Hedging and Boosting in Albanian, British, and Italian Online News Articles in Europe

Vincenzo Dheskali

Abstract

Modality (and its subcategory modalization) has the function of mitigating statements at some intermediate point between positive polarity (it is, does) and negative polarity (it is not, does not), with various degrees of indeterminacy (cf. Halliday 1985, Halliday and Matthiessen 2014: 144-176). This indeterminacy covers the area of probability and is expressed with items that Holmes (1990) and Hyland (1998) refer to as hedges and boosters. These items are relevant because writers achieve approval by finding the right balance between statements that convey confidence of reliable knowledge (using boosters) and the tentativeness to convey doubt and adequate social interrelations (using hedges) (cf. Hyland 1998b). The aim of this comparative paper is to investigate the usage of hedges and boosters in Albanian, British and Italian online news articles on the refugee crisis. I compare categories and subcategories of author-related and proposition-related hedges (e.g. believe, possibly, approximately) and boosters (e.g. show, certainly, completely) as well as relevant interrelated aspects such as their positioning, orientation, manifestation, and prosody of modalization. My paper will interweave a pragmatic categorization (cf. Prince, Frader, Bosk 1980), a semantic division (cf. Lafuente Millán 2008) and Systemic Functional Theory (cf. Halliday 1985, Halliday and Matthiessen 2014) in an innovative combination. I have compiled three compatible small corpora of Albanian, British and Italian online articles from well-known media sources with genres varying from reports to editorials. The results indicate that hedges and boosters have similar frequencies in the British and Albanian corpora, but that hedges were more prominent (more than double) than boosters in the Italian corpus. I conclude that British and Albanian journalists showed full commitment as well as tentativeness and partial commitment to the proposition in a balanced manner, while Italians were more careful and tentative in their writing.

Keywords: journalistic writing, corpus analysis, modalization, hedge, booster, Albanian online articles, British online articles, Italian online articles, refugee crisis

1. Introduction

Our language is a “system network” and Communication contains choices in relation to language (cf. Halliday and Matthiessen 2014: 23). Language appears in connection to a larger context and discourse and therefore can be considered as a social practice in context (cf. Fairclough and Wodak 1997). My paper seeks to
analyse the language of journalism by considering its context, metalinguistic aspects, and related systemic functional aspects. Specifically, this paper focusses on newspaper discourse, as these might be involved in the creation and reproduction of social inequalities (cf. Richardson 2006: 9).

Journalism fulfils several functions. It is said to work as a means of entertainment, a form of business, an entity that enables “citizens to better understand their lives and their position(s) in the world” (Richardson 2006: 6-7). A single news item often interweaves various ‘voices’ or representations of the discourse from different persons (cf. Fairclough 1995: 79). An analysis of ‘framing’ reveals how surrounding elements of the reporting discourse might have an impact on the way this discourse is perceived by the audience. Framing can be very manipulative (cf. Fairclough 1995: 83). For instance, the journalist might show partial agreement with the reported statements and question their truthfulness by employing hedges, such as ‘the minister suggests/claims’ instead of employing ideational forms, as in the example ‘the minister said’. These and other similar aspects of media discourse can be studied with a Critical Discourse Analysis approach (see Fairclough 1995, Fairclough and Wodak 1997, van Dijk 1993, 2001, Wodak and Chilton 2005 etcetera). Critical Discourse Analysis, also referred to as CDA, deals with the analysis of how political and social inequalities are shown in and recreated through discourse (cf. Wooffit 2005: 137). In my paper, I will borrow from parts of CDA, such as the analysis of modality.

Fairclough (1995: 106) states that ‘if something is explicitly present in a text, it may be informationally foregrounded or informationally backgrounded.’ Journalists might choose to emphasise and minimise different parts of their texts, depending on where they want to lead the audience. ‘The informational structure of clauses’ represented by word positioning (e.g. medial, thematic etcetera) is one of the factors that foreground or background information (cf. Fairclough 1995: 120). I believe that hedges act as means of backgrounding information and boosters of foregrounding information. Hedges (e.g. might, possibly, almost etcetera) are ‘devices which withhold complete commitment to a proposition, allowing information to be presented as an opinion rather than fact’ (Hyland, 1998, Hyland 2011: 199). The opposite of hedges are boosters, which are mostly adverbs or adverbials that provide an increased authority or force to the claims (cf. Bondi 2008: 32). There is a widely accepted need for a more in-depth analysis of the pragmatic and semantic features of boosters (Hyland 2000a, 2000b, Conrad, Biber 2000, Precht 2003, cited in Bondi 2008: 33), especially in Albanian research this area can be further developed. Therefore, both relevant forms of author-commitment will be analysed quantitatively and qualitatively in my paper.

Now, I will briefly introduce the refugee crisis, which has been a central phenomenon and topic treated by the media world-wide in the last months and years. It is a major movement of people towards the European Union, for political, war, economic reasons etc. According to a text published by the European Commission:
‘In the last 2 years, Europe has experienced the greatest mass movement of people since the Second World War. More than 1 million refugees and migrants have arrived in the European Union, the large majority of them fleeing from war and terror in Syria and other troubled countries’ (European Commission 2016).

Without the need to go into a detailed analysis, the main point here is that the refugee crisis had a large impact on Europe, which probably led to political, social, and economical changes. The crisis definitely relates to Albania as one of the Balkan countries involved and to Great Britain and Italy as destination countries (with these countries exercising different degrees of legal and societal acceptance of refugees). Therefore, I chose online articles from media sources of these countries for my analysis, as it is interesting to see how and if their different national approaches to refugees are reflected in their newspaper articles. Similarly to the European Commission, I will use the term refugee in my paper, since it is a well-known term that many of the articles from my corpora used and because I consider the term ‘migrant’ as too generalising and perhaps misleading.

However, I purposively compared the different terms refugee and migrant according to the choices of different news sources between the use of the term refugee and the more general term migrant, especially in the British context. In the articles from my corpora, The Guardian used the term refugee whilst the BBC used the term migrant. In my opinion, the use of the term migrant supports the idea that not all the asylum-seekers are escaping for political or war reasons, but that many of them are leaving their countries for economic reasons. Therefore, restrictive national approaches against these ‘immigrants’, rather than refugees, might come across as fairer to the audience. Meanwhile, the use of the term ‘refugee’ indicates support to these people’s reasons for leaving their home countries. Although further discussions of this terminology go beyond the scope of this paper, it shows how language in media discourse can be used to influence opinions, foreground or background information and possibly, manipulate.

In the following section (Section 2), I will first provide a detailed explanation of the key concepts of my paper, namely polarity, modality and modalization, prosody of modalization, socio-semiotic processes, journalistic writing, hedges and boosters, approximate meanings, positioning in the clause complex, orientation and manifestation. Moreover, I will explain my trinocular perspective on the topic and the way I will interweave three semantic, pragmatic and systemic functional systems into it (Section 3). Secondly, I will illustrate how I compiled the corpora, selected the words for the analysis and I will provide the characteristics of my corpora in Section 4. I will present and explain the quantitative results of my research in Section 5. Findings will be accompanied by contextualised examples of hedges and boosters. Lastly, I will draw conclusions and present constrains and further suggestions in Section 6.
2. Literature Review

2.1. A Trinocular Perspective

As Halliday and Matthiessen (2014) suggest, I will adapt a trinocular perspective on analysing the hedges and boosters of the journalistic articles. I will look at hedges and boosters from above, to see what they mean, which requires including the semantic and contextual level of the analysed items. Moreover, I will look at them from below, to see with which words they are expressed and what the lexical choice of the author is, which for hedges can be endless. Lastly, I will look at them from round about, to see how they co-occur with each other. I will not look at their co-occurrence with other elements but only their co-occurrence with each other e.g. hedge-Booster or even hedge-hedge, Booster-booster. This phenomenon is referred to as ‘prosody of modalization’ (Halliday 1970, 1979, cited in Halliday and Matthiessen 2014: 177) or ‘clustering’ (Hyland 1998a).

