedges (*probably, approximately, might*) are used for the purpose of withholding the authors’ complete commitment to the presented information [Hyland 2017: 20]. The authors use them to distance themselves from their research while maintaining a credible image [Schmied 2018: 9]. According to Prince et al. [1980: 6–20], they either explicitly indicate authors’ partial commitment towards a proposition (by acting as shields, e.g. *may, according to*) or partial membership of any entity or being into a semantic category (by acting as approximators, e.g. *circa, more or less*). Boosters (*totally, definitely, it is clear that*) are used to emphasize strength or an author’s sureness regarding the given information [Hyland 2017: 20]. They indicate either the writer’s full commitment towards the truthfulness of the proposition (“author-related emphazizers”, in Quirk et al.’s [1985] terminology), or the complete membership of an item within a semantic category (“proposition-related intensifiers”, as per [Quirk et al. 1985]). The main categorization of boosters, which is used in my study, was introduced by Quirk et al. [1985] and followed by Bondi [2008]. Further classifications of hedges and boosters (see [Prince et al. 1980; Lafuente Millán 2008]) will not be the focus of this paper.

In Albanian, there are no terms referring to hedges and boosters apart from the slightly mentioned term for hedges *pjesëza të dyshimit* ‘particles of doubt’ in the first volume of the “Grammar of the Albanian Language” (Alb. *Gramatika e gjuhës shqipe*) [Agalliu et al. 2002: 316–317], which I do not find accurate or appropriate because of its negative connotation regarding hedges; they do not only show doubt as this term implies, but also strategic modesty, negotiation of knowledge, etc. For hedges in Albanian, I suggest the term *shprehje mbrojtëse* ‘protecting expressions (hedges)’, based on their pragmatic function of protecting the author from possible criticism. It is an equivalent of the German expression “Heckenausdruck” and similar in meaning to “shields”. For boosters, I introduce the term *përforcues* ‘booster’ to refer to items that show an author’s full commitment to the truth of the proposition. I transferred this term to the field of Linguistics after finding it in a general sense of amplifying in Stefanllari’s [1999/2007] English-Albanian Dictionary. While these two terms can be useful for Albanian researchers in the future, in this paper the widely known English versions “hedges” and “boosters” will be used.

The overall structure of the paper takes the form of five sections, including this *Introductory* one. In *Section 2*, there is a focus on the system of modality in Albanian, current Albanian corpora, as well as the system of hedges and boosters and their syntactic positioning. *Section 3* contains detailed information about the compilation of corpora for the study and the paper's research question. *Section 4* provides a quantitative analysis and related discussion. Finally, conclusions and the most interesting findings are drawn in *Section 5*.

2. Theoretical framework

2.1. Modality and its subcategories in the Albanian language

Modality is a system of meaning which is conveyed in various domains of language and which is “centrally concerned with the speaker's attitude towards the factuality or actualization of the situation expressed by the rest of the clause” [Huddleston 2002: 172–173]. According to [Halliday 1985; Halliday, Matthiessen 2014], modality is further divided into modulation, which covers proposals and offers, and modalization, which covers probability and usability. This division is partially applied in my paper. A parallel category to modality is intensity, which has three degrees, low (indicated by *partly*), high (*almost*) and total (*totally*) [Halliday, Matthiessen 2014: 189]. However, this subsection (and paper as a whole) focuses more on modality, particularly modalization, than intensity.

In Albanian, modality is clearly defined and described by Breu [2009], who compares the modalization functions of Albanian modal auxiliaries. Modality and modalization, modulation and degree devices are briefly explained in the two volumes of the well-known Grammar of the Albanian language from the Albanian Academy of Sciences [Agalliu et al. 2002; Çeliku et al. 2002]. In the second volume, hedges and boosters are explained in more detail as words or groups of words that show the degree of the statements' truthfulness. They can present the statement as completely true, certain, necessary, without doubts, possible, supposed, likely, etc. The speaker expresses his or her strong sureness, strong conviction on the complete truthfulness of the facts or his or her doubts, limitations and tentativeness. However, no equivalent term to “hedges” and “boosters” is mentioned. For example, Çeliku et al. explain how through the use of *sigurisht* ‘surely’,

speakers or writers offer a logical assessment of their statements and indicate their truthfulness [Çeliku et al. 2002: 367]. Though Çeliku et al. make no use of the term “booster” for *sigurisht* ‘surely’, they prove that the phenomenon is present in Albanian (e.g., as in English). While I agree with their statement on *sigurisht* ‘surely’, there are many pragmatic aspects of its use, e.g. face-saving, that were not mentioned. Additionally, Çeliku et al. provide useful examples of other hedges and boosters in Albanian. Some of them are studied in my paper, such as the hedge *ndoshta* ‘probably’ and the boosters *patjetër* ‘of course’ and *natyrisht* ‘certainly’ [Çeliku et al. 2002: 368]. Similarly, in the first volume of the same book, Agalliu et al. [2002: 306] mention the examples of *mbase* ‘perhaps, maybe’, *kushedi* ‘who knows’, *ndoshta* and its variant *ndofta* ‘probably’. I analyze only the more common, academic and standard version *ndoshta*.

In the next subsection, the discussion will move from the theoretical concepts to the corpora used for this analysis.

