UNIVERSIDAD AUTÓNOMA DE MADRID

A CORPUS-BASED ANALYSIS OF MODALITY IN THE WRITING OF SPANISH UNIVERSITY STUDENTS OF ENGLISH AT DIFFERENT PROFICIENCY LEVELS

Rebeca García González
71503964-T
07/09/2011

Abstract

This paper will examine the quantity and variety of grammatical resources that Spanish university students of English use to express modality in argumentative essays and their evolution, if any, at different proficiency levels. In particular, I will look not only at the prototypical realization of modality in English, that is to say, at modal auxiliaries, but also at other grammatical categories which can be and, indeed, are also employed to convey the semantic notion of modality, such as lexico-modal verbs, some lexical verbs, modal idioms and certain adverbials and adjective phrases.

Regarding modality among Spanish university learners of English, several studies have been carried out, mainly by the SPICLE team –the Spanish writing research team at the Universidad Complutense de Madrid. Their focus has been on the close analysis of modal verbs and their main goal has been determining to what extent factors, such as typological differences between the student's L1 and L2, L1 transfer of rhetorical conventions, developmental stages and the instructional factor, influence the students' use of modal verbs (Neff, J. *et al*., 2003; 2004).

However, no study has so far attempted to closely examine what grammatical strategies students employ to convey modality at different proficiency levels. In my opinion, this should be the starting point in organizing the teaching of this linguistic domain and also, in informing the elaboration of teaching and learning materials. Moreover, in the long term, this type of study could form a basis on which more theoretical work within the field of second language acquisition could be built.

As Aijmer, a pioneer in the field of corpus linguistics, said in her study of modality in Swedish learners of English (Aijmer 2002: 74), "before it is possible to speak of universal tendencies, more learner groups and perhaps also different proficiency levels should be

investigated". To this end, I will examine a section of the WRICLE corpus (Rollinson and Mendikoetxea, 2010), written by Spanish university learners of English with a wide range of proficiency levels. For the annotation of the corpus and analysis of data, UAM CorpusTool will be used (O'Donnell, 2008), since it enables automatic annotation of grammatical patterns, and provides statistical analyses.

My initial assumption is that, as students' proficiency level rises, both the quantity and variety of modal markers will also experience an increase since, in my view, a rising in the level of proficiency necessarily involves not only an increasing acquisition of the grammar and lexis of the language, but also of its pragmatics as well as the conventions of the genre (i.e argumentative essays). As Spanish native speakers, all the students whose texts are analyzed in the study are taught the pragmatic differences between English and Spanish, in particular, when writing argumentative essays.

In particular, when learning how to write this type of essays in English, these students are trained to qualify their claims so as to sound more indirect and, hence, more objective. Thus, my hypothesis is that the more advanced the students, the more qualification and tentativeness they will employ, which would involve a larger number and types of modal markers employed. However, as I will explain in the results section, the data obtained seem to defy my original conception. A closer look at the proportions and tokens used in each grammatical class established in this study as capable of expressing modality might shed some light on the apparently paradoxical results.

TABLE OF CONTENTS:

Pages

1. INTRODUCTION

In the last few years the area of modality has received wide interest among linguists, as shown by the high increase in the number of studies carried out on this topic, which not only analyze English native speakers/writers but also the interlanguage of learners of English (Selinker, 1972). In addition, with the emergence of corpus linguistics, new paths have been opened up, since this discipline allows us to examine, compare and contrast an unprecedented amount of data, taken from real language, either from native speakers and/or from language learners.

As a result of this, new types of research have emerged. For instance, Contrastive Interlanguage Analysis (CIA) (Granger, 1998) compares different linguistic domains in both native and non-native performance to see to what extent they differ. This line of research has yielded successful results mainly due to the collection of a pan-European non-native corpus known as ICLE –the International Corpus of learners of English-, led by Sylviane Granger and the Université du Louvain (Granger, 1993).

It is the important pedagogical applications this kind of studies may have that lead us to investigate here the use, types and realization of the most salient grammatical means to convey the semantic notion of modality. Since this is a corpus-based study and, therefore, the amount of data that will be examined is quite large, I will here set out to provide a quantitative analysis of the grammatical resources used by Spanish learners to express modal meanings, and to see their evolution as they rise in proficiency.

Thus, my study differs from others that have been done in the same field of modality, such as Aijmer (2002), Neff *et al.* (2003; 2004) and Soufleros (2010) –which I will review in the following section, in that I will examine the students' use of a larger number of grammatical

structures capable of expressing modality -which I will explain in the method section, and their evolution in quantity and realization across different proficiency levels. Hence, my focus here is on Spanish university learners and what they do at the different stages of their learning process regarding the grammatical expression of modality.

My assumption is that as students improve their level of English, they will display more quantity as well as more variety of modal markers in their writing. That is to say, it appears logical to expect more evaluation on the writer's part -expressed by means of a wider range of structures, in the more advanced students of English. Also, I presume that, as learners' proficiency level rises, there will be a movement away from verbal to non-verbal modality, since verbal modal markers such as modal auxiliary verbs are considered by most authors as the "prototypical" means of modal expression (Halliday, 1985; 1995; Halliday and Matthiessen, 2004).

2. REVIEW OF THE LITERATURE

2.1 Basic notions: mood and modality

To define modality, the first point we should make, following Huddleston and Pullum (2008) is that modality is a semantic category. Other authors, such as Palmer (1986; 2001), consider it as both a semantic and a grammatical category, within which mood and modal systems are subsumed as sub-categories. Furthermore, modality is generally associated with tense and aspect in that they are categories of the clause and usually marked within the verbal group. However, as Palmer put forward (1986; 2001) if tense and aspect are, notionally, "concerned with characteristics of the event" –its time and nature, respectively-, modality refers to "the status of the proposition that describes the event".

In his typological account of modality, Palmer (1986; 2001; 2003) distinguishes between two basic means that the languages throughout the world use to express modality, namely, mood and modal systems. Roughly speaking, mood is seen as a binary system, where there are two available possibilities, either "non-modal" or "modal", codified by the indicative and the subjunctive, respectively. Other labels for "non-modal" and "modal", which are pervasive in the literature, are "realis" vs "irrealis" or "asserted" vs "non-asserted (Palmer, 2001; Greenbaum *et al.*, 1985; Halliday, 1985; Downing & Locke, 2003; Halliday and Matthiessen, 2004). On the contrary, modal systems are typically composed of "a set of modal forms" (Palmer, 2001).

If Romance languages like Spanish, Italian or French make use of mood to realize the semantic category of modality, other languages like English rely solely on a modal system. This argument is used by several authors such as Palmer (2003), Plank (1984 as cited in Nuyts, 2001) and Lightfoot (1979) to conclude that both categories, mood and modal systems, are mutually exclusive. Nevertheless, Palmer points out the case of German, which seems to be a case in point in its use of both sub-categories, mood and a modal system (Palmer 2001; 2003), to encode modality.

2.2. The English Modal System

In this section I will focus on the English language and, thus, provide an overview of the main grammatical elements it offers to express modal meanings.

2.2.1. Modal Auxiliary Verbs

To begin with, it should be noted that English has "a modal system of modal forms, since the subjunctive mood has virtually disappeared" (Palmer 2003:4) in contrast to Romance

languages. According to authors like Coates (1983), Greenbaum *et al*. (1985), Halliday (1985) and Palmer (1979; 1986; 2001; 2003), among others, the prototypical grammatical elements to express modality are modal auxiliary verbs.