2.2. The Systems of Polarity and Modality

Polarity represents a choice between the options yes and no. Within a proposition, we are confronted with the choice to assert (positive polarity: it is, do) or deny (negative polarity: it is not, do not). In English, the finite operators (modals or verbs) can convey positive or negative meanings (cf. Halliday 1985: 85, Halliday and Matthiessen 2014: 101). In between these options (negative or positive polarity) there are intermediate degrees. There are different types of ‘indeterminacy’ labelled collectively as ‘modality’, that act between words such as ‘sometimes’ or ‘maybe’. Modality serves both propositions and proposals, but I am only concerned with the first. Modality is divided into modulation, which includes obligation, permission etc. and modalization, the latter of which covers probability and usuality (cf. Halliday 1985: 86, Halliday and Matthiessen 2014: 176). I will exclusively focus on modalization as it is the only area that covers hedges and boosters. Moreover, I will only concentrate on declarative sentences as they were the main source of hedges and boosters in my corpora. Below, I will explain in detail the part of modality that I focus on, modalization.

2.3. Modalization

Within the proposition, with which we assert (positive polarity: it is) or deny (negative polarity: it is not) something, there are two kinds of intermediate possibilities: (i) degrees of probability: possibly/probably/certainly, that convey ‘either yes or no’ and contain various levels of likelihood; and (ii) degrees of usuality: sometimes/usually/always that express ‘both yes and no’, i.e. sometimes yes, sometimes no, and contain various levels of oftenss (cf. Halliday 1985: 335, Halliday and Matthiessen 2014: 177). In my paper, I will mainly treat probability, as it is the most frequent phenomenon in my corpora. These degrees or ‘modal operators’ represent hedges and boosters (cf. Holmes 1990, Hyland 1998a, 1998b). The
modal operators (within positive polarity) can further fall into three values: low (possibly), medium (probably) and high (certainly) (cf. Halliday 1985: 337, Halliday and Matthiessen 2014: 145). These values will be included in my paper, but not analysed according to this ranking separately. After having briefly introduced the system of modalization, I will explain the items that convey this system: hedges and boosters.

2.4. Socio-Semiotic Processes of Journalistic Writing

Figure 1 below shows a general view of socio-semiotic processes, with the main focus on the categories of sharing and reporting and the subcategory of categorizing into which journalistic writing falls.

Within processes of sharing, we find editorials (opinion pieces), and within reporting, especially chronicling, we find recounts, such as news reports. Within the subcategory of categorizing, we also find reports with a descriptive and taxonomic nature. News articles are both tenor-oriented and field-oriented processes. They are tenor-oriented because they often aim at changing social roles but also are influenced by the institutional roles of the people who write the article, the people who sponsor the news sources and many other roles that might affect the final product. But they are also field-oriented in the sense that they aim at
building field by, for instance, developing knowledge on a specific topic through reports etc. Below, I will explain one of the devices that help to achieve the aims of these tenor-oriented and field-oriented processes, namely hedges.

2.5. Author Commitment: Hedges

Hedges are numerous linguistic devices that protect the reputation of writers and save novice writers from possible criticism. They either indicate the writers’ low level of commitment toward the truthfulness of the given information, or the partial membership of an item to a semantic category. Hedges such as *possibly*, *probably* and *I think* represent ‘devices which withhold complete commitment to a proposition, allowing information to be presented as an opinion rather than fact’ (Hyland 1998, Hyland 2011: 199). In this paper, I investigate the wide phenomena of ‘hedging’ and ‘boosting’, with a specific focus on ‘hedges’ and ‘boosters’. Hyland (1998a: 1-3) separated the terms ‘hedging’ and ‘hedge’ with the former conveying “lack of complete commitment” or “a desire not to express commitment categorically” and the latter as the devices used to express this phenomenon. But the author did not explicitly distinguish between the two terms and seemed to present them as the same (e.g. by explaining hedging through Lakoff’s definition of fuzzy ‘words’). However, I want to point out that hedging is different from hedges. The former represents the wide phenomenon of modalising clauses and clause complexes, expressing e.g. vagueness, which may cover many parts of a clause complex or whole paragraph. The latter stands for lexical and nonlexical items that express various degrees of probability and tentativeness. They could be compared to the difference between weather and climate (cf. Halliday and Matthiessen 2014: 27). Climate (similarly hedging and boosting) is the weather (similarly hedges and boosters) seen from a wider sense of time, which is an internal and external phenomenon expressed in the form of weather. The first to use the term ‘hedge’ was G. Lakoff, who in 1972 referred to it as ‘words whose job it is to make things fuzzy or less fuzzy’ (Lakoff 1973: 471). Even though Lakoff’s semantic approach to hedges has been often criticised by current research (often with a pragmatic focus), it remains a valuable contribution to the concept.

Additionally, Prince, Frader, Bosk (1980: 85) clarify that hedges reflect (at least) two ways of expressing fuzziness, by offering ‘perhaps the most detailed classification of hedges in English to date’, as Crismore and Vande Kopple (1997: 225) declared. Prince, Frader, Bosk divide hedges into ‘approximators’, where the prepositional information is hedged (e.g. *John is some sort of a bad boy*) and ‘shields’, where the author himself/herself is hedged (e.g. *In my opinion, this is a spider*). Moreover, approximators are subdivided into rounders, which convey quantificational approximation of a certain item (e.g. *He is around 50 kg*) and adaptors which modify the lexical meaning of an expression (e.g. *The house is sort of yellow*) (Prince, Frader, Bosk 1980: 9-10). Shields are divided into ‘Attribution Shields’ which means passing the authorship and responsibility to another person (e.g. *Smith claims*) and ‘Plausibility Shields’, where the author directly indicates that
he/she is not fully committed to the truthfulness of the information (e.g. I think this is a hedge) (Prince, Frader, Bosk 1980: 11). Various scholars have contrasted the categorization that I am using by Prince, Frader, Bosk (1980) by stating that it has no significant contribution and that sometimes approximators function in the same pragmatic manner as shields and therefore, differentiating between them in actual texts is very difficult (cf. Skelton 1988, Varttala 1999, 2001). Others such as Lafuente Millán (2008) have partially used the categorization by proposing changes. Hyland (1998a: 4) divided hedges into lexical (e.g. modals: may, might) or non-lexical (comments on limitations of results: [...] it is difficult because of the difficulty of obtaining suitable [...] crystals). In my paper, I only concentrate on lexical hedges, as non-lexical ones were not frequently used in my corpora and are not the purpose of this paper.

2.6. Approximative Meanings

Lafuente Millán (2008: 77) divided approximators into four ‘approximative meaning’ categories: approximate quantity (e.g. roughly, approximately, about, around, circa, almost (when almost precedes a numerical information)), approximate degree (e.g. quite, somewhat, somehow, sort of, kind of, almost (in connection to lexical meanings)), approximate frequency (e.g. occasionally, often, rarely, sometimes, usually, almost (when almost collocates with e.g. always, never, etc.), and approximate limitation (e.g. partly, mostly, predominantly, nearly, almost (when almost appears in connection to finites-e.g. almost cries)). All the underlined approximators represent my additions to Lafuente Millán’s (2008: 77) list of examples. Lafuente Millán referred to hedges only to the items that Prince, Frader, Bosk (1980) called shields and placed approximators in a parallel main category (with a different pragmatic force). He recognised the pragmatic functions of approximators which he described as overlapping and difficult to pinpoint, but mainly focused on the semantic meanings of approximators (cf. Lafuente Millán 2008: 77). Similarly, I see approximators as mainly semantic, and shields as more pragmatic, but I rank them both as subcategories of hedges, since they both have pragmatic functions. With a similar focus on approximators, Mihatsch (2010: 10) draws a comparison between Italian, French, Spanish and Portuguese corpora of spoken language to find notable unidirectional changes leading from ‘lexical’ approximators to ‘quantificational’ ones, displaying various cognitive processes, communicative purposes and distinct origins, all leading the author to employ rounders and adaptors. It is clear that lexical approximators refer to adaptors and approximative meanings of degree and limitation and quantificational ones refer to rounders and approximative meanings of quantity and frequency.