2.2. Corpus Linguistics: Existing corpora of Albanian and Albanian English

My study is grounded in the field of Corpus Linguistics. Corpus Linguistics, according to McEnery and Hardie [2012: i], is “the study of language data on a large scale, computer-aided analysis of very extensive collections of transcribed utterances or written texts.” A corpus is a “systematic compilation of naturally occurring language” [Friginal 2018: 12].

Corpus studies are very rare in the Albanian context, especially for academic writing. This makes it essential to compile corpora of Albanian and Albanian English. The earlier important initiatives for building corpora of Albanian and Albanian English have resulted in a few existing projects, which will be briefly presented below. Firstly, the corpora of Albanian will be presented and secondly, the ones of Albanian English and academic writing will be the main focus.

At the Friedrich-Alexander-University in Erlangen-Nürnberg, Besim Kabashi is developing an Albanian corpus of spoken and written discourse. So far, it has reached approximately 150 million words and mostly contains material from the Albanian press, as well as books on Medicine, Economy, Geography, and History [Kabashi 2016, 2018]. Another large corpus of Albanian is the Albanian National Corpus, which is available online and can be freely used. The corpus was started by the end of 2011 by a group

of researchers from St. Petersburg (led by Alexander Yu. Rusakov) and Moscow and in 2013 consisted of 16,804,058 words [Morozova, Rusakov 2013: 87]. In 2019, a new version of the corpus was launched reaching a larger size of 31.12 million words [ANC]. During scientific discussions with specialists in the Albanian language, Prof. Rami Memushaj and Besim Kabashi, I learned about the existence of another corpus of Albanian created by the researchers from the Department of the Albanian Language at the Beijing Foreign Studies University (personal communication, December 6, 2016).

A corpus of 60 dissertations written in Albanian was compiled by Armela Panajoti [2015]. It is called “The Albanian Corpus of Dissertations (ACD)” and consists of approximately 2,800,000 words. The corpus data were retrieved from the online repository of the University of Tirana. The dissertations belong to the fields of Language and Literature, Natural Sciences, and Economic Sciences [Panajoti 2015: 180–181]. Bledar Toska [2015: 163–165] built the Albanian Dissertation Corpus (ADC) with a total of 9,668,002 words for his paper. ADC contains 160 dissertations which are written in Albanian and pertain to several categories such as Economics, Humanities, Natural Sciences, and Social Sciences (40 dissertations from each category). Finally, Jacinta Edusei [2015: 119ff] gathered another corpus of Albanian student academic writings consisting of circa ten BA theses written by Albanian students in English. It is named the AlCorpus and it includes 71,342 words.

It is important to note that the previously mentioned corpora, together with my own corpora described in 3.1, are the first and the only corpora of Albanian and Albanian English. They are all different from my corpora of Albanian students’ academic writing including different genres and disciplinary domains.

2.3. Classification of hedges and boosters

As the most relevant concepts of my paper, the system of hedges and boosters will receive particular attention in this subsection. *Figure 1* includes the higher delicacy levels of modalization, modulation and degree of intensity as well as the lower delicacy levels of hedges and boosters, with their subcategories and related examples.

The highest levels of modal assessments and deicticity are represented by modality and intensity, which are intermediate points between positive

and negative polarity. Modality includes different levels of modalization and modulation whereas intensity includes different degrees, or levels.

Figure 1 shows booster categories such as emphasizers (high probability) and intensifiers (total degree). The same lexeme can be used as an emphasizer or intensifier, depending on the context. The example (1) from my Albanian English corpus and my own example (2) illustrate this phenomenon.

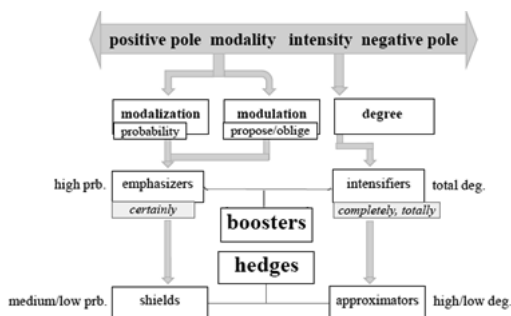


Figure 1. A summary of the key concepts of this article

In the example (1), *definitely* means “without doubt, absolutely”. Through this semantic category, the BA student of Languages and Literature is boosting her commitment towards the truthfulness of the proposition. She appears to be trying to persuade the reader and avoid possible doubts about her claim on *the internet* being *the fastest-growing tool*. As a result, *definitely* functions as an authorial emphasizer. In contrast, the next example (2), includes *definitely* as an intensifier which conveys the meaning of “completely, without exceptions”. If *definitely* is replaced with *completely*, the meaning of the clause complex will remain the same. The same overlaps appear for shields and approximators as it will be shown in the following paragraph.

- (1) *The internet is **definitely** the fastest-growing communication tool [...]*
[CARE13FBL_24]
- (2) *One can separate **definitely** one class from another class.* [author’s example]

Below boosters in Figure 1, there are hedge categories such as shields (medium/low probability) and approximators (high/low degree). Example (3) includes an impersonal usage of the booster *it seems*. Usages such

as *it seems* act as shields, which create a distance between the authors and their texts. The next example (4) shows the approximator *totally* used with a narrow scope and modifying only *trustfully*. The narrow scope is a feature of approximators as well as intensifiers (see example (2)) in contrast to shields and emphasizers, which usually have a larger scope (see examples (1) and (3)). Approximators hedge the propositional level of the clause by conveying partial membership in a certain category (see *Section 1*). They have further subcategories which are not the focus of this study (see [Prince et al. 1980]).