Following Huddleston and Pullum (2008), they can be defined as those elements that exhibit the following formal features:

(a) They take negation directly, eg. *cannot/can't, mustn't*.

(b) They take inversion without DO-support, eg. *Will you, may I?*

(c) They allow "code" or "stranding", that is, they can function as an operator without a main verb in reduced clauses, eg. *John can swim and so can Bill*.

(d) They allow emphatic positive, carrying the stress on speech, eg. *Mary WILL solve the problem.*

(e) No -s form for the third person singular of the present tense (*cans, *musts)

(f) No non-finite forms (*to can, *musting)

(g) No co-occurrence (*may will)

It should be noted that the first 4 characteristics are shared by all auxiliaries (Huddleston, 1976; Coates, 1983; Greenbaum *et al.*, 1985; Downing and Locke, 2003; Palmer, 2001; Aijmer, 2002) and are known in the literature as the "NICE" properties –standing for "negation, inversion, coda and emphasis". The last three features, from (e) to (g), are specific to modal auxiliaries, which allow us to distinguish them from the auxiliary verbs BE, HAVE and DO (Huddleston, 1976; Palmer, 1979 and 2001; Coates, 1983; Greenbaum *et al.*, 1985).

However, as is generally known, clear-cut categories are not the norm in languages, where indeterminacy and ambiguity play a central role. One of the reasons is that some elements belonging to a certain class, under certain conditions may behave as elements of a different class. This is what Greenbaum *et al*. (1985: 136) adduce to explain the behavior of a number of verbs "whose status is in some degree intermediate between auxiliaries and main verbs".

Using a gradience model with modal verbs at one end, and full lexical verbs at the other, as Table 1 below shows, Greenbaum *et al.* (1985: 137) include a cline of verbs which go from (a) central modals to (f) full lexical verbs with nonfinite clauses as complements (e.g. hope, begin), through (b) marginal modals, (c) modal idioms, (d) semi-auxiliaries, and (e) catenatives. Those verbs on top of the scale are semantically closer to auxiliary verbs, since their meanings are "associated with aspect, tense and modality" (1985: 136).

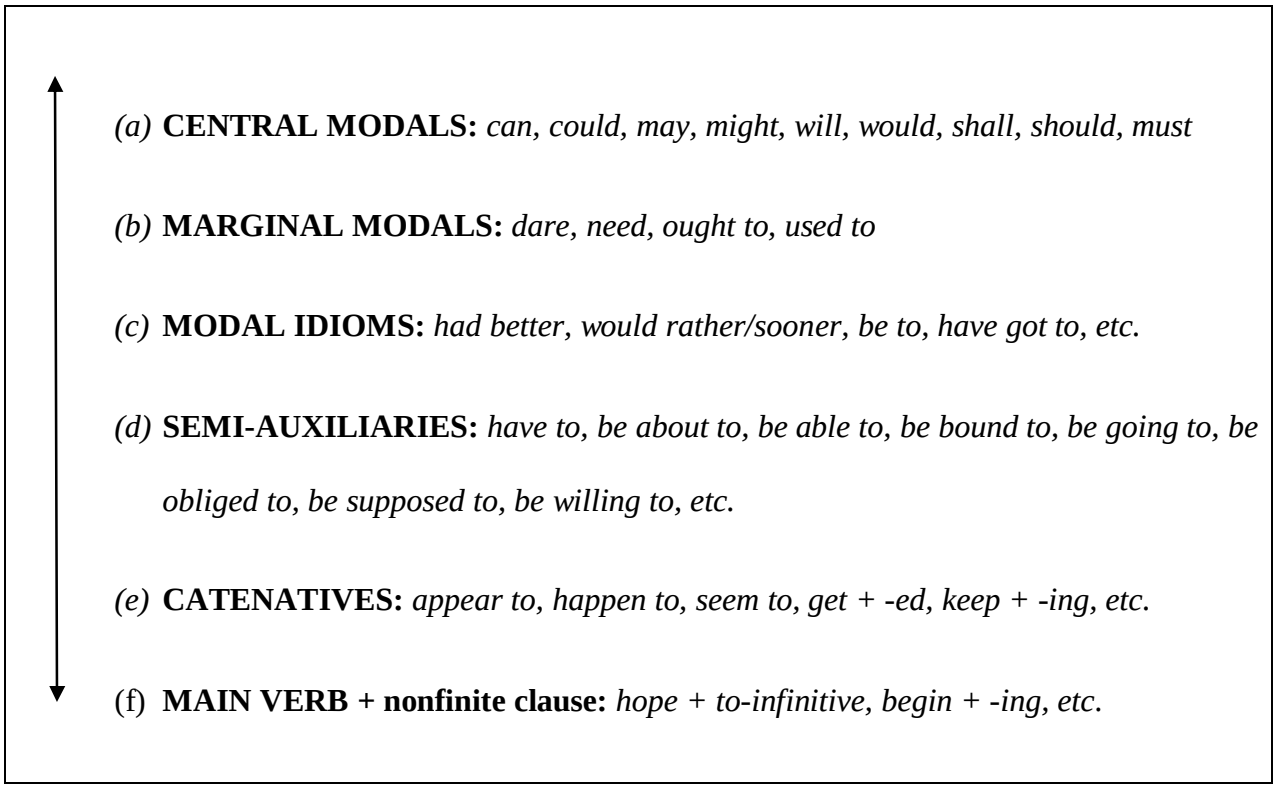


Table 1: The auxiliary verb – main verb scale (Greenbaum *et al*., 1985: 137)

To determine the status of these challenging verbal categories, from (b) to (e) in Table 1 above, Greenbaum et al. provide grammatical arguments which consist of applying the formal criteria mentioned above as the characteristics that determine auxiliary verbs –the NICE properties-, on the one hand, and those defining modal auxiliaries, on the other. Obviously, as we go down in the scale, the tests show more ambivalent results (see Greenbaum *et al*., 1985: 138-148 for a full account). In the method section I will explain in detail which of these categories I will include in my analysis and which ones I will not.

At this point I should bring back to mind Palmer's claim that "modal systems are typically composed of a set of modal forms" (2001: 4). As he explicitly claims, modal auxiliary verbs are not the only means available to express modality since "modal forms" include a wide variety of grammatical resources. Other linguists make similar claims, for instance, Halliday (1985), Greenbaum *et al* (1985), Hoye (1997), Nuyts (2001), Aijmer (2002), Downing & Locke (2003), Halliday and Matthiessen (2004).

2.2.2 Other modal grammatical elements

Most of the linguists who write about modality acknowledge that English has a rich and complex modal system, where, apart from modal auxiliary verbs, other grammatical elements such as lexical verbs, adverbs, adjectives and noun groups can be used to convey some kind of modality. Furthermore, and even more surprisingly, the combination of some of these elements, for instance, modal auxiliaries and adverbs -also known as "modal synergy" in the literature (Hoye, 1997), provides the speaker with even more resources to express a wide variety of modal nuances.

As it follows, the elements that the speaker can employ to express modality in English are numerous and highly dependent on context. Also, their consideration as modal expressions depends on the particular theoretical model of modality that is being used. Hence, in sections 2.3

and 2.4 I will provide an overview of the theoretical frameworks I consider most relevant in the field of modality. Moreover, in the method section (section 3) I will go back to the literature when I explain in detail which elements I have included within each grammatical class.

2.3 The semantics of modality

When looking at the semantics of modality, it should be mentioned that there is no consensus in the literature either on the number of types or on the actual types of modality. This, it seems to me, provides evidence of the complexity of the linguistic domain of modality. In this section I will offer a brief overview of the most salient and recurrent categories found in the literature whereas in the following section (2.4) I will present the models of selected linguists, which have most influenced the framework of the present study.