2.7. Author Commitments: Boosters

Hyland (1998b: 236-238) claims that boosters or certainty markers ‘allow writers to project a credible image of authority, decisiveness, and conviction in their views’. Similarly, Holmes (cf. 1982: 18-20) refers to boosters as lexical elements
that every writer can use to convey the idea of self-confidence regarding his or her statements. In the same line, Diani (cf. 2012: 60) presents the examples of certainly, of course, clearly and it is clear, as items that reviewers use to persuade their readers by means of the reviewers’ belief in the logical strength of his/her argument. Within the same context of Italian researchers, Bondi (2008: 32) declares that emphatics (Crismore’s term (1989) for boosters) are mostly adverbs or adverbials which provide an increased authority or force to the claims. She extends her discussion by presenting the widely accepted need for a more in-depth analysis of the pragmatic and semantic features of emphatics (Hyland 2000a, 2000b Conrad, Biber 2000, Precht 2003, cited in Bondi 2008: 33). In my opinion, boosters provide certainty and truthfulness of writer's claims when they are accompanied by necessary evidence. Boosters, together with hedges, are also seen as engagement and stance elements (Lee and Deakin 2016: 22). I agree that boosters and hedges have functions of engaging with the readers. However, in this paper my main aim is to categorise and investigate systemic functional features of hedges and boosters. Researchers often consider the concepts of hedges and boosters as two items that interact and that sometimes are even inseparable (see Grabe and Kaplan, 1997: 155, cited in Vázquez and Giner, 2009: 221). This explains why I analyse both concepts and a adapt comprehensive perspective to see how they cluster together. However, others like Kaltenblöck (2010), by analysing spoken British English, have ranked I think as a booster, as an addition to its other functions, such as that of approximator, shield and structural device. While in my corpora, the alpha clause I think occurred only as a plausibility shield, I do not agree that it can also function as a booster. It would mean that the same item causes two contrastive rhetorical effects, strengthening and weakening the proposition. This is similar to one of the weaknesses of Lakoff’s definition of hedges who claimed that hedges can present and reduce fuzziness at the same time. In journalistic writing, hedges and boosters represent modal choices that convey the judgements, attitudes, or political beliefs of the speakers/writers (cf. Richardson 2006: 62).
2.8. Orientation and Manifestation

The cycle matrix in Figure 2 that I created and adapted to the hierarchy of orientation and manifestation, shows the different orientations and manifestations of modalization according to Halliday (1985: 333-336) and Halliday and Matthiessen (2014: 689).

Halliday (1985) and Halliday and Matthiessen (2014: 162) divided orientation as either subjective or objective and manifestation as either explicit or implicit. The journalists can project the propositional information as a fact and mark subjectivity (e.g. *I believe*) or objectivity (e.g. *it is possible*). Forms such as *It is obvious* and *I think* represent projecting clauses within hypotactic clause complexes. They are clearly separate and primary ‘alpha clauses’ instead of the usual function as modal elements within the clause (e.g. probably). There are other intermediate forms between explicit and implicit realizations. They are conveyed through half-way propositional phrases such as *in my opinion, in all probability* (cf. Halliday 1985: 333). However, these forms were not considered for my study as they are rarely used in my corpora. As we can see in the picture above, these categories are interrelated. It includes both hedges and boosters, as they may fit to all the indicated categories such as subjective-explicit. The authors of these terms only used 1st person forms (like plausibility shields) such as *I think*, as examples of subjective-explicit forms, but I added 3rd person forms (like attribution shields) such as *Smith suggests*, *Smith demonstrates* as they also show a direct link between the proposition and its author (*Smith, in this case*). This link is subjectively ranked (because the author is taking an explicit stance) as subjective from the author who does not fully agree to its content. I will continue with positioning, which is another relevant aspect of the Systemic Functional Theory covered in my paper.
In this section I will briefly describe the positions hedges and boosters (not only) might take in the clause complex. I have used Halliday’s (1985) and Halliday and Matthiessen’s (2014) term ‘clause complex’ to refer to relations of clauses which are also referred to as sentences. Some of the positions derive from the Systemic Functional Theory. As Fairclough states, the ‘element at the beginning of a clause is called its theme’ (Fairclough 1995: 120). The subject is usually marked as theme unless is a good reason for a different choice. (cf. Halliday 1985: 45). One of these good reasons is presenting the whole proposition as modalised and tentative. For instance:

- **Ndoshta nesër Evropa nuk do të jetë më evropiane […]’** (CAA15EXP_3).
- **Perhaps tomorrow Europe will not be European anymore.**

Furthermore, when a word is placed in the final position of a clause, it is called ‘afterthought’ or ‘information focus position’ (Halliday and Matthiessen 2014). The example below contains an afterthought represented by *he* warns.

- **Nëse ky emigracion shumë i rrezikshëm nuk do të ndalohet, Evropa mund të bëhet një Afrikë e re, paralajmëron ai’** (CAA15EXP_3).
- **If this very dangerous migration will not be stopped, Europe might turn into a new Africa, he (former Lybian leader, Muamar Gadafi) warns.**

The journalist might place a word (e.g. a name, adverb etc.) as a ‘theme’ or ‘afterthought’ because it represents information that he or she wants to foreground. With ‘neutral’, I am using Halliday and Matthiessen’s (2014: 187) term to refer to the position directly next to the finite operator which might be a modal or a verb. An example would be:

- **He always wants to win.**

With ‘medial’, I refer to the position in the middle of the sentence but not next to the finite. It might be separated by commas, e.g.

- **[… and, obviously without any reason, they attacked us.**

This is not necessarily the case, however. This term was only fleetingly mentioned by Halliday and Matthiessen, for instance, to refer to the general act of placing an ‘adjunct of intensity’ in the middle of the sentence (Halliday and Matthiessen 2014: 188) or to explain how ‘comment adjuncts’ can follow the element that is prominent (Halliday and Matthiessen 2014: 190). Perhaps they used it as an alternative and more general term for neutral, but there is clearly a difference between neutral and medial. When it is separated by commas or in parenthesis, it counts as an external comment of the author within the sentence, also referred to as a parenthetic form. ‘Parenthetic’ refers to independent proposition that interferes in a sentence by interrupting its syntactical structure. A parenthetic form is usually surrounded by commas, parentheses or dashes.
3. Research Focus: Interweaving Three Systems

As I explained in the section on hedges, I consider hedges and boosters as devices that indicate the author’s complete or partial commitment to the given information. Hedges either indicate the journalist’s low level of commitment toward the truthfulness of the given information, or the partial membership of an item in a semantic category. In the case of boosters, they either indicate the journalist’s full commitment towards the truthfulness of the proposition, by using the category that I will introduce as author-related boosters, or the complete membership of an item in a semantic category, by using the category that I am referring to as proposition-related boosters. I will apply these theoretical concepts with the help of three main systems which I will explain below.

I consider the study by Prince, Frader, Bosk (1980) as a relevant division between the semantic (approximators) and pragmatic (shields) aspects of hedges, and will apply it to my paper. Prince, Frader, Bosk (1980) indirectly mentioned the booster completely within the approximators’ category, hinting at a similarity between proposition-related boosters and approximators. Moreover, I consider proposition-related boosters (e.g. completely, entirely, totally etc.) as devices with a similar semantic (but also pragmatic) function like approximators (e.g. around, almost, partly etc.), even though the former conveys full membership in a semantic category and the later expresses partial membership in a semantic category. As for shields, they show a direct connection between the journalist and its text, just as author-related boosters do. Therefore, I will adapt Prince, Frader, Bosk’s system of hedges to analyse both hedges and boosters and introduce new terms for boosters.

Additionally, for a more detailed division of approximators, I will introduce Lafuente Millán’s categorization even though differently from the researcher, I do not consider approximators as separate from hedges. Approximators are mainly semantic, but they still have interpersonal functions (see Skelton 1988, Lafuente Millán 2008) which are not feasible to explore for the purposes of my paper. Lafuente Millán’s categories will not be introduced to an analysis of proposition-related boosters as it is within the scope of my paper.