- (3) *Everywhere we look, it seems that technology is constantly at our fingertips.* [CARE12FBL_27]
- (4) *Remaining **totally** trustfully to ones origin and ones memory in relation with smells, tastes and customs as it recalls home, the place you were born and with which one maintains a particular attachment independently from the place one is living.* [CARE12MPL_33]

3. Methodology

3.1. Corpus compilation

In this methodology section, important aspects such as compilation of Albanian and Albanian English corpora of academic writings, formulation of the research question, data preparation and analysis will be elaborated on. This will offer an accurate overview of the different aspects of my study.

For this paper, two corpora of Albanian (CAR: Corpus of Albanian Research) and Albanian English (CARE: Corpus of Albanian Research in English) are used. In the corpora, there is a general presence of all sorts of student academic writing genres (see *Table 1*, p. 63). The different levels start from BA term papers, which are the first substantial academic writing pieces required from students nowadays, to PhD theses, which are written pieces of at least 200 pages that are produced after about eight years of studying at university [Schmied 2015: 12]. A very important aspect of CAR regarding its compatibility is the similar number of texts written by male and female authors.

Table 1. Genres and gender in CAR and CARE including their respective number of texts and word totals (AntConc)

CAR			CARE		
Genre and author (m/f)	# texts	# words	Genre and author (m/f)	# texts	# words
PhD theses	52	2,285,498	PhD theses	9	337,184
			MA theses	10	144,225
			MA term papers	8	22,214
			BA theses	13	110,054
			BA term papers	1	2,620
Males	26	1,108,837	Males	28	485,807
Females	26	1,176,661	Females	13	130,490
Total (2010–2015)	52	2,285,498	Total (2009–2015)	41	616,297

The discussion will move from the different academic writing genres to the disciplinary domains of my corpora (indicated as “disciplin. domain” in *Tables 2* and *3* below). The CAR corpus (*Table 2*) is balanced for several variables. An equal number of texts (8 per disciplinary domain) was included for the domains of Social Studies, Language and Literature, Chemistry, Biotechnology and Physics, Economics, and Mathematics and Informatics. Only Medicine has less texts (only 4). In CAR, Mathematics and Informatics have a considerably smaller number of words than Economy and Physics. Physics (with Biotechnology) as well as Mathematics and Informatics are all considered as “hard” sciences and have generally a considerable focus on numbers. Despite this similarity, Mathematics and Informatics have the lowest number of words from all “hard” sciences. It is interesting to observe how writers of Languages and Literature have used more words to convey their ideas than the ones from Social Studies. It can be seen that females (indicated as f in *Tables 2* and *3*) have used slightly more words than males (indicated as m in *Tables 2* and *3*). However, in Languages and Literature, males have used 349,466 words in total, while females have only used 256,090 words.

The compatibility of CARE with CAR and its internal balance are limited (see *Table 3*). This is also true for the different disciplinary domains, where there is only one PhD thesis in each one of the disciplinary domains of Chemistry, Physics, Mathematics and Informatics. While ranking papers according to different disciplinary domains such as Physics and Chemistry,

I took a record of the specific disciplinary domain or subdomain where they belonged to.

Table 2. Disciplinary domains in CAR including their respective number of papers and word totals per gender type (AntConc)

CAR						
Disciplin. domain	# text	# words	# text m	# words m	# text f	# words f
Lang. & Lit.	8	605,556	4	349,466	4	256,090
Social Stud.	8	483,872	3	185,435	5	298,437
Medicine	4	103,037	2	50,734	2	52,303
Chemistry	8	214,619	4	100,821	4	113,798
Biotechno. & Physics	8	334,607	3	125,470	5	209,137
Economics	8	361,906	4	170,547	4	191,359
Math. & Inf.	8	181,901	6	126,364	2	55,537
Total	52	2,285,498	26	1,108,837	26	1,176,661

Table 3. Disciplinary domains in CARE including their respective number of texts and word totals per gender type according to AntConc

CARE						
Disciplin. domain	# text	# words	# text m	# words m	# texts f	# words f
Lang. & Lit.	33	293,899	20	166,634	13	127,265
Social Stud.	3	127,883	3	127,883	0	0
Chemistry	1	69,513	1	69,513	0	0
Physics	1	14,979	1	14,979	0	0
Economics	2	75,173	2	75,173	0	0
Math. & Inf.	1	34,850	1	34,850	0	0
Total	41	616,297	28	489,032	13	127,265

3.2. Data selection and preparation for analysis

The following paragraphs will concentrate on the general framework of my quantitative comparison as well as the stages of preparing and analysing the data. The stages include selecting the occurrences for the analysis,

copying them to Excel, preparing them for the statistical test and analyzing them according to different variables.

The general framework of my study includes a comparison between the academic writing in L1 (Albanian) and the academic writing in L2 (English). The data from the two corpora will be investigated and compared in different parts of my analysis.

For this paper, I used the concordance program AntConc 3.4.4 [Anthony 2014], which enabled me to search and retrieve the instances of usage of the selected lexemes and their equivalents. After inserting all TXT files (separately for each corpus), I mostly made use of the KeyWord in Context (KWIC) function, which allowed me to see and check all instances.