The two primary types of modality one can encounter in the literature are one "containing an element of will" and another "containing no element of will", which are respectively labeled as "**deontic**" and "**epistemic**" modality (Jespersen, 1924 as cited in Hoye, 1997: 41).

Greenbaum *et al.* in their discussion of the semantics of modality in English also point at this element of will, which they rename as "human control", as the factor that divides modality into two different types, which they term "**extrinsic"** and "**intrinsic"** modality (1985: 219), which correspond to epistemic and deontic, respectively. Their view is that extrinsic modality does not involve human control of events and they see the defining element of this type of modality to be the human judgment of "what is or is not likely to happen". On the contrary, intrinsic modality entails "some kind of intrinsic human control over events" and, therefore, includes the notions of "permission", "obligation", and "volition" (1985: 219-220).

The same distinction but again with different names is presented in Palmer (1979; 1986; 2001; 2003), where he distinguishes between **propositional** and **event** modality. If propositional

modality involves “the speaker’s attitude to the truth-value or factual status of the proposition”, event modality is concerned with “the speaker’s attitude towards a potential future event”. It, therefore, refers to “events that are not actualized, events that have not taken place but are merely potential” (Palmer 2001:8). For him, propositional modality includes epistemic and evidential modality, whereas event modality covers deontic and dynamic modality. I will examine these notions in the next subsections.

So then, most authors agree on establishing two broad types of modality, even though their labels may vary. In addition to the terms already mentioned, other names such as “**modality**” -or its improvement for clarity’s sake as “**modalization**” in Halliday and Matthiessen (2004)- and “**modulation**”, coined by Halliday (1985), or “**epistemic**” and “**root**” modality (Coates, 1983; Talmy, 1988 and Sweetser, 1990 as cited in Nuyts, 2001) are also frequently used in the literature. For ease of presentation, I will here use the traditional concepts of epistemic and deontic modality. To these I will add a third category, dynamic modality, following authors such as Nuyts (2001).

a) Epistemic modality

Following Nuyts, epistemic modality can be defined as the speaker’s “estimation of the likelihood that (some aspect of) a certain state of affairs is true/false in the context of the possible world under consideration” (2001: 21-22). Also, he mentions that “this estimation of likelihood can be situated on a scale going from certainty that the state of affairs applies to certainty that it does not apply, with intermediary positions on the positive and the negative sides of the scale” (2001: 22). This cline is reflected in all accounts of epistemic modality (see Halliday (1985), Greenbaum *et al* (1985), Palmer (1979; 1986; 2001; 2003), Nuyts (2001), Aijmer (2002), Halliday and Matthiessen (2004)).

Some authors include the notion of "evidentiality" within epistemic modality, like Bybee and Fleischman (1995) and Palmer (2001; 2003). If epistemic modality concerns the speaker's assessment of the probability of the event, evidentiality implies, following Chafe (1986: 262), "attitudes toward knowledge" (e.g., whether it is reported by someone, derivable from evidence, etc.). Palmer even distinguishes between two types of evidential modality, namely, "sensory" – which may be "visual, non-visual and auditory", and "reported" (2001: 9).

Besides this, Chafe, in his account of evidentiality, points out that a common trait of all the expressions conveying "evidentiality" is "their refusal to restrict themselves to evidence in a narrow sense" and "their extensions and shifts among various epistemological considerations" (1986: 262). Furthermore, he offers a schema to help us understand his position regarding "evidentiality" (see table 2 below), which involves various notions.

The first notion he puts forward is KNOWLEDGE, which he defines as "the basic information whose status is qualified in one way or another by markers of evidentiality" (1986: 262). On his schema, reproduced below, knowledge appears in the middle and vertically. Then, knowledge may be seen by a speaker as "more or less RELIABLE", as he visually shows "with the suggestion of a continuum from the most reliable knowledge, at the top, to the least reliable, at the bottom" (1986: 263).

In addition, he considers "various MODES OF KNOWING", that is to say, "various ways in which knowledge is acquired". Among them, he includes "BELIEF, INDUCTION, HEARSAY, and DEDUCTION" (1986: 263). As he explains, the source "for belief is problematic, for induction is EVIDENCE, for hearsay is LANGUAGE, and for deduction is a HYPOTHESIS" (1986: 263). Chafe highlights that their position in the schema does not involve their predominance over each other.

Finally, as can be seen on the right-hand side of table 2 below, he includes a distinction concerning the fact that "knowledge may be matched against VERBAL RESOURCES or EXPECTATIONS".

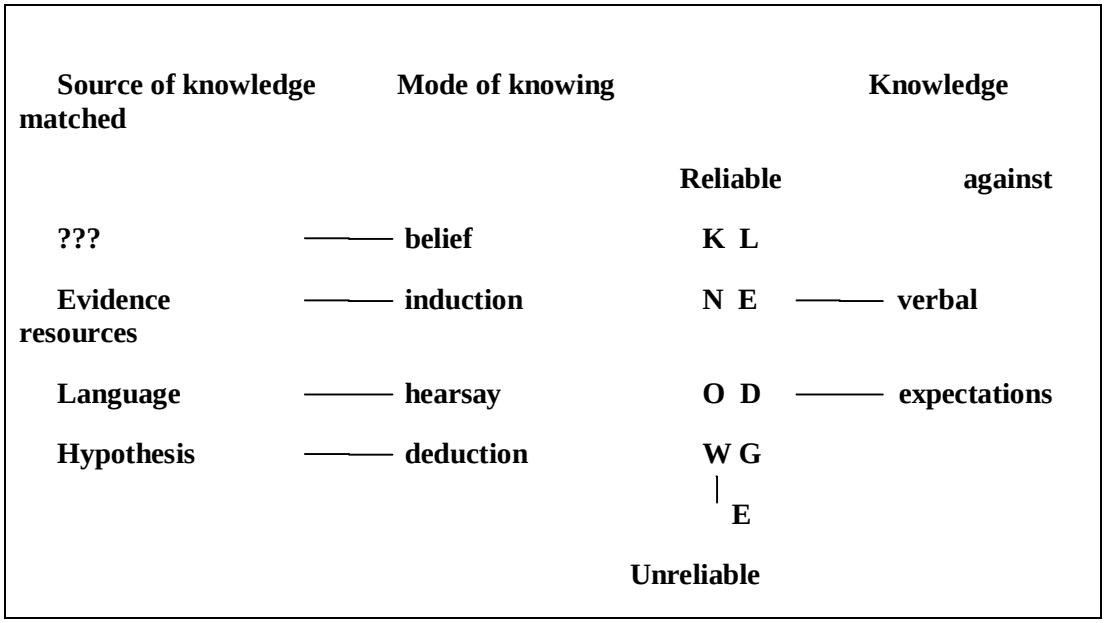


Table 2: Chafe's schema of "evidentiality" (Chafe 1986: 263)

On the contrary, Nuyts (2001), like other authors, excludes from modality in general and from the epistemic category in particular, the realm of "evidentiality", which he considers a clearly distinct category, although he acknowledges the difficulty to draw the borderline in some challenging cases.