In a combination with Prince, Frader, Bosk’s system, Lafuente Millán’s (2008) study and Halliday’s and Halliday and Matthiessen’s modalization aspects, I will try to develop their concepts. Even though their systems are a valuable initial guide to the study of hedges and boosters, there are many details where a researcher can extend on, which I will explain in the methodology part. Moreover, it is challenging to interweave these systems in a systemic, semantic and pragmatic perspective. Until now, there has been no previous study that connects these systems.
4. Methodology

4.1. Research Questions

Four major research questions (see below) have been developed in order to compare attitudes of Albanian, British and Italian journalists found in their online articles on the refugee crisis:

1. What is the frequency and what are the types of hedges and their semantic and pragmatic subcategories used in Albanian, British and Italian (L1) online articles?
2. What is the frequency and what are the types of boosters and their semantic and pragmatic subcategories used in Albanian, British and Italian (L1) online articles?
3. What are the differences in use between hedges, boosters and their semantic subcategories in all the corpora?
4. What choices do journalists from my corpora make regarding Systemic Functional aspects such as positioning, orientation, manifestation and prosody of modalization?

The first two research questions seek to analyse in quantitative and qualitative terms the use of hedges, boosters and related semantic and pragmatic subcategories like approximators in articles from well-known Albanian, British and Italian media sources. The third research question compares the uses of hedges, boosters and their semantic subcategories in all the corpora. The fourth research question combines both hedges and boosters to discover their type of co-occurrence (prosody of modalization in Halliday and Matthiessen’s terms), frequency and type of orientation/manifestation (e.g. if subjective or objective, implicit or explicit) and where journalists position these hedges and boosters to foreground or background information.

I hypothesise that hedges (and related subcategories) will be generally underrepresented in all the corpora and that Italians will use significantly more hedges than Albanians and have a similar amount to British articles. Furthermore, I assume that the data will also show a general low frequency of boosters (and their subcategories) and that Italians will use significantly more boosters than Albanians and a similar amount to the British. These hypotheses derive from my experience as a native speaker of Albanian and a proficient speaker of Italian and my long-term contact with both cultures. Therefore, I posit that Albanians use modalization (and its subcategories) less frequently than Italians and British which I consider a highly modalised language. Moreover, I draw the hypothesis that hedges will be generally more frequently used than boosters and the same will occur for their subcategories. Lastly, I predict that in all the corpora, there will be a hedge-booster type of occurrence of the prosody of modalization and modalization elements will mainly occur in a neutral position in more objective-implicit forms.
4.2. Word (Category) Selection

I selected the hedges and boosters of my study after searching for all the possible (lexical) hedges and boosters and focusing exclusively on the ones that were more frequent. I searched for them with the concordance program AntConc (Anthony 2014). The other aspects that I considered were the lexical and semantic categories that appear as hedges (see Hyland 1998a) and boosters. In this process, I borrowed from Hyland’s (1998) categories, Prince, Frader, Bosk (1980) division of hedges and Halliday’s (1985) and Halliday and Matthiessen’s (2014) types of orientation and manifestation. Therefore, I included epistemic nouns represented by the hedge *possibility* but also by the booster *fact*. Adverbs were represented by, for example, *probably, almost* etc. as hedges and *certainly, surely* etc. as boosters. As it is generally believed by the scientific community, hedges are more frequent variables in word categories and researched than boosters (see Bondi 2008). Boosters mainly appear as adverbs (cf. Bondi 2008: 32). Apart from adverbs, another word class chosen are epistemic verbs such as the cognitive verb *think*, which act as hedges and show bi-clausal realizations (in Halliday’s terms) such as *I think*. I furthermore decided for *show*. These forms are both author-related, according to Prince, Frader, Bosk’s philosophy. I also included modals such as the booster *will* in contrast to *may* and *can*. Additionally, objective-explicit forms such as *it is possible* and *it is clear/obvious* were compared. Lastly, I included an attribution shield (according to) and proposition related boosters such as *completely* in comparison to approximators like *almost*.

4.3. Corpus Compilation

The three small corpora were compiled by collecting online articles on the refugee crisis through the search engine Google. I copied the texts of every selected article into Word files which I then transferred into TXT files (final version for analysis). I cleaned the data by deleting every photo, direct quote, additional comment and anything else which did not represent main body of text to be analysed.

The file names contain an acronym of the corpus of Albanian articles (CAA), corpus of British articles (CBA) or corpus of Italian articles (CIA), year of publication as well as the initial letter of the name of the news source. They do not contain a letter indicating the genre because of a lack of a clear distinction. Hence, a file name could look like following: CIA16CR_3. Here, CIA stands for corpus of Italian articles and it is followed by the year, 2016. The letters CR indicate the news source’s name, Il Corriere Della Sera, and the number 3 differentiates it from other articles.
4.4. Sample

Corpus Linguistics, as McEnery and Hardie (2012: i) state, is ‘the study of language data on a large scale, computer-aided analysis of very extensive collections of transcribed utterances or written texts’. Additionally, they state that a corpus should be representative, and large enough for the analysis. As I will explain below, even though my corpora are not too large in size, they are appropriate for the large number of systemic, pragmatic and semantic aspects that I analyse and my detailed focus on qualitative aspects of my data. The corpora are also balanced in various aspects. The three small corpora were compiled by collecting online articles on the refugee crisis. For this paper, I have compiled an Italian corpus, an Albanian corpus and a British corpus with articles from well-known news sources such as The Independent, BBC and La Repubblica. I researched the refugee crisis in its European, Balkan-related (Macedonian, Albanian, Greek etcetera) Italian and British contexts. All the articles chosen either treated the refugee crisis from a national perspective/context or explained it according to the general European context. Of course, very interesting and diverse perspectives are revealed by comparing, for instance, Albanian online news sources with British online news sources, where the former often revealed a pro-refugees tendency and the later often revealed an against-migrants tendency. I purposively used these different terms according to the preference of many Albanian news sources towards the word refugee and the use from some British news sources of the more general term migrants. In the following tables, I provide a more detailed picture of my small corpora. I have included the name of news sources, their websites, the average length per paper, the number of papers and the number of words. In my paper, I have used the symbol # to refer to the amount (number) of a certain item which is presented in a table.

<table>
<thead>
<tr>
<th>Albanian news sources</th>
<th>News websites</th>
<th>average length</th>
<th>#articles</th>
<th>#words</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlbInfo</td>
<td><a href="http://www.albinfo.ch">www.albinfo.ch</a></td>
<td>465</td>
<td>2</td>
<td>930</td>
</tr>
<tr>
<td>Gazeta Express</td>
<td><a href="http://www.gazetaexpress.com">www.gazetaexpress.com</a></td>
<td>579</td>
<td>5</td>
<td>2896</td>
</tr>
<tr>
<td>Ora News</td>
<td><a href="http://www.oranews.tv">www.oranews.tv</a></td>
<td>357</td>
<td>4</td>
<td>1429</td>
</tr>
<tr>
<td>Telegrafi</td>
<td><a href="http://www.telegrafi.com">www.telegrafi.com</a></td>
<td>726</td>
<td>1</td>
<td>726</td>
</tr>
<tr>
<td>Java News</td>
<td><a href="http://www.javanews.al">www.javanews.al</a></td>
<td>311</td>
<td>2</td>
<td>623</td>
</tr>
<tr>
<td>Dw.com</td>
<td><a href="http://www.dw.com">www.dw.com</a></td>
<td>558</td>
<td>2</td>
<td>1116</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td></td>
<td><strong>483</strong></td>
<td><strong>16</strong></td>
<td><strong>7720</strong></td>
</tr>
</tbody>
</table>

Table 1: Overview of Albanian online news articles and their respective news sources.
As I previously stated in the abstract, the corpora consist of a similar number of words (approximately 10,000 words each), articles (around 13 each), articles per journal (mainly 2 or 3), source of articles (similar proportion of news sources) and years of publication (2015-2017). All the corpora consist of either editorials (also referred to as opinion pieces), or reports and comments, but a clear division was not made for the purposes of my paper. Generally, most papers were taken from the year 2016, but in the Albanian corpus, most papers were from 2015. By observing the tables above, we can easily notice the smaller average length per paper for Albanian articles, which was the smallest in size (311.5 words) of all the articles compared. The average length of this article is around 5 times less than the highest scoring paper/average length of all from Il Post with 1513 words. As we can see, the Italian news sources had the largest number of words (10544) and the Albanian news sources had the lowest (7720). This might relate to differences in the audience preferences. However, it also depends on the genres included. Even though I did not provide a clear division of genres in my paper, maybe the higher presence of editorials (generally considered as short) in one corpus would make this corpus smaller as compared to one that would consist of more news reports.
To have comparable numbers, I used relative frequencies per 10,000 words for all the selected hedges and boosters.