Since some of the selected lexemes had more than 1,000, 2,000 or even 3,000 occurrences, I used a system that enabled me to analyse large numbers of lexemes and instances. The system that I created consists of specific rules for copying the occurrences from AntConc to the Excel table:

- 1) until 479 instances per lexeme and corpus, all instances are copied and analyzed,
- 2) if the total number of instances per lexeme exceeds 479 instances, only every second instance is copied and analyzed,
- 3) if the total number of instances per lexeme is between 750 and 1,500, only every 3rd instance is copied and analyzed,
- 4) if the total number of instances per lexeme is between 1,501 and 2,000, only every 5th is copied and analyzed,
- 5) if the total number of instances per lexeme is between 2,000 and 2,500, only every 6th of the instances is copied and analyzed,
- 6) if the total number of instances per lexeme is between 2,501 and 3,000, only every 7th instance is copied and analyzed,
- 7) if the total number of instances per lexeme is between 3,001 and 3,500, only every 8th instance is copied and analyzed.

While preparing the data for the analysis, I manually copied every single clause complex (also referred to as sentence). At all, I copied approximately 15,000 occurrences of hedges and boosters from AntConc to Excel.

Two statistical tests were used for the quantitative analysis. The broadest test, the logistic regression, was conducted through the program Rbrul 3.1.1 [Johnson 2009]. Gelman and Hill [2007: 79] refer to the widely used logistic regression as “the standard way to model binary outcomes (that is,

data y_i that take on the values 0 or 1)". I opted for this statistical test due to the binary nature of my dependent variables, hedges or boosters.

The results are represented as log odds values, proportions, and n values. One valid way to interpret logistic regression coefficients is by using odds ratios. The ratio of two odds, $(p_1 / (1-p_1)) / (p_2 / (1-p_2))$, is referred to as the odds ratio [Gelman, Hill 2007: 82]. Using the log odds enables us to present infinite +/- values. The log odds can be easily updated with new data. The n value represents the total number of tokens for hedges and boosters for a specific level of a variable such as hedges and boosters in CARE. The proportion shows the fraction of the total sum that contains a certain attribute. It can be also interpreted in terms of percentages. To prepare the data for the logistic regression through Rbrul, empty cells, non-applicable slots and categories of the included independent variables with insufficient or no data were excluded. They mostly belonged to the CAR corpus. This ensured accurate results from the test.

Another relevant test that I used to measure differences in the usage of specific hedges and boosters across the Albanian L1 and L2 student academic writing corpora is the log-likelihood and effect size calculator [Rayson, Garside 2000]. It provided several values. For this paper, I only present log likelihood (henceforth LL) and odds ratio (henceforth OR) values. The OR effect size shows the statistical impact on this significance. The LL value indicates the significance of the difference. These tests provided valuable results, which are presented in *Section 4*.

My quantitative analysis will concentrate on the following comparisons:

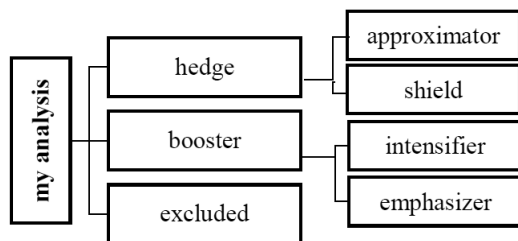


Figure 2. A scheme of my categories for the analysis

My aim is to compare the frequencies of these relevant categories and test their statistical significance. The following subsection will focus on an essential element for scientific research, namely the research question.

3.3. Research question

This subsection introduces the research question, which aims at comparing attitudes of Albanian students found in their academic writings:

What are the differences in the frequency of hedges, boosters and their author commitment categories across the corpora of Albanian student academic writings in L1 (Albanian) and L2 (English)?

The research question focuses on hedges, boosters and their author commitment categories. These categories can be pragmatic or authorial, such as shields and emphasizees, as well as semantic or propositional, such as approximators and emphasizees. To answer this research question, several statistical tests and quantitative comparisons will be presented (see *Section 4*). After presenting the corpora and the research question, I proceed to the discussion of the results of the quantitative analysis.

4. Quantitative analysis

4.1. The distribution of hedges and boosters in CAR

This subsection will focus on the usage of hedges and boosters in CAR, which was measured through the overarching statistical test logistic regression. The widely used logistic regression was used due to the binary nature of my dependent variables, hedges or boosters. I only reveal the results for hedges, which allow inferences on boosters as well. About 15 English lexemes (mostly adverbs) and circa 35 Albanian equivalents were analyzed (see *Appendix*).

The first aspect that will be discussed is corpus. Since several categories, e.g. fronted, afterthought etc., were unevenly distributed and had either zero (empty cells) or very low occurrences for many words, excluding them was the only way to ensure accurate results from the logistic regression through Rbrul [Johnson 2009]. Moreover, any kind of interaction between the categories was excluded from the analysis. These problems were particularly prominent in CARE. The CARE corpus had to be excluded from the further analysis due to the uneven results for categories that could not be used for the statistical test. Hence, the resulting data about hedges (and boosters) in CAR will be compared.

Before discussing the results, it is necessary to explain briefly the data types included in the following *Table 4*. First, the table includes the log odds ($\log [p/(1-p)]$), which is the logarithm of the odds ratio of hedges. Positive values represent the higher occurrence of hedges, negative values — lower occurrence of hedges as compared to boosters. The benefits of using the log odds are that they present infinite +/- values and can be easily updated with new data. Second, the *n* value indicates the total number of tokens for both, hedges and boosters. Third, the proportion shows the fraction of the total sum that contains a certain attribute. It can be also interpreted in terms of percentages (see the details in [Gelman, Hill 2007; Chatterjee, Hadi 2006]).