Moreover, Nuyts explains that "evidentiality" merely concerns "the speaker's indication of the nature (the type and quality) of the evidence invoked for the state of affairs expressed in the utterance" (2001: 27). Thus, it is obvious that the basic difference he draws between the epistemic and the evidential semantic notions is that the latter lacks an evaluation or judgment on the part of the speaker regarding the state of affairs contained in the proposition.

b) Deontic modality

The other main type of modality, mentioned above, is deontic modality. Hoye considers it as expressing "the possibility or necessity of acts in terms of which the speaker gives permission or lays an obligation for the performance of actions at some point in the future" (1997: 43). This is why Jespersen talks about the existence of an "element of will" (as cited in Hoye 1997: 41) and Greenbaum *et al.* point at the notion of "human control" as crucial to define deontic modality (1985: 219).

Nuyts claims that deontic modality involves "an evaluation of the moral acceptability, desirability, or necessity of a state of affairs" and that it is concerned with the notions of "allowance", "permission", and "obligation" (2001: 25). In the same vein, Palmer highlights that in deontic modality there are conditioning factors which are "external to the relevant individual", therefore, deontic modality relates to "obligation and permission, emanating from an external source" (2001: 9-10).

However, a large number of authors, for instance, Perkins (1983 as cited in Nuyts, 2001), Palmer (1986; 2001; 2003), Nuyts (2001), split up the broad category of deontic modality into two different classes, namely, **deontic** and **dynamic** modality.

c) Dynamic modality

Palmer puts forward that the crucial element that defines dynamic modality is the "internal" conditioning factors that constrain the speaker's evaluation of a state of affairs and, thus, for him dynamic modality relates to "ability or willingness, which come from the individual concerned" (2001: 10).

Also, Nuyts considers that dynamic modality entails "an ascription of a capacity or a need to the subject-participant in the state of affairs, or of a situation-internal potential or necessity for him to do something" (2001: 25). Consequently, Nuyts claims that the inclusion of both deontic and dynamic in just one term is "unfortunate".

Moreover, he also argues that dynamic modality is "fully agent-oriented" whereas deontic modality is "as much speaker-oriented as agent-oriented" since "the permission, obligation, etc. is usually directed at the agent in the state of affairs, but the speaker is manifestly present as either the issuer of the permission, the carrier or messenger of a permission, etc" (2001: 25-26).

In the following section, I will examine all these notions in greater detail when offering a summary of the theoretical models that have inspired my own framework on modality.

2.4. Relevant theoretical models of modality

I will now review the theoretical framework of the authors I have followed in this study, that is, Halliday (1985; 1995; Halliday and Matthiessen, 2004) and Nuyts (2001).

2.4.1. Halliday's model

At this point, it is worth offering an overview of Halliday's ideas on modality, since I will basically rely on his framework in my analysis of the data. To begin with, he defines modality as "the various kinds of indeterminacy that fall between the positive and negative poles, like *sometimes* or *maybe*" (1985: 86; Halliday and Matthiessen 2004: 147). That is to say, for Halliday modality is also a semantic category which would cover the grey area in between the extreme poles, YES and NO.

Besides this, and following his approach to language, we should mention the distinction he establishes between "propositions" (i.e. statements and questions) and "proposals" (i.e. offers and commands), since they bring about two categories within modality. On the one hand, regarding propositions, the grey area between "it is so" and "it is not so" would include various degrees of probability and different degrees of usuality. For Halliday, these two in-between areas (i.e. probability and usuality) constitute **modality**, strictly speaking (1985: 86), although later on, he comes up with the notion of **modalization** for the sake of clarity (Halliday and Matthiessen 2004: 147). This category is equivalent to what is known in the literature as "epistemic" modality.

On the other, concerning proposals, the middle ground between "do it" and "don't do it" includes several degrees of obligation and inclination. He labels these two scales as **modulation** (1985: 86; Halliday and Matthiessen 2004: 147), to be distinguished from modality (1985), or rather modalization (2004) –I will here use the latter term to avoid ambiguity, as defined just above. Modulation is, therefore, a different label for "deontic" modality, understood as a broad category (including "dynamic" modality). Halliday considers both modalization and modulation within a wider semantic category which he labels modality and which, thus, corresponds to the traditional notion of modality.

As for the grammatical means by which modalization and modulation can be conveyed in the language, he mentions several resources. In the case of modalization, he considers three basic means: firstly, finite modal operators in the verbal group (namely, modal auxiliaries); secondly, modal adjuncts (i.e. "certainly"); and, thirdly, the combination of both categories together (e.g. "That will probably be John)[1] (1985: 86-88; Halliday and Matthiessen 2004: 147).

[1] It is worth mentioning that these are the grammatical categories Aijmer (2002) looks at in her examination of modality among Swedish learners of English.

In turn, modulation can also be expressed by means of a finite modal operator (that is, modal auxiliary verbs). In addition, it can be encoded by an expansion of the predicator (such as a passive verb, e.g. "*you're supposed to know that*"; or an adjective, e.g. "*I'm determined to win*"). Unlike modalization, modulation does not allow combinations of different categories (1985: 86; Halliday and Matthiessen 2004: 147). Table 3 summarizes Halliday's model of modality (1984: 87).

COMMODITY EXCHANGED	**SPEECH FUNCTION**		**TYPE OF INTERMEDIACY**		**TYPICAL REALIZATION**	**EXAMPLE**
Information	Proposition	Statement Question	**Modalization**	Probability	-Finite Modal Operator -Modal Adjunct -Both the above combined	*-They must have known* *-They certainly knew* *-They certainly must have known*
				Frequency	-Finite Modal Operator -Modal Adjunct -Both the above combined	*-It must happen* *-It always happens* *-It must always happen*
Goods and Services	Proposal	Command	**Modulation**	Obligation	-Finite Modal Operator -Passive Verb Predicator	*-You must be patient!* *-You´re required to be patient!*
		Offer		Inclination	-Finite Modal Operator -Adjective Predicator	*-I must win!* *-I´m determined to win!*

Table 3 Halliday's model of modality (adapted from Halliday 1985: 87)

As we can see, the grammatical means that can be used to encode modulation are a bit more varied than in the case of modalization –or rather, epistemic modality. This could be the reason why most studies on modality have focused on epistemic modality, or Halliday's

"modalization", rather than on deontic modality, or Halliday's "modulation". The latter appears to be a "fuzzy" domain in comparison with the much more restricted area of epistemic modality.

With the sole purpose of illustrating the ample variety of grammatical elements that English speakers have at hand when expressing modality, I will offer Halliday's "systemic paradigm" to express "probability" (1985: 332-33; Halliday and Matthiessen 2004: 148-49). He establishes, on the one hand, the dichotomy subjective (*I´m certain, must*) vs. objective modality (*it is certain, certainly*) and, on the other, explicit vs. implicit modality.

If by means of elements such as *I'm certain,* the speaker explicitly presents "the source of the conviction as a subjective judgment on the speaker's part", by the use of expressions like *it is certain* "the source of the conviction is explicitly said to be objective" (2004: 149). Along with these, Halliday explains that the source of the conviction may be left implicit (i.e. *must, certainly*). However, these last implicit modal markers "also differ along the subjective and objective dimension" since "the adverbial form *certainly* is a way of objectifying the speaker's evaluation" whereas "the verbal form *must* carries a subjective loading", as if "it is the speaker's own judgement on which the validity of the proposition is made to rest" (2004: 149-50).