5. Analysis

<table>
<thead>
<tr>
<th>hedge</th>
<th>rel. freq. per 10,000 words</th>
<th>booster</th>
<th>rel. freq. per 10,000 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>mund (can, may, might)</td>
<td>16.83</td>
<td>do të (will)</td>
<td>46.63</td>
</tr>
<tr>
<td>ndoshta (probably)</td>
<td>3.88</td>
<td>sigurisht/me siguri (surely, for sure)</td>
<td>2.59</td>
</tr>
<tr>
<td>mundësi (possibility)</td>
<td>0</td>
<td>fakt* (fact*)</td>
<td>2.59</td>
</tr>
<tr>
<td>mend* (think*)</td>
<td>0</td>
<td>trego* (show*)</td>
<td>2.59</td>
</tr>
<tr>
<td>sipas (according to)</td>
<td>25.90</td>
<td>qart* (clear*)</td>
<td>2.59</td>
</tr>
<tr>
<td>rreth* (around)</td>
<td>5.18</td>
<td>patajëtër (certainly)</td>
<td>0</td>
</tr>
<tr>
<td>thuaj* (almost)</td>
<td>1.29</td>
<td>ndjeshëm (significantly)</td>
<td>0</td>
</tr>
<tr>
<td>më shumë se (more than)</td>
<td>5.18</td>
<td>tërësisht (entirely)</td>
<td>1.29</td>
</tr>
<tr>
<td>plotësisht (completely)</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>ka mundësi (it is possible)</td>
<td>0</td>
<td>është e qartë/dukshme (it is clear/ obvious)</td>
<td>1.29</td>
</tr>
<tr>
<td>total # hedge</td>
<td>58.29</td>
<td>total # booster</td>
<td>63.47</td>
</tr>
</tbody>
</table>

Table 4: Hedges and boosters in the Albanian (CAA) corpus of online news articles per 10,000 words.

The word *mund* was divided according to its three main (English) equivalents, *can* (7.77), *may* (5.18) and *might* (3.89). Of course, only the instances where these words occurred as hedges were counted. Instances of *mund* occurring as *can* were slightly higher than *may* and *might*, and journalists rather used *may* and *can* forms instead of a stronger hedge such as *might*. It was impossible to completely avoid subjectivity, because it is difficult to decide when *mund* referred to e.g. a stronger hedge (*might*), an open possibility (*may*), or a more objective choice (*can*). The main equivalent for *significantly*, *ndjeshëm* had 0 instances, while a rarer form, *në mënyrë të ndjeshme* (*in a significant manner*), was employed with a relative frequency of 3.88. It is interesting that authors indicated the percentage of probability with the form *me 99% siguri* (*with 99% certainty*). As expected, the attributive shield *sipas* (*according to*) showed a significantly higher frequency than all the other hedges and boosters, except from *do të* (*will*). The hedged source was usually a political actor, another media and once, even the words of a person (*according to his word*). The positioning (e.g. neutral, fronted) of *according to* was analysed as for other single-word hedges, but it might need a different categorization, which is not the purpose of this study. From a general perspective, Albanians employed more boosters than hedges, with a slight difference. This result is interesting because hedges are widely believed to be more frequent than boosters.
In the example (1) above, the author has chosen a more complex and probably rare form for significantly, which is në mënyrë të ndjeshme. The same structure and meaning might be found in Italian with in modo significativo. However, this structure is not too frequent in English. The form does not create a prosody of modalization with the temporal will here, since they are not next to each other. As for its position, it is neutral because it is directly followed by the finite reduce.

Cases of do të thotë (it means) were excluded and there were no medial occurrences because Albanian does not allow the use of adverbs between a finite operator (will) and another finite (verb) such as will surely happen, a very low prosody of modalization. Regarding clustering, only once ndoshta (maybe) was used in a cluster with do të (will).

In the following instance (2), do të means would, representing an equivalent that expresses both, low degree (would) and high degree (will). These cases were separated.

2. Sipas Ivanov po qe se vendi i  tëj do të kishte pritur për urdhrat e BE-së [...]. (CAA16DW_8)
   • According to Ivanov, if his country would have waited for orders from EU [...].

In example (3) below, a variation do ta of do të was employed by the author of Gazeta Express, to convey certainty and full commitment to the proposition. However, it is an erroneous use of the booster because one cannot use the booster will to modify a generalization that concerns a whole continent, Europe. In my opinion, will should be used as a booster when there is the necessary evidence to support it, rather than trying to cause panic and be sure about results that are far beyond the perception of a single journalist. I conclude that this statement about the refugee crisis shows biases and erroneous foregrounding.

3. Kriza e ‘refugjatëve’ do ta tjetërsoj dhe shndërroj Evropën në një shoqëri kozmopolite të emigracionit. (CAA15EXP_3)
   • The ’refugee’ crisis will completely change and transform Europe into a cosmopolitan society of migration.

Now, I will focus on the Italian equivalents of this group of hedges and boosters to reveal their (quantitative) similarities and differences to the Albanian and British corpora.
Table 5: Hedges and boosters in the Italian (CIA) corpus of online news articles per 10,000 words.

<table>
<thead>
<tr>
<th>hedge</th>
<th>rel. freq. per 10,000 words</th>
<th>booster</th>
<th>rel. freq. per 10,000 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>potrebbe* (may, might)</td>
<td>10.43</td>
<td>*rà/*ranno (will)</td>
<td>18.01</td>
</tr>
<tr>
<td>può (can)</td>
<td>0.94</td>
<td>certamente (certainly)</td>
<td>0.95</td>
</tr>
<tr>
<td>probabilmente (probably)</td>
<td>4.74</td>
<td>sicuramente/di sicuro (surely/for sure)</td>
<td>1.88</td>
</tr>
<tr>
<td>possibilità (possibility)</td>
<td>0</td>
<td>fatt* (fact*)</td>
<td>7.58</td>
</tr>
<tr>
<td>pens* (think*)</td>
<td>0.94</td>
<td>mostr* (show*)</td>
<td>0</td>
</tr>
<tr>
<td>secondo (according to)</td>
<td>21.81</td>
<td>chiar* (clear*)</td>
<td>2.84</td>
</tr>
<tr>
<td>circa (around)</td>
<td>13.28</td>
<td>sensibil* (significantly)</td>
<td>0</td>
</tr>
<tr>
<td>quasi/appeana (almost)</td>
<td>3.79</td>
<td>interamente (entirely)</td>
<td>0.94</td>
</tr>
<tr>
<td>più di (more than)</td>
<td>10.43</td>
<td>completamente (completely)</td>
<td>0</td>
</tr>
<tr>
<td>è possibile (it is possible)</td>
<td>0.94</td>
<td>è chiaro/ovvio (it is clear/obvious)</td>
<td>0</td>
</tr>
<tr>
<td>total # hedges</td>
<td>70.18</td>
<td>total # boosters</td>
<td>32.25</td>
</tr>
</tbody>
</table>

For will, *rà was used for singular uses of every verb that uses this suffix (of future tense) to boost the given information and *ranno for all the plural forms of will in Italian. Again, we see an obviously high use of modal forms such as rà, ranno (will) and potrebbe (may, might) but a surprisingly low occurrence of può (can), with a relative frequency of only 0.94, since the Albanian equivalent of can had 7.77. It is worth mentioning the high (as usual) occurrence of secondo (according to), circa (around), and più di (more than) on the hedging part and fatt* (fact*) for the boosting part. The contrast between epistemic nouns such as fatt* (fact) with 7.58 and possibilità (possibility) with 0 occurrences, is very striking. Especially considering that the total number of hedges outscored by far (more than double) the number of boosters in the Italian context of journalistic articles. These results indicate that the Italian journalists preferred to convey tentativeness and an approximative meaning with their statements, rather than showing confidence and commitment to their sentences. The higher use of hedges (in the Italian articles) was an expected result. Yet, it did not occur in the other contexts, as we saw from analysing the Albanian articles.