Table 4. Hedges (and boosters) in CAR

Corpus	Logodds	N	Proportion
CAR	0.412	2445	0.536

The results in *Table 4* clearly show that hedges were favored in CAR. Out of 2445 hedges and boosters in CAR, 53.6% of them were hedges. Moreover, the log odds show a positive value of 0.412. This shows that Albanian students prefer to appear less committed or more approximate in their academic texts in Albanian. The result is statistically significant. While appearing confident does not necessary mean having a confident personality, it does show how authors want to project themselves in their writings. It is one of the most important results of my study.

4.2. The pairwise comparison between CAR and CARE in their occurrences of hedges and boosters

Table 5 shows the results of a brief pairwise comparison between Albanian (CAR) and Albanian English (CARE). In this comparison, there are two opposing sets of the binary results. Various meanings will be compared between the two corpora. In this section, meanings such as PROBABLY and SEEM, that refer to the same semantic set across Albanian (e.g. (a) *ndoshta*, (b) *qartësisht, në mënyrë të / të tregosh qartë*) and English ((a) *probably*, (b) *clearly*) are always written with capitalized letters. The examples or lexemes referring to only one language are written in italics.

While the meanings like PROBABLY, SEEM and CLEARLY show a lower use than expected in CAR in comparison to CARE, the other

meanings show the opposite. These differences are proven to be statistically significant by their log-likelihood (henceforth LL) values, which are generally higher than the 95th percentile or the critical value of 3.84. Furthermore, KIND OF, ENTIRELY, SURELY and CERTAINLY show high odds ratio (henceforth OR) values (more than 1), which establish a strong association between these words in Albanian and English. Out of this collection of words, PROBABLY is the only “outlier” since it has an LL value of only -2.28. This, added to the low size effect expressed by the OR value of 0.82, demonstrates an insignificant difference and association in the use of PROBABLY across the two corpora. Since the negative LL values of CLEARLY and SEEM are very high, it is surprising that their OR results are very low. This might indicate an error of the calculator or simply a low effect of CAR on CARE.

Table 5. Hedges and boosters in CAR and CARE showing their log-likelihood and size effects

Hedge	O1	%1	O2	%2	LL	OR
KIND OF	273	0.01	32	0.01 +	24.70	2.30
PROBABLY	230	0.01	76	0.01 -	2.28	0.82
SEEM	207	0.01	184	0.03 -	128.32	0.30
Booster						
ENTIRELY	185	0.01	18	0.00 +	22.54	2.77
SURELY	194	0.01	8	0.00 +	50.08	6.54
CERTAINLY	103	0.00	17	0.00 +	3.95	1.63
CLEARLY	35	0.00	109	0.02 -	194.75	0.09

There are three meanings, namely SURELY, CERTAINLY and KIND OF, which were used significantly more often in the mother tongue than in English and had relatively high OR values. It is interesting how the English words expressing the two similar meanings CERTAINLY and SURELY were employed much less than their Albanian counterparts. There is no convincing argument on the reasons for the low use in English. In this case, I would attribute it to the writers’ idiosyncratic preference for certain lexemes and rejection to others. This may be conscious in the sense that they purposively opt for a different adverb with a similar meaning. They may have even been at some point exposed to a certain linguistic element and decided to use it without thinking much of its semantic and pragmatic aspects. Indeed, writers preferred to employ CLEARLY as a booster in their texts in English, but not SURELY.

4.3. The distribution of hedges and boosters according to their propositional and authorial categories

The last variable that was analyzed is author's commitment, including author-related usages (indicated as authorial, including emphasizees and shields) and proposition-related usages (indicated as propositional, including intensifiers and approximators). Definitions of these usages and contextualized examples were provided in the *subsection 2.3*. A complete list of analysed propositional and authorial hedges and boosters is provided in *Appendix*. The results are summarized in *Table 6*.

Table 6. The variable of author's commitment influencing the usage of hedges (and boosters)

Author's commitment	Logodds	N	Proportion
Propositional	0.332	4716	0.508
Authorial	-0.332	3372	0.305

On the one hand, hedges were significantly more frequent in the proposition-related category. The higher proportion of the proposition-related category demonstrates that the majority of propositional meanings consisted of hedges. This is also shown by the positive log odds value of 0.332. On the other hand, hedges were less frequent in the author-related level. This can be seen from their low proportion of only 0.305, which shows that only 30.5% of authorial usages were hedges and the rest were boosters. It is also demonstrated by the negative value of -0.332, which is the exact opposite of the odds value of propositional usages which was 0.332. The analysis of author's commitment shows that Albanians prefer to establish a direct connection to their writing while boosting and mostly hedge on the propositional level. Hedging the results of their studies, general quantities and the intensity or usuality of various lexemes is a preferred practice by Albanians. Showing a complete level of commitment and a direct author-proposition relation through boosters is a prominent phenomenon in my corpora.

As a final point about *Table 6*, I would like to mention that the n-values of propositional usages were considerably higher than the authorial ones. This is an outstanding result, considering the presence of many typical authorial usages (OF COURSE, OBVIOUSLY, CERTAINLY), in particular of boosters. However, it needs to be mentioned that the exclusion of the

fronted bi-clausals (initial clauses such as BELIEVE, SUGGEST, SEEM, DEMONSTRATE and PROVE belonging to two clausal relations), which are exclusively authorial usages, influences this result.