	Subjective	**Objective**
Implicit	*must*	*certainly*
Explicit	*I´m certain that*	*It is certain that*

Table 4 Systemic paradigm of probability (Halliday and Matthiessen 2004: 150)

Furthermore, Halliday puts forward the notion of "metaphors of modality" to refer to the process by which modality is encoded as "a separate, projecting clause in a hypotactic clause complex, which becomes the primary clause" (1985: 332; 2004: 613). And, as a piece of evidence, he applies several grammatical tests like tagging the clause:

i) *I think it's going to rain, isn't it?*

ii) *I think it's going to rain, don't I?*

Obviously, and quoting Halliday (1985: 333; 2004: 614), only i) is "equivalent to *It's probably going to rain (*tag *isn't it?)"*, whereas ii) corresponds to the proposition *John thinks that X*" (tag *"doesn't he?").*

Halliday says that further evidence for the existence of modality metaphors is that negation can be transfered from the embedded to the main clause without any change of meaning or problems of sense. He offers the following examples:

i) ***Subjective:***

I think Mary *doesn't* know/I *don't* think Mary knows

ii) ***Objective:***

It's likely Mary *doesn't* know/It *isn't* likely Mary knows

If we compare them, we will conclude that the meanings of the two sentences in either i) or ii) are identical. In his analysis, Halliday claims that "since modality is being dressed up as a proposition, it is natural for it to take over the burden of yes or no." (1985: 333; 2004: 616)

Finally, Halliday suggests that what determines the grammatical means that will be used to express modality is "orientation", which he defines as "the distinction between subjective and objective modality, and between the explicit and implicit variants" (2004: 619). The following figure illustrates these four variables (see Table 5 below):

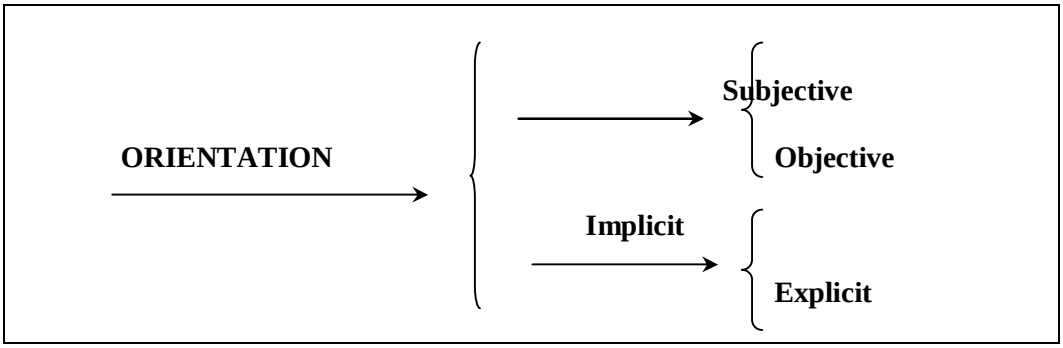


Table 5 System of types of orientation in modality (2004: 619)

2.4.2. Nuyts' theoretical considerations

Nuyts' (2001) conceptual framework of modality, or more specifically, of epistemic modality not only goes back to some of the traditional semantic categories already mentioned in section 2.3 (i.e. epistemic, deontic and dynamic modality), but also comes up with some important notions, which I consider extremely useful for clarifying the "fuzzy" domain of modality.

Drawing on the work of previous authors such as Lyons (1977), Nuyts analyses certain grammatical categories, namely, modal adverbs, predicatively used modal adjectives, mental state predicates (*think, believe*), and modal auxiliaries, when studying epistemic modality cross-linguistically. He argues that all these grammatical means can be used to "express epistemic modality in most or all West European languages" (2001: 29).

More specifically, he examines three languages, i.e. Dutch, German and English. He uses Dutch and German corpora for his analysis, however, in the case of English he relies on translations from those corpora and on his knowledge of and about the language, which might be considered a shortcoming of his conclusions. In any case, his argumentation is thoroughly explained, logical and well-supported.

Besides the grammatical structures above mentioned, he acknowledges the existence of other grammatical forms used to express epistemic modality such as nouns, modal particles and certain tenses in West European languages. Nonetheless, he does not take them into account because of their low frequency, even though some of them are present in some of the languages he looks at –like nouns in English.

Nuyts puts forward an important conceptual clarification in regards to epistemic modality: he re-analyses the traditional distinction between subjective and objective (epistemic)

modality (Lyons 1977, among others) as subjective vs. intersubjective (epistemic) modality. Using the following example:

(1) Alfred may be unmarried,

Nuyts reflects on the two interpretations Lyons makes of it, namely, "subjective" modality – in which "the speaker is uncertain about the (hypothetical) fact that Alfred is unmarried"- or "objective" modality –in which "the speaker knows there is a mathematically computable chance that Alfred is unmarried, for example, because Alfred belongs to a community of ninety people of which there are thirty unmarried, hence there is one chance in three that he is unmarried" (2001: 33)

Nuyts, however, offers a different analysis. On the one hand, there is "the speaker's evaluation of the probability of the state of affairs", that is to say, the "epistemic qualification". On the other, there is "his characterization of the status or quality of the sources (evidence) for that qualification", which pertains to the realm of "evidentiality". Hence, Nuyts concludes that "the dimension of intersubjectivity in not inherent in the epistemic domain but belongs in the evidential domain". Moreover, he also denies the existence of two types of epistemic modality, subjective and objective. For him what exists is "an interaction of an epistemic with an evidential qualification" (2001: 35).

Also, he applies this same reanalysis to deontic modality, with independent expression forms such as "In my view/opinion/mind", "if you ask me", "to me", which would express subjective deontic modality as opposed to "it is well known", "you know", which can be used to qualify information about a state of affairs as intersubjective (2001:36).

Another important distinction that Nuyts makes within epistemic modality is between performative and descriptive expressions. The former convey "the speaker's current attitude

towards a state of affairs" which implies his actual commitment to the evaluation at the time of producing the utterance whereas the latter merely "report on an epistemic qualification of a state of affairs without involving speaker commitment to it at the moment of speaking" (2001:39).

To illustrate this, he offers the following examples (2001:39):

a) *I think they have run out of fuel.*

b) *John thinks they have run out of fuel.*

c) *I thought they had run out of fuel.*

The first example clearly shows the "speaker's evaluation of a state of affairs to which (s)he is fully committed at the time the utterance was produced". This is what Nuyts calls "performative" epistemic expressions (2001:39). However, b) and c) differ from a) in that they involve either the "speaker expressing someone else's evaluation of a state of affairs without showing his/her own commitment to it" –as in b)- or the "speaker reporting his own past evaluation of a state of affairs but giving no indication of his current commitment to it" –as in c) (2001:39).

These two last types of epistemic expressions, (b) and (c), are labeled as "descriptive" by Nuyts and he does not consider them to be purely modal. It is important to mention that Nuyts extrapolates this distinction of performative vs. descriptive modal expressions to the domain of deontic modality (2001:40).

Regarding the particular grammatical categories Nuyts examines, he establishes a difference between mental state predicates and other categories of predicates "sometimes used to express a similar kind of speaker qualification". Within the latter group, he mentions predicates such as *say* and *hear*, which have no epistemic meaning but "a purely evidential one of hearsay" (2001:109)

However, when dealing with mental state predicates, Nuyts explicitly mentions that this class is more open and, therefore, more ambiguous than the modal adverbs and adjectives. One of the reasons he offers is the role of evidentiality in the inherent meaning of the mental state predicates, which I will peruse below. Another is the strength of the evaluation expressed by verbs such as *think, believe, suppose* or *guess*, which is rather vague. For Nuyts, these mental state predicates simply express "it is somewhere on the positive side of the epistemic scale, beyond mere possibility" (2001:111).