Almost all the numerous instances of can in Italian are conveyed as deontic (able to etc.) and dynamic (obligation etc.) meanings. Apart from that, several hedges within interrogatives were found and excluded as they represent a minority of instances which are not part of my analysis.

4. **Più di 7 mila persone sono intrappolate a Indomeni - un numero quattro volte superiore alle capacità dei due campi di accoglienza locali - […]** (CIA16MSG_6)
   - *More than 7 thousand people are stuck in Indomeni – a number which is four times more than the capacities of the two local camps - […]. (my translation)*

5. **Non più di cento al giorno, per la precisione, […]** (CIA16LS_11)
   - *Not more than one hundred per day, to be precise […] (my translation)*
The first example (4) above shows more than an interestingly fronted, and subsequently thematic position in the clause complex. It is definitely not a usual position for more than or any other approximator including *around, almost, approximately* etc. The second example (5) was excluded from the analysis.

I decided that all the instances including *rà and *ranno* were neutral because of the fact that they are suffixes that are attached to every verb in order to add the future or boosting *will* to it. It is very rare to find a neutral attached to the finite and this was not mentioned in the ‘Introduction to Functional Grammar’ (2014). Below in (6), the Italian journalist is boosting the sentence by using the objective-implicit *surely* in a prosody of modalization from its clustering with *will*. In the Italian part, *surely* is followed by *will not be*, while in its English translation, it naturally requires to be placed in between *will* and *not be*. Lastly, it is a direct negative form as the negative aspect belongs to the proposition and not to the modal operator (see Halliday and Matthiessen 2014: 693).

6. *La commissione ha dato indicazioni che sicuramente non verranno convertite in legge al cento per cento, ma [...] (CIA16LS_13)*

- The commission has given indications that *will surely* not be converted into laws with one hundred percent, but [...].

Now, I will analyse *surely* and other modalization elements in the British corpus.

<table>
<thead>
<tr>
<th>hedge</th>
<th>rel. freq.</th>
<th>booster</th>
<th>rel. freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>may</td>
<td>2.95</td>
<td>will</td>
<td>30.48</td>
</tr>
<tr>
<td>can</td>
<td>2.95</td>
<td>certain*</td>
<td>0.98</td>
</tr>
<tr>
<td>probably</td>
<td>0</td>
<td>surely/for sure</td>
<td>0</td>
</tr>
<tr>
<td>possibility</td>
<td>0</td>
<td>fact*</td>
<td>1.96</td>
</tr>
<tr>
<td>think*</td>
<td>2.95</td>
<td>show*</td>
<td>1.96</td>
</tr>
<tr>
<td>according to</td>
<td>6.88</td>
<td>clear*</td>
<td>0</td>
</tr>
<tr>
<td>around</td>
<td>1.96</td>
<td>significantly</td>
<td>0</td>
</tr>
<tr>
<td>almost</td>
<td>4.91</td>
<td>entirely</td>
<td>0</td>
</tr>
<tr>
<td>more than</td>
<td>15.73</td>
<td>completely</td>
<td>0.98</td>
</tr>
<tr>
<td>it is possible</td>
<td>0</td>
<td>it is clear/obvious</td>
<td>0</td>
</tr>
<tr>
<td>total # hedges</td>
<td>32.45</td>
<td>total # boosters</td>
<td>36.39</td>
</tr>
</tbody>
</table>

Table 6: Hedges and boosters in the British (CBA) corpus of online news articles per 10,000 words.

In general, the British corpus had a low occurrence of hedges and boosters. The obviously low usage of objective explicit forms such as *it is clear that, it is possible* could be noticed in all the corpora. However, in the British corpus, they were totally absent. By comparing epistemic nouns in all the three corpora, we clearly see that the epistemic noun *probability* was never used as a hedge in any of them, while the word *fact*, as a booster, showed various frequencies, e.g. 7.58 in Italian. It is worth emphasizing the large difference between approximators such as *almost* (4.91), *around* (1.96) and *more than* (15.73) and their equivalent forms, proposition-related boosters such as *entirely* (0), *significantly* (0) and *completely* (0.98). This comparison shows a significantly higher occurrence of approximators rather
than the almost absent proposition-related boosters. Differently from Albanians and Italians, the British journalists did not make a high use of the modals may and can as hedges, with both forms only revealing a relative frequency of 2.95. For the highest ranking word, will, quotes interrogatives (as for all the words) and temporals were excluded and there were found a few medial occurrences (2.95) where the finite (verb) and will were divided by an adverb. This is different from Albanian, where no instances occurred in a medial position.

As in the case of Albanian articles, British articles showed a slightly higher quantity of boosters than hedges, revealing that journalists preferred to show commitment to their statements (use boosters) rather than presenting a low degree of confidence in terms of probability and usuality. But as I previously stated, these differences are not significant for the Albanian and British contexts. Perhaps, it is better to say that they employed these forms to a similar degree.

7. In the same year, more than a million migrants applied for asylum - although applying for asylum can be a lengthy procedure so many of those given refugee status may have applied in previous years. (CBA16BBC_1)

8. The UK can, and has a responsibility to, do more to assist refugees along each step of their passage to safer grounds. (CBA15PJ_4)

In example (7), can functions as a hedge to express a subjective-implicit assessment of the author to the probability of its statement. Here, the author had the option to use an unmodalised statement (e.g. is), but chose to modalise the statement regarding it being difficult to generalise on the length of applying for asylum. It also stands in a neutral position. Contrastively, the second clause complex (8) shows a deontic (ability) meaning of can in relation to the UK’s decision to only accept a small number of refugees, rather than expressing probability. This and similar examples were excluded from the analysis. However, this contrast demonstrates how hedges and boosters need a contextual manual analysis in order to draw conclusions because instances of the same word often differ.

<table>
<thead>
<tr>
<th>provenience of articles (corpora)</th>
<th>modalization element</th>
<th>categorization of hedge/booster</th>
<th>rel. freq. per 10.000 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albanian articles</td>
<td>hedge</td>
<td>shield</td>
<td>46.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>approximator</td>
<td>11.66</td>
</tr>
<tr>
<td>British articles</td>
<td>hedge</td>
<td>shield</td>
<td>16.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>approximator</td>
<td>22.62</td>
</tr>
<tr>
<td>Italian articles</td>
<td>hedge</td>
<td>shield</td>
<td>41.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>approximator</td>
<td>27.5</td>
</tr>
<tr>
<td></td>
<td>booster</td>
<td>author-related</td>
<td>35.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposition-related</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>booster</td>
<td>author-related</td>
<td>30.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proposition-related</td>
<td>1.89</td>
</tr>
</tbody>
</table>

Table 7: Categorization of hedges and boosters in the three corpora per 10,000 words.
The journalists from all the three nationalities were more inclined towards creating a direct connection between themselves/source(s) and their texts (using shields or author-related boosters) rather than using forms related to the proposition (using approximators or proposition-related boosters). This difference was significant. The main influence in these higher quantities derives from the high use of modals which I ranked as plausibility shields and attribution shields such as according to. This general tendency might reveal how the journalists from my samples establish a direct link between the text and its source(s)/author(s). They still used a more implicit, detached reporting style in their online articles. This style is proven by their frequent use of modals such as *will, may, can* in all the corpora, which represent subjective-implicit forms. It is also shown by the frequent occurrence of attribution shields, like *according to*, where authors directly or indirectly quoted other news sources, political actors etc., without fully agreeing to their statements. An exception was the case of British journalists who used more approximators (22.62) than shields (16.71) but the difference of frequencies was not significant.