Testing the variable of author's commitment revealed that Albanian writers show complete commitment to their writing while boosting (by using author-related boosters) and show lack of commitment (hedge) on the propositional level (by using proposition-related approximators). This is one of the most important findings of this paper. It leads to the conclusions, limitations and suggestions for further research.

5. Conclusions

This quantitative research contributed to two widely unexplored areas of research within the Albanian academic context, namely academic writing and the persuasive devices called hedges and boosters. The research question posed in the paper was: What are the differences in the frequency of hedges, boosters and their author commitment categories across the corpora of Albanian student academic writings? The comparison CAR and CARE demonstrated that PROBABLY, SEEM and CLEARLY had a significantly lower occurrence than expected in CAR in comparison to CARE. Meanwhile, ENTIRELY, SURELY, CERTAINLY and KIND OF showed the opposite. The most striking result is the similarity between CAR and CARE regarding the use of SURELY, CERTAINLY and KIND OF. Both comparisons showed that these three meanings were used significantly more often in the mother tongue than in English. Albanian does not provide an equal form for CERTAINLY, but it offers an equal form for SURELY. Therefore, it is surprising that the English versions of these two items were used significantly less frequent than their Albanian equivalents. The writers' low familiarity with CERTAINLY appeared as a possible factor influencing this difference. However, no convincing argument was found about the difference between CAR and CARE for *surely*. I attribute it to the writers' idiosyncratic choices of certain items and rejections to others. The analysis of author's commitment revealed that Albanians establish a direct connection to their writing while boosting (by using author-related boosters) and mainly hedge on the propositional level (by using proposition-related approximators).

In summary, boosters were more explicit, authorial and initial elements of the clause complex whereas hedges were more implicit, propositional and

located in-between the clause complex. The findings on the usage of proposition-related hedges and author-related boosters, the significantly lower frequency of PROBABLY, SEEM and CLEARLY in CAR as compared to CARE and the higher frequency of ENTIRELY, SURELY, CERTAINLY and KIND OF in CAR in comparison to CARE enabled me to answer the paper's research question.

Finally, a few limitations and suggestions for further research need to be emphasized. Some limitations of the paper were the unbalance and small size of the CARE corpus as well as the limited discussion on dependency, differences in disciplinary domains and student academic writing genres. In any case, compiling and analysing Albanian and Albanian English corpora represents a valuable contribution, which is worth pursuing in further research on hedges and boosters across different disciplinary domains and student academic writing genres.

List of abbreviations

ACD — Albanian Corpus of Dissertations; ADC — Albanian Dissertation Corpus; BA — Bachelor's level (studies); CAR — Corpus of Albanian Research; CARE — Corpus of Albanian Research in English; LL — Log Likelihood; L1 — First Language; L2 — Second Language; MA — Master's level (studies); OR — Odds Ratio; SFG — Systemic Functional Grammar.

References

- Agalliu et al. 2002 — F. Agalliu, E. Angoni, Sh. Demiraj, A. Dhrimo, E. Hysa, E. Lafa, E. Likaj. *Gramatika e gjuhës shqipe* [Grammar of the Albanian Language]. Vol. 1. Tiranë: Akademia e Shkencave e Shqipërisë, 2002.
- ANC — M. Morozova, A. Rusakov, T. Arkhangelskiy. Albanian National Corpus. Available at: <http://albanian.web-corpora.net/> (accessed on 01.12.2019).
- Anthony 2014 — L. Anthony. *AntConc* (Version 3.4.4) [Computer Software]. Waseda University. Tokyo, Japan. Available at: <http://www.laurenceanthony.net/> (accessed on 10.09.2017).
- Bondi 2008 — M. Bondi. Emphatics in academic discourse: Integrating corpus and discourse tools in the study of cross-disciplinary variation. A. Ädel, R. Reppen (eds.). *Corpora and Discourse: The Challenges of Different Settings*. Amsterdam; Philadelphia: John Benjamins Publishing Company, 2008. P. 31–55.
- Breu 2009 — W. Breu. Modals in Albanian. B. Hansen, F. de Haan (eds.). *Modals in the Languages of Europe: A Reference Work*. Berlin: Mouton de Gruyter, 2009. P. 229–266.