As I have already hinted at, one of the central features of this class, according to Nuyts, is that they conflate an epistemic and an evidential meaning. This combination is not so evident, although it is always present, in the case of *think, believe* or *doubt,* which Nuyts considers "the prototypical members of the category of mental state predicates expressing epistemic modality". However, it appears more clearly when closely analyzing the meaning of other mental state predicates like *know, suppose,* and *guess* (2001:112).

Thus, he provides the following definitions: *know,* "I consider it (quite) certain/I am convinced that something is the case, although I have no hard proof for it and can only say so on the basis of intuition"; *suppose,* "I consider it likely/I expect that something is the case because I have good evidence for it"; and *guess,* "I consider it probable that something is the case, even though I have no firm evidence for it" (2001:111).

Therefore, he concludes that all of these entail an epistemic as well as an evidential meaning. The only difference between the former –*think, believe* and *doubt*- and the latter group –*know, suppose* and *guess*- is that the epistemic meaning is more central in the former whereas the evidential is more prominent in the latter, but both meanings are inherently present in all of these mental state predicates (2001:112).

Nuyts suggests an interesting semantic characteristic of the mental state predicates, which is their "non-qualificational" meaning (2001:113). Even though this meaning is different for each predicate, he attempts to characterize it by offering the following general definitions, either "being in mental state X" or "performing mental process X" (2001:113).

In particular, he provides the following non-qualificational definition for each mental state predicate: *think* is "the mental process of reasoning with knowledge"; *believe* is "the mental state of faithfully trusting information (from someone); *guess* is "the mental process of tentatively estimating something"; *know* is "the mental state of having knowledge about something"; *suppose* is "the mental process of hypothetically putting forward something in order to be able to use it as a premise in a reasoning process" (2001:113).

Moreover, Nuyts also shows how this semantic difference between a qualificational and non-qualificational meaning of the mental state predicates has its reflection in the syntax. Focusing only on *think* and *believe* for space constraints, he provides the typical syntactic patterns in which the non-qualificational meaning appears. For instance, in the case of *think*, it is triggered "when it occurs intransitively", as in "*Shut up, I am thinking"*; "when combined with a few fixed prepositions", as in *"I am thinking of going home today"*; or "when introducing a quote", as in *"I thought: Oh no! Not again!"* (2001:116)

Concerning *believe*, its non-qualificational meaning occurs when it is used "with a simple transitive pattern", as in "*I believe you/this story*", or "with a complement to be interpreted as "I accept that…", as in *"I believe that this story is true";* or, like *think*, "when combined with a few fixed prepositions", as in "*I don't believe in miracles*" (2001:116).

Conversely, the qualificational meaning of all the mental state predicates seems to arise in two basic syntactic patterns. These are either when they take a that-clause complement, i.e. "*I think/believe (that) they have run out of fuel*"; or a "reduced variant of it", as in *"I think/believe*

so"; or when the mental state predicates are "used parenthetically", as in "*It is dangerous, I think/believe, to run out of fuel in a desert*" or "*It is dangerous to run out of fuel in a desert, I think/believe*" (2001:117). As it follows, Nuyts only considers the qualificational meaning of the mental state predicates as expressing epistemic modality.

Table 6 below offers a summary of Nuyts' theoretical model of modality:

Grammatical elements expressing (epistemic) modality	**Tokens within MSP capable of expressing epistemic modality**	**Semantic distinctions he draws within MSP**	**Semantic notions of MSP he considers as modal**
Modal Adverbs Predicatively used modal adjectives Mental state predicates (MSP) → Modal auxiliaries	think, believe, doubt, know, suppose and guess	Descriptive vs. Performative Qualificational vs. Non-qualificational	Performative Qualificational

2.5. Previous corpus-based studies on modality

After having offered an overview of the most relevant theoretical models of modality that inspire my analysis, I will now present three studies, which bear a resemblance to this in being corpus-based examinations of the use of modality in the writing of university learners of English. However, they differ from mine in various respects, as I will explain below.

2.5.1. Aijmer's corpus-based analysis of modality

Aijmer's work (Aijmer, 2002) should be highlighted since her seminal analysis of modality among learners of English stands out as a landmark. It is a contrastive and corpus-based account of the use of modality among non-native learners of English. In particular, she analyzes the range and frequency of modal forms employed in argumentative essays written by advanced Swedish learners, their meanings and uses.

The type of grammatical structures expressing modality she focuses on are modal auxiliaries, modal adverbs and their harmonic and disharmonic combinations as well as certain lexical verbs with modal meaning such as "I think" or "I believe" (Aijmer, 1996). Thus, she basically relies on Halliday's framework (e.g., Halliday, 1985; 1994; Halliday and Matthiessen, 2004).

To that end, she primarily uses the Swedish section of the ICLE corpus but also, on occasion, the French and German sections, along with the LOCNESS -the Louvain Corpus of Native English Essays- and the British section of the LOB corpus –the London-Oslo/Bergen corpus (Granger, 1998). In her exploration and comparison of data she highlights that all the above mentioned grammatical categories were overused by the Swedish learners (Aijmer, 2002).

However, what is relevant here is that Aijmer herself explicitly mentions the risks of considering that students overuse all the expressions of modality when only a certain number of categories were analyzed in her study (2002). In fact, grammatical categories such as adjective phrases or other lexical verbs different from *I think/believe*, which can be used with a modal meaning, were not taken into account in her investigation.

Thus, it could well be the case that learners overused certain structures, the ones included in her analysis, at the expense of others, which have been disregarded in her study. Nevertheless, as Aijmer herself points out, undertaking a corpus-based study of modality is a "complex

process", because of the various means that the English language can use to express this semantic category (2002).

2.5.2. SPICLE corpus-based examination of modality

In the same vein, but working with the Spanish section of ICLE, the SPICLE research team has also examined modality in the academic writing of Spanish university students (Neff *et al.* 2003; 2004). In particular, they have analyzed their use of modal verbs –along with reporting verbs- expressing epistemic modality –they include evidentiality- to see whether they differ from the patterns displayed by similar non-native students of English –with a different L1, and American university students. Their final goal is to ascertain the role that L1 transfer, second language constraints, developmental process and teaching might play in the acquisition of this linguistic domain (2003; 2004). In their study, they do not distinguish between proficiency levels.

Their findings show that, on the one hand, the modal auxiliary verb CAN is overused by all non-native writers, but the highest frequencies are found among Spanish and Italian students. They resort to L1 transfer and, possibly, to a teaching effect to account for it. On the other, the modal auxiliaries COULD, MAY and MIGHT are underused, especially by Spanish students.

In this case several factors are put forward to explain these tendencies, for instance, typological differences between Spanish and English, the "transfer of sociolinguistic norms" for formal writing from students' L1 to their L2, and an ineffective teaching method (2003; 2004). Their recommendations are improving the teaching of this particularly challenging linguistic area by explicitly teaching modal verbs in their authentic context of use and in a higher number of contexts (2003; 2004).

2.5.3. Soufleros' corpus-based investigation of modality

Analyzing the Greek section of the ICLE corpus –GRICLE and a Spanish learner corpus collected at the Universidad Autónoma de Madrid –WriCLE, Soufleros attempts to see whether the student's first language (and its culture) has an effect on their expression of modality in their academic writing (Soufleros, 2010). To do so, he analyses the systemic categories of modality established by Halliday (Halliday, 1984; 1994; Halliday and Matthiessen, 2004), namely, type, orientation, value and polarity in a number of essays dealing with the same topic. There is no distinction of proficiency levels in his analysis.