9. *This is one of the reasons that Germany opened its borders completely to asylum seekers as Angela Merkel and other German officials recognize that the influx of refugees from Syria will actually replenish their ageing working population.* (CBA15PJ_4)

The question here is if I would have the choice to leave out *completely* without changing the meaning of this clause complex. I think *completely* is used here to intensify the action and therefore, it works as a proposition-related booster, since the item is adding strength to the proposition represented by Germany, the borders and asylum seekers.

![Figure 3: Categorization of approximative meanings in the three corpora per 10.000 words.](image)
Figure 3 shows that only approximators of degree and quantity were used by the journalists in all the corpora. In my opinion, these two categories are the most frequent categories in reports, editorials and other similar writing. Italian journalists used approximators of quantity the most, with a relative frequency of 26.56, closely followed by British journalists with 20.65 and with circa half of these usages, the Albanian writers with 10.36. The difference between the Albanian journalists and the other two nationalities with regards to approximators of quantity is significant. It seems that the Albanian journalists did not prefer to hedge their quantitative propositions and offer approximative assessments of quantity as much as the Italian and British journalists did. However, the lines and markers (or categories) of this scatterplot are partly influenced by the general frequencies of approximators. Approximators of degree were very low in all the corpora. For example, Italian journalists used them with a relative frequency of 0.94, British journalists 1.97 and Albanians 1.3.

It is necessary to emphasise that these subcategories were not used for boosters, but only for the type of hedges referred to as approximators, in Prince, Frader, Bosk’s terms. I made the decision to create a similar division of author-related and proposition-related boosters, since they function in the same manner as hedges, even though Prince, Frader, Bosk (1980) indirectly mentioned the booster completely within the approximators’ category and one could choose to put proposition-related boosters as approximatorsI considered approximators only as a hedging category and I proposed the term proposition-related boosters for their boosting counterparts.

10. The International Organization for Migration (IOM) estimates that more than 1,011,700 migrants arrived by sea in 2015, and almost 34,900 by land. (CBA16BBC_1)

11. Quasi tutti i motivi che l’anno scorso hanno spinto centinaia di migliaia di siriani, nordafricani e abitanti del Medio Oriente a emigrare sono rimasti irrisolti: […] (CIA16IP_7).1

- Almost all the motives that last year made hundreds and thousands of Syrians, Northern Africans and citizens of the Middle East to migrate have remained unsolved: […]

12. Winter and high seas have not stopped them, and when the weather gets better there are almost certainly going to be even more on the move. (CBA16GR_8)

Lafuente Millán (2008: 77) claimed that the division of approximative meanings into limitation, quantity, frequency and degree, made the variation between them become more significant. This division is a valuable help to distinguish between types of approximators, but the matter seems to be more complex than that. The

---

1 All translations into English in this article are my own.
word *almost* proves my point. In Lafuente Millán’s division, *almost* would probably fall into the category of limitation. However, I will show how *almost* falls into different categories. In (10), *almost* is followed by a number and therefore, it falls into the quantity category. In the next example (11), *almost* still belongs to the category of quantity, with *tutti* (*all*) replacing the number 34.900. But in (12), *almost* is followed by *certainly*, modifying a high degree of probability to indicate that the prediction is nearly, but not completely certain. It is clear that here, *almost* works as an approximator of degree. Perhaps, Lafuente Millán only referred to the meaning of *almost* in itself, as an isolated item regardless of context. But one cannot ignore the surrounding context/proposition of approximators if one is to analyse their meaning.

As I hypothesised, hedges (and their subcategories) were generally underrepresented in all the corpora and showed a higher frequency in Italian articles, as compared to British and Albanian articles. Surprisingly, the difference was not significant with Albanian articles, but rather with British articles. Regarding the subcategories of hedges, the journalists from all my corpora employed significantly more shields than approximators. Exceptionally, the British used slightly more approximators than shields. The type of use was mainly implicit, with modals and attribution shields such as *according to*. Lastly, approximators of degree and quantity were the only forms used consistently by all writers, with a considerably higher amount (around 10 times the size of ‘degree’) of approximators of quantity. Differently from Lafuente Millán’s (2008) categories, *almost* proved to fall into both categories of degree (modifying adverbs) and quantity (modifying numbers). Interestingly, the Albanian equivalent *mund* expressed meanings of *can, may* and *might*. With these statements, an answer to the first research question was provided.

As I predicted, boosters (and related subcategories) had a generally low occurrence in all my corpora. In contrast to my hypothesis, boosters were used with a significantly higher amount (almost double) in the Albanian articles, in comparison to the British and Italian articles, which showed a similar use of these forms. Additionally, in all the corpora, author-related boosters were more frequent than proposition-related boosters, with a significant difference. The Albanian equivalent *do të* expressed both low value (*would*) and high value (*will*). With these statements an answer to the second research question was provided.

I furthermore found out that hedges and boosters (and their subcategories) showed similar frequencies in the British and Albanian corpus, but hedges (and their subcategories) were preferred (more than double) over boosters in the Italian corpus. My research hypothesis was not supported by the results from the British and Albanian corpora, but it was by the Italian corpus. With these statements an answer to the third research question was provided.
Overall, objective-explicit forms are the most highly used category. The Albanian corpus showed the highest use (63.47) of subjective-implicit forms (modals) from all the corpora, which was also the highest category and the highest in the British corpus. The Italian corpus had a significantly higher frequency of objective-implicit forms (54.06) when compared to all the other categories in this corpus. It needs to be pointed out that subjective-explicit forms represented the second highest result (22.76) for this corpus with attribution shields, such as secondo (according to), increasing the frequency of this category, while ‘direct’ plausibility shields such as In my opinion and I think were never used by the journalists. Obviously, objective-explicit forms such as It is possible and it is clear scored the lowest frequencies in all the corpora, except in the Italian one where they had the second lowest score (13.28). The journalists were mainly subjectively oriented (circa 174) rather than objectively oriented (circa 124). From a general overview on manifestation, it seems implicit forms were significantly higher than explicit forms. This brings me to the statement that when hedging or boosting, the Albanian, British and Italian journalists did not prefer to show themselves or draw a direct connection between them as authors and the text. Even though I previously analysed how authors tended to create a direct link between the text and its sources/authors, we saw how this link mainly referred to ‘implicit’ modals or attribution shields (e.g. according to), rather than explicit first person I forms. In fact, they rather used adverbs and modals instead. This might be interrelated to an aimed objective and impersonal writing style/reporting of events.
13. Fino ad oggi gli incidenti legati alla presenza dei rifugiati si contano sulla punta delle dita, ma una presenza più visibile della polizia sarebbe **probabilmente** utile. (CIA15BC_1)

- Until today, the incidents related to the presence of the refugees are very rare in number, but a more visible presence of the police would be **probably** useful.

14. **Probabilmente** è la più grande sfida dell’Unione europea in tutta la sua storia, tanto più grande perché […] (CIA16LS_13)

- **Probably** is the biggest challenge of the European Union in all its history, mostly the biggest because […]

These two examples (13,14) clearly show the interrelation between positioning and manifestation. **Probably** is a plausibility shield in both cases, but it is placed in a neutral position (next to the finite operators *would* and *be*) in the first clause complex and in a fronted (+thematic) position in the second clause complex. This changes the manifestation of the hedges used with the first **probably** being objective-implicit and the second appearing as objective-explicit. Meanwhile, the orientation does not change. This analysis goes against Halliday and Matthiessen’s system where all adverbs are objective-implicit. However, it is possible that the author placed **probably** in the beginning (as fronted) of the clause complex to show a more direct link between him or her as an author and the uncertainty of his or her strong assertion which even included the superlative *the biggest*. The author emphasised that all the following information is hedged, the manifestation shifts from implicit to explicit. It is also interesting that the fronted **probably** also appears next to the finite *is* (as if it was neutral) but this discussion requires a more in-depth analysis of different cases.