- Chatterjee, Hadi 2006 — S. Chatterjee, A. S. Hadi. *Regression Analysis by Example*. 4th edition. New Jersey: John Wiley and Sons, 2006.
- Çeliku et al. 2002 — M. Çeliku, M. Domi, S. Floqi, S. Mansaku, R. Përnaska, S. Prifti, M. Totoni. *Gramatika e gjuhës shqipe [Grammar of the Albanian Language]*. Vol. 2. Tiranë: Akademia e Shkencave e Shqipërisë, 2002.
- Dheshkali n. d. — V. Dheshkali. *A Corpus-based Comparison of Albanian and Italian Student Writings in L1 and English as an L2: Hedges and Boosters as Modalization by Degree*. PhD thesis. Chemnitz: University of Chemnitz, n. d.
- Edusei 2015 — J. Edusei. *Code glosses in student writing: A comparative study of Albanian and German BA theses*. J. Schmied (ed.). *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillier Verlag, 2015. P. 119–131.
- Fraser 2010 — B. Fraser. *Pragmatic competence: The case of hedging*. G. Kaltenböck, W. Mihatsch, S. Schneider (eds.). *New Approaches to Hedging*. (Studies in Pragmatics 9). North America: Emerald Group Publishing Limited, 2010. P. 15–34.
- Friginal 2018 — E. Friginal. *Corpus Linguistics for English Teachers: New Tools, Online Resources, and Classroom Activities*. New York: Routledge, 2018.
- Gelman, Hill 2007 — A. Gelman, J. Hill. *Data Analysis Using Regression and Multilevel / Hierarchical Models*. Cambridge: Cambridge University Press, 2007.
- Halliday 1985 — M. A. K. Halliday. *An Introduction to Functional Grammar*. 1st Edition. London: Edward Arnold, 1985.
- Halliday, Matthiessen 2014 — M. A. K. Halliday, C. M. I. M. Matthiessen. *An Introduction to Functional Grammar*. 4th Edition. London: Routledge, 2014.
- Holmes 1990 — J. Holmes. *Hedges and boosters in women’s and men’s speech*. *Language and Communication*. 1990. Vol. 10. Iss. 3. P. 185–205.
- Huddleston 2002 — R. Huddleston. *Mood and modality*. R. Huddleston, G. K. Pullum (eds.). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 2002. P. 172–208.
- Hyland 1998a — K. Hyland. *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins Publishing Company, 1998.
- Hyland 1998b — K. Hyland. *Boosting, hedging and the negotiation of academic knowledge*. *Text*. 1998. Vol. 18. Iss. 3. P. 349–382.
- Hyland 2003 — K. Hyland. *Second Language Writing*. Cambridge: Cambridge University Press, 2003.
- Hyland 2017 — K. Hyland. *Metadiscourse: What is it and where is it going?* *Journal of Pragmatics*. 2017. Vol. 113. P. 16–29.
- Johnson 2009 — D. E. Johnson. *Getting off the goldvarb standard: Introducing Rbrul for mixed-effects variable rule analysis*. *Language and Linguistics Compass*. 2009. Vol 3. Iss. 1. P. 359–383.
- Johnson 2017 — D. E. Johnson. *Getting off the goldvarb standard: Introducing Rbrul for mixed-effects variable rule analysis*. *Language and Linguistics Compass*. 2017. Vol. 3. Iss. 1. P. 359–383.
- Kabashi 2016 — B. Kabashi. *Building an Albanian text corpus for linguistic research*. Paper presented at the international conference “Corpus-Based Approaches to the

- Balkan Languages and Dialects”, Institute for Linguistic Studies of the Russian Academy of Sciences. St. Petersburg, December 5–7, 2016.
- Kabashi 2018 — B. Kabashi. The neologisms in the AlCo. Paper presented at the international conference “Balkan Languages and Dialects: Corpus-based and Quantitative Studies”, Institute for Linguistic Studies of the Russian Academy of Sciences. St. Petersburg, Russia, October 18–20, 2018.
- Lafuente Millán 2008 — E. Lafuente Millán. Epistemic and approximative meaning revisited: The use of hedges, boosters and approximators when writing research in different disciplines. S. Burgess, P. Martín-Martín (eds.). *English as an Additional Language in Research Publication and Communication*. Bern: Peter Lang, 2008. P. 65–82.
- McEnery, Hardie 2012 — T. McEnery, A. Hardie. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2012.
- Morozova, Rusakov 2013 — M. Morozova, A. Rusakov. Korpusi elektronik i shqipes: përpunimi, përbajtja dhe përdorimi. B. Rugova (kryered.). *Materialet e punimeve të Seminarit XXXII Ndërkombëtar për Gjuhën, Letërsinë dhe Kulturën Shqiptare. Prishtinë, 19–30 gusht 2013* [Proceedings from The International Seminar for Albanian Language, Literature, and Culture XXXII. Prishtina, August 19–30, 2013]. Prishtinë: KOHA Print; Fakulteti i Filologjisë, 2013. F. 85–96.
- Toska 2012 — B. Toska. Studim përqasës i konektorëve në tekstet argumentuese të gjuhës angleze dhe asaj shqipe [A study on connectors in argumentative texts in English and Albanian]. PhD. Tirana: University of Tirana, 2012. Available at: <http://www.doktoratura.unitir.edu.al> (accessed on 03.05.2018).
- Toska 2015 — B. Toska. Structural, textual and commentary linkers in Albanian PhD dissertations: a pilot study. J. Schmied (ed.). *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillier Verlag, 2015. P. 163–177.
- Panajoti 2015 — A. Panajoti. Authorial identity in PhD theses in Albania(n). J. Schmied (ed.). *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillier Verlag, 2015. P. 177–187.
- Prince et al. 1980 — E. F. Prince, J. Frader, C. Bosk. On hedging in physician-physician discourse. Paper presented at the AAAL Symposium on Applied Linguistics in Medicine. University of Pennsylvania, Pennsylvania, December, 1980.
- Quirk et al. 1985 — R. Quirk, R. Greenbaum, G. Leech, J. Svartvik. *A Comprehensive Grammar of the English Language*. London: Longman, 1985.
- Rayson, Garside 2000 — P. Rayson, R. Garside. Comparing corpora using frequency profiling. A. Kilgariff, T. Berber Sardinha (eds.). *Proceedings of the workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000). Hong Kong, October 1–8, 2000*. Hong Kong: Association for Computational Linguistics, 2000. P. 1–6.
- Schmied 2015 — J. Schmied. Graduate academic writing in Europe in comparison: a research-based approach to metalanguage and genre. J. Schmied (ed.). *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillier Verlag, 2015. P. 1–24.