His findings appear to support his initial assumptions, since Spanish-speaking learners systematically use more modalization and positive polarity than Greek students. Soufleros, thus, claims that students' first language and culture plays a crucial role in the semantic profiles of modality (Soufleros, 2010). However, in the light of the results he obtains, he concludes that not all the examined categories of modality are equally influenced by the student's mother-tongue and culture. Only modalization and value –which are of great importance in the argumentative genre- show a significant difference in the writing of the Greek and Spanish students (2010).

3. METHOD

In this section I will deal with three main aspects of my study: first, the learner corpus I have used (section 3.1); second, the semi-automatic software employed in the analysis of data (section 3.2); third, the criteria underlying my selection and analysis of data (section 3.3).

3.1. A corpus-based study: the WriCLE corpus

For this study I have used a section of the WriCLE corpus (Rollinson and Mendikoetxea, 2010), which is a corpus of essays written by Spanish university students learning English at the Universidad Autónoma de Madrid (UAM). WriCLE stands for ***Wri**tten **C**orpus of **L**earner **E**nglish*. The corpus was collected by Paul Rollinson during 2005-2008, and consists of 719 essays containing approximately 710,000 words[2].

The corpus was collected as part of the WOSLAC project (Chocano *et al.,* 2007) and consists of essays submitted as class assignments within Academic Writing courses in the first and the third year of the English Studies degree. Paul Rollinson, the teacher, normalized the submitted text in accordance with the process used in the ICLE corpus (Granger, 1993): all personal data, titles, footnotes, endnotes, Graphics, maps and bibliographies were stripped out, and quotations and references were replaced with <Q> and <R> respectively. The essays are stored in electronic format and range from 500 words up to 2,000 words.

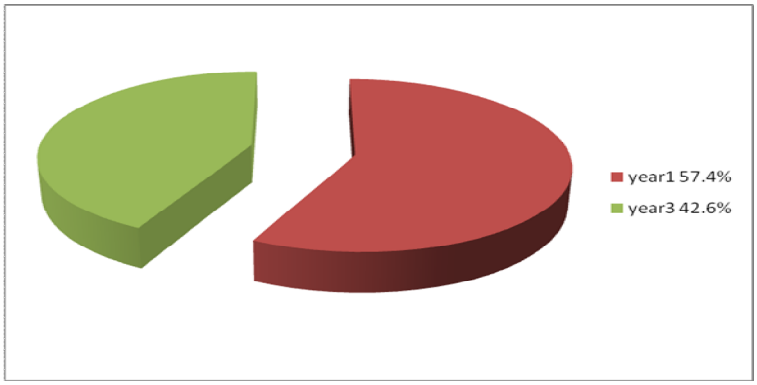
Various metadata were collected as well. These include a *Learner Profile* and a *Release forms/Essay Profile*. The former contains the age, gender, language background, English language proficiency[3], etc. of the students whereas the latter was provided by the learner for each essay, detailing the resources they employed to produce the piece of writing. The form also includes a section where the student grants permission for the essay to be used for research purposes.

The essays include both discussion and personal opinion essays on controversial and current topics, such as abortion, euthanasia, the legalization of marijuana or gay marriage. Since it is a learner corpus, mistakes will obviously appear in some of the examples I will provide in the following sections, even though they do not constitute our focus here.

[2] Although 752 essays were collected, 33 were eliminated because the learner's L1 was not Spanish.

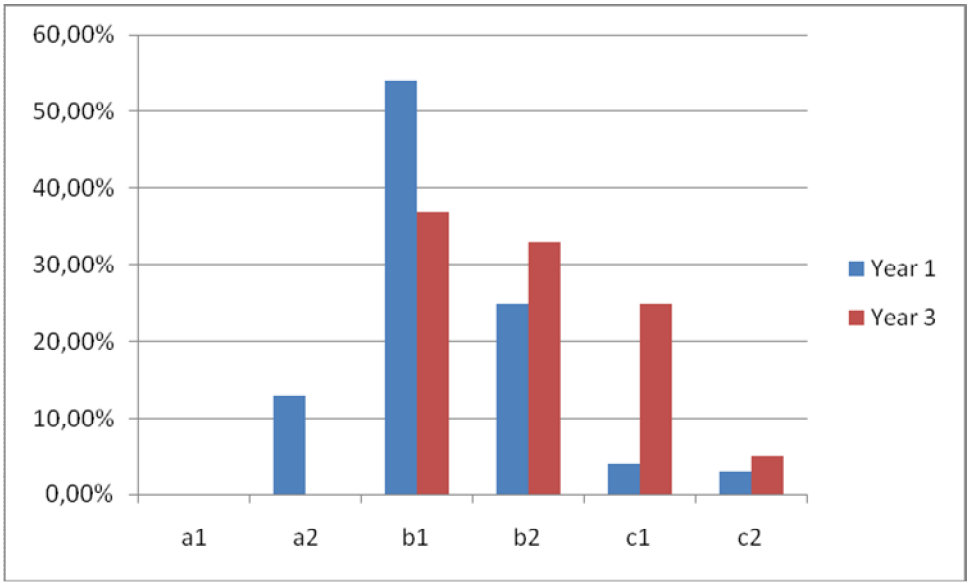
[3] All the students took the *Oxford Quick Placement Test* (UCLES, 2001) at a time close to the writing of the essays, which allowed to ascribe them a CEFR proficiency level.

As I have mentioned above, only a section of WriCLE has been analyzed for this study. My corpus, therefore, consists of 458 essays or 445,776 words. This is basically the full WriCLE corpus except for the texts which lacked the placement test results. Of the 458 essays, 263 were written by first-year students and 195 by third-year students. In percentages, the former represents 57.4% whereas the latter is 42.6% of the total number of essays examined. Graph 1 displays these percentages.



Graph 1: Percentages of essays written by first- and third-year students

It is important to highlight that first and third-year students can, and indeed have, a wide variety of levels of proficiency. That is to say, being a first-year student does not necessarily mean having a lower level of proficiency than a third-year student. Likewise, being a third-year student does not necessarily involve a higher proficiency level than a first-year learner, as Graph 2 below illustrates:



Graph 2: Levels of proficiency across academic years at university

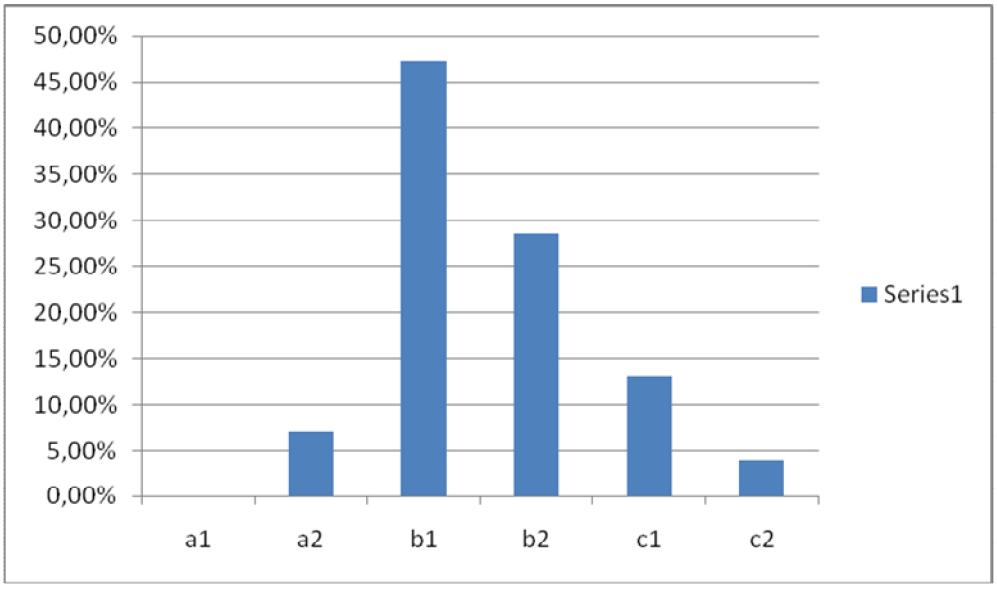
Since the main goal of this study is to see which grammatical structures students make use of to express modality at different proficiency levels, the essays will be grouped and examined according to their authors' level of proficiency, regardless of the academic year in which they were written. To this end, the metadata included in WriCLE, which offer the proficiency level of the writers, were looked into.

I should mention that the section of WriCLE I have analyzed does not contain longitudinal data. This is, therefore, a cross-sectional study. As for the proficiency levels of the writers of the texts examined, none of them have the lowest proficiency level, namely, A1, following the Common European Framework of Reference for Languages (CEFRL) (Council of Europe, 2001). The lowest level of proficiency attested in the section of WriCLE analyzed is A2, which represents 7.2%, and is visually illustrated in Graph 3 below.

Almost half of the total number of students analyzed (47.2%) are B1 students, which comes as no surprise if we consider that Spanish high-school students are expected to reach that level by the end of "Bachillerato". Since more than a quarter of the students (28.6%) are classified as

B2, we can safely state that the majority of the writers have an intermediate command of English.

As for the advanced levels, only 15% of the students can be considered as such, with the highest percentage of students in C1 (13.1%) and a mere 3.9% in C2.



Graph 3: Levels of proficiency in the section of WriCLE used (in percentages)

3.2. Procedures: UAM CorpusTool

For both the automatic grammatical analysis and manual inclusion and exclusion of elements, I have used UAM CorpusTool (O'Donnell, 2008) which allows manual and automatic annotation of collections of text at multiple annotation layers. Figure 1 below shows the main window of UAM CorpusTool with the section of the WriCLE corpus used open. The window shows that 7 annotation layers are defined, from left to right:

- Document: where features relating to the document as a whole and its writer are recorded. In particular, the kind of information it contains is the proficiency level of the writer (according to

the CEFRL, Council of Europe (2001)), their university year, gender and native language (for this study always Spanish), and the language of the document (in this study, always English).

- Error: This layer is used to record the manual annotation of errors, which in this case has not been used.

- Grammar: This layer is used to record the automatically generated grammar analysis of each sentence.

- STNDFParse: To produce our grammatical parse, UAM CorpusTool first parses each sentence using the Stanford parser. The Stanford parse trees are recorded as a layer on their own.

- Sentence: Each text is automatically segmented into sentences, and this annotation layer records the start and end of the sentences. The interface allows the user to manually correct wrong segmentation made by the software. I have not coded modal markers in relation to the sentence, but to the clause. However, the sentence level is used to allow us to make cross-level queries regarding sentence units (e.g., sentence containing nonfinite-clause would find all sentences containing a non-finite-clause).

- Modality: I used pattern matching (concordance patterns), followed by the elimination of false matches, to get the software to automatically assign the type of grammatical element a particular modal marker is, using the taxonomy I will present in section 3.3.

- Word: It parses every word to enable us to do searches of single tokens.

To the right of the list of layers, the window shows the files included in the corpus. For each file, there is a button for each of the layers, and pressing one of these buttons brings up the editing window for the file at that layer.

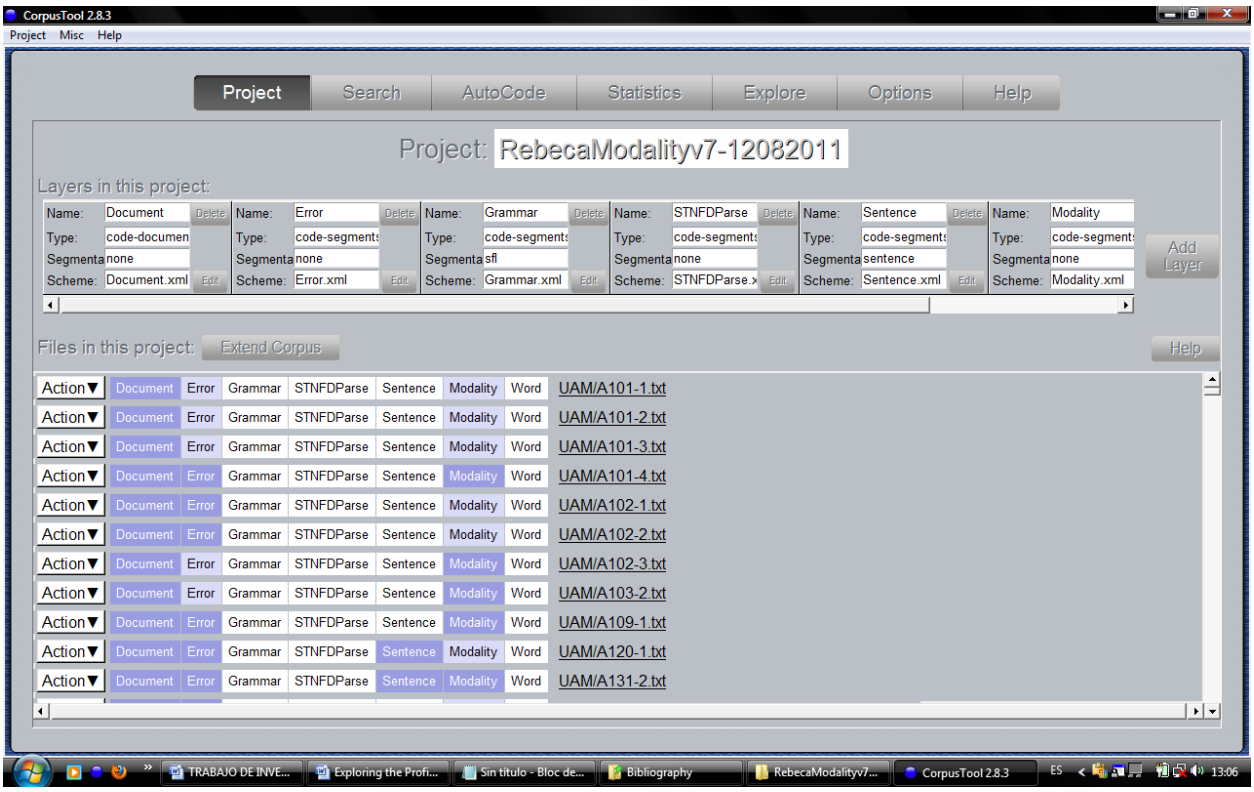


Figure 1: Main Window of UAM CorpusTool with the section of the WriCLE corpus used open

One of the main advantages of working with UAM CorpusTool is its enormous flexibility. This feature is crucial in order to undertake this study since, as I will explain in the following section, the grammatical elements included in my analysis as modal markers depend, to a great extent, on their grammatical environment to have a modal meaning.

For instance, and only for the sake of clarification, the lexico-modal verb *have to* behaves as a modal marker in contexts where it equates to the modal auxiliary *must*. However, its use in instances such as "This *has to* do with the fact that homosexuality is still a taboo in our society" (UAM/A173-2) cannot be considered as modal and, therefore, has been manually excluded from my analysis.

Also, the restricted uses of the lexical and lexico-modal verbs included in my examination have been manually selected as well as the distinction between different grammatical categories. That is, elements such as *likely* or *probable* can