![Figure 5: Positioning of hedges and boosters in the clause complex for the three corpora per 10,000 words.](image-url)
Figure 5 shows that in the Albanian corpus there was a significantly high occurrence of hedges and boosters in a neutral position (81.61) compared to all the other positioning categories in all the corpora. As I expected, this category showed the highest presence in all the corpora. Secondly, in the same corpus there was a considerable number of hedges and boosters placed in a thematic (24.61) and fronted (16.84) position. It is interesting that these positions had higher amounts than the medial position, and a significantly higher amount than parenthetic. It is interesting that in the Italian corpus, medial (37.94) is almost as equally high as neutral (38.88), followed by thematic (23.71), fronted (17.07) and parenthetic (12.33). It is clear that the Italian journalists from this sample preferred to hedge or boost in the middle of a clause-complex rather than at the beginning of it. Obviously, the parenthetic position had a very low frequency in all the corpora. It is surprising that the afterthought (at the end of the clause-complex) category exclusively appeared in the British corpus, even though with only a relative frequency of 3.93.

It must be made clear that finite operators (verbs) such as think could not be included into the frequencies of this table. These categories depend on the word’s positioning in relation to verbs, so it does not make sense to include verbs themselves. I did include finite operators such as modals in the analysis of the positioning, since unless they are separated by an adverb, they are always followed by a verb (81.61).

There was a clear difference between Albanian and Italian articles, where according to was mainly used in a fronted, thematic position as well as medial, as compared to the British articles where this hedge was mainly found as an afterthought (at the end of the sentence).

Lastly, there was a hedge-booster type of prosody of modalization. Objective-implicit forms were most frequently used in all the corpora, with a slight difference from subjective implicit forms. Clearly, journalists showed a significantly higher implicit manifestation, rather than an explicit one. In addition, they preferred to place hedges and boosters in a neutral position, which was significantly different than the other positions in the British and Albanian articles. It is surprising that approximators often appeared in a fronted (+thematic) position in the clause complex. In addition to Halliday’s (1985) and Halliday and Matthiessen’s (2014) theory, I found out that adverbial hedges and boosters (e.g. probably) can become objective-explicit rather than their original form, objective-implicit. These statements prove my hypothesis and represent an answer to the fourth research question.

6. Conclusion

In a comparison between online news articles from well-known news sources of Albania, Great Britain and Italy, the use of hedges and boosters showed variation not only in their general usage, but also in their positioning, subcategories, orientation and manifestation. As I hypothesised (research question 1), hedges (and their subcategories) were generally underrepresented in all the corpora and showed a
higher frequency in Italian articles, as compared to British and Albanian Articles. It is surprising that the difference was not significant in Albanian articles, but rather in British articles only. I furthermore found out that the texts in two corpora contained significantly more shields than approximators. Only the British used slightly more approximators than shields. This type of use was mainly implicit, however. Moreover, the journalists employed only approximators of degree and quantity in all the three corpora, with a significantly higher preference for approximators of degree. I assume that this was due to the unclear number of refugee arrivals, which created a need for more approximators of quantity, such as almost, around, and more than. It is interesting that I discovered three main (English) ‘hedging’ equivalents for the Albanian word mund, which were can (7.77), may (5.18) and might (3.89). Differently from Lafuente Millán’s (2008) categories, almost proved to fall into both categories of degree (modifying adverbs) and quantity (modifying numbers).

Regarding the second research question, boosters (and related subcategories) had a generally low occurrence in all my corpora, as I predicted. In contrast to my hypothesis, Albanian journalists employed boosters with a significantly higher amount (almost double) than British and Italian journalists, who both made use of these forms considerably less. In addition, in all the corpora, author-related boosters had more occurrences than proposition-related boosters, with a significant difference between them. A brief qualitative overview revealed that Albanian journalists foregrounded information erroneously, by using boosters to modify overgeneralizing propositions with superlatives. As an exception, the Albanian equivalent do të expressed both low value (would) and high value (will).

Similarly to the Albanian articles, British articles contained a slightly higher frequency of boosters than hedges. However, since the difference is not significant, I conclude that these forms were used in a similar amount. This represents the balanced manner chosen by Albanian and British journalists between expressing tentativeness and partial commitment, as well as showing confidence and full commitment to the propositions made in their online articles. In contrast, the results indicated that the Italian journalists were more likely to express tentativeness and an approximative meaning in their clause-complexes. This statement derives from the fact that the total frequency of hedges (and their subcategories) in Italian out-scored by far (more than double) the frequency of boosters (and their subcategories). My research hypothesis was not supported by the findings from the British and Albanian corpora, but it was supported by the findings from the Italian corpus.

As I mentioned in the previous paragraphs, the Albanian, British and Italian journalists created significantly more direct connections between their statements and their source/author (using shields or author-related boosters), rather than using forms related to the proposition (approximators or proposition-related boosters). This connection was mainly established by using implicit forms like modals or attribution shields like according to, instead of explicit’ first person I forms, however. This enables me to assess that the journalists aimed at an objective and implicit writing style while reporting/analysing events.
Regarding manifestation (the fourth research question), the data showed that implicit forms were more frequent than explicit forms, with a significant difference. Additionally, the journalists were more subjectively-oriented than objectively-oriented. This orientation is interesting since journalism should be more related to objectivity, but the subjective orientation is just one aspect of subjectivity. From a general perspective, the subjective-implicit category was the most frequently occurring category and objective-explicit forms like *It is possible* and *it is clear* showed the lowest frequencies in all the corpora, except Italian. Furthermore, a hedge-booster type of prosody of modalization was discovered. As hypothesised, journalists placed hedges and boosters in a neutral position with a larger range for all the three corpora. This choice was followed by a similar (in frequency) placement in a thematic and medial position. Therefore, I conclude that the journalists preferred to hedge more within the sentence rather than in an initial or ending position of the clause-complex, which weakens the strength of hedges. It is surprising that often approximators appeared in a fronted (+thematic) position in the clause complex. In addition to Halliday’s (1985) and Halliday and Matthiessen’s (2014) theory, it was found out that adverbial hedges and boosters (e.g. *probably*) become objective-explicit rather than their original form, objective-implicit.

There are some constrains that have to be mentioned. While compiling my corpora, the accessibility of data was a problem. Certain online news sources, like the Albanian ‘Panorama’, did not allow the computer option of copying their data. Furthermore, in analyses of this kind, it is impossible to completely avoid subjectivity. For instance, it was sometimes difficult to find a clear-cut distinction between similar cases from the meanings of *can*, *might* and *may* of the Albanian word *mund*. Another limitation is the compatibility of my data. As Schmied (2011: 16) states, the compatibility of data has always been a ‘major problem of comparative research in academic writing’. My corpora were small in size, and the exact same number of articles was not taken from every included news source. Furthermore, the corpora do not contain a clear quantification of its genres like how many editorials, reports, comments etcetera there are. The various genres and disciplines in which the papers have been written also have an impact on the writing style (Matarese 2013: 178). Therefore, because these corpora are not ideally stratified, this could have led to skewed results. However, the similar number of words, articles per journal, amount of news sources included per corpus is strength of this paper, and contributes to the reliability of the data used. Moreover, the concordance program AntConc 3.4.4 (2014) also contains some limitations. When I compared the amount of hits in the Key Word in Context (KWIC) tool, it was slightly diverse from the amount of hits in the Word List tool.

Further similar research should be conducted with larger corpora and more genres, as this would add more evidence to the use of hedges and boosters in international online news sources. Further research can be underpinned with a comparison of various genres of journalistic writing. From a theoretical perspective, it would be valuable to include Prince, Frader, Bosk’s (1980) subdivisions of shields (attribution and plausibility shields) and approximators (rounders and adaptors) in the
quantitative analysis and draw a direct comparison to Lafuente Millán’s categories. This would contribute to the research into two distinct but similar phenomena.

References