- Schmied 2018 — J. Schmied. A global view on writing research articles for international journals. J. Schmied, M. Hofmann, A. Esimaje (eds.). *Academic Writing for Africa: The Journal Article*. Göttingen: Cuvillier Verlag, 2018. P. 1–18.
- Stefanllari 1999/2007 — I. Stefanllari. English-Albanian Dictionary. Reprint. Kent: Bay Foreign Language Books, 2007. (Original work published in 1999).

Appendix. The list of the hedging and boosting equivalents in Albanian and Albanian English

Note 1. The main principles of selection were high frequency in my corpora, pragmatic similarity, semantic similarity according to dictionaries (COD, LDOCE, and the Oxford Dictionary (for English) and FGJSSH and QEP for Albanian), previous literature e.g. [Fraser 2010: 22ff; Hyland 1998a: 154–177; Prince et al. 1980: 6ff; Toska 2012; Trajkova 2015: 145ff] and suggestions by experts (Memushaj, Kabashi, personal communication, December 6–7, 2016).

Note 2. The lexemes that are in **bold** indicate the lexemes selected on the basis of my perception/competence, while the ones that are not in bold refer to the ones added from previously mentioned literature sources. The words in *italics* are additional equivalents, which I did not include in this study.

Table 7. Hedging epistemic adverbs in Albanian and English

Hedge: epistemic adverb	
English lexeme	Albanian equivalent / translation
approximately, <i>nearly</i> , almost, generally, relatively, roughly, about, around, more or less, circa, loosely	afërsisht, afersisht, për afersisht, perafersisht , <i>në mënyrë të përafërt, me afërsi, pak a shumë, rreth, nja, afër</i>
almost, <i>nearly</i> , very nearly, about, around, approximately, essentially, not far from, in the vicinity of, substantially, near to	thujajse, pothujajse, pothujaj, po thujajse , <i>me përafërsi, gati-gati, afërsisht, gati, rreth</i>
nearly, almost, approximately, roughly, practically, virtually, roundly, in essence/ effect/substance, closely, just about, circa	gati , gati-gati (rare, spoken), <i>shumë afër, pothujajse</i>
kind of, sort of, somewhat, rather, moderately, to some extent	një lloj, disi, njëfarë

Hedge: epistemic adverb	
English lexeme	Albanian equivalent / translation
possibly, <i>conceivably</i> , <i>likely</i> , <i>perhaps</i> , <i>probably</i> , <i>maybe</i> , <i>not impossibly</i>	ka mundësi, ka mundesi , është e mundshme , eshte e mundshme , mundet, mundësisht , mundesisht , ka të ngjarë, ka te ngjare (not found), është e mundur (<i>it is possible</i>), <i>po të jetë e mundur</i>
probably, <i>presumably</i> , <i>seemingly</i> , <i>apparently</i> , <i>perhaps</i> , <i>in all likelihood</i> , <i>in all probability</i>	ndoshta, me/ka shumë mundësi , me/ka shume mundesi , ka shumë të ngjarë , ka shume te ngjare (not found), <i>mbase</i> , <i>ka mundësi</i> , <i>ka të ngjarë</i> , <i>kushedi</i>

Table 8. Boosting epistemic adverbs (emphasizers) in Albanian and English

Booster: epistemic (emphasizing) adverb	
English lexeme	Albanian equivalent
clearly, <i>without any doubt</i> , <i>apparently</i> , <i>certainly</i> , <i>definitely</i> , <i>distinctly</i> , <i>evidently</i> , <i>obviously</i> , <i>audibly</i> , <i>incontestably</i> , <i>incontrovertibly</i>	qartësisht , në mënyrë të/të tregosh qartë , <i>me qartësi</i> <i>qartas</i> , <i>qartazi</i>
of course, <i>as expected</i> , <i>by all means</i> , <i>certainly</i> , <i>definitely</i> , <i>indeed</i> , <i>indubitably</i> , <i>naturally</i>	patjetër, patjeter , <i>medoemos</i> , në mënyrë të detyrueshme/padiskutueshme; <i>me çdo kusht</i> , <i>me siguri</i>
definitely, <i>certainly</i> , <i>absolutely</i> , <i>clearly</i> , <i>decidedly</i> , <i>doubtless</i> , <i>obviously</i> , <i>unequivocally</i> , <i>undeniably</i> , <i>unquestionably</i> , <i>without doubt</i> , <i>surely</i>	absolutisht , në mënyrë absolute, ne menyre absolute (not found), <i>pa asnjë përjashtim</i> , <i>plotësisht</i> , <i>krejt</i>
obviously, <i>unmistakably</i> , <i>apparently</i> , <i>certainly</i> , <i>clearly</i> , <i>definitely</i> , <i>evidently</i> , <i>noticeably</i> , <i>visibly</i> , <i>surely</i>	dukshëm , dukshem , duket qartë , duket qarte (not found), <i>natyrisht</i> , (<i>në mënyrë të</i>) dukshme
surely, <i>without doubt</i> , <i>absolutely</i> , <i>assuredly</i> , <i>certainly</i> , <i>indeed</i> , <i>inevitably</i>	sigurisht , <i>me siguri</i>
certainly, <i>without doubt</i> , <i>absolutely</i> , <i>assuredly</i> , <i>exactly</i> , <i>of course</i> , <i>unquestionably</i> , <i>surely</i>	natyrisht , <i>siç është e natyrshme</i> , <i>sigurisht</i> , <i>pa dyshim</i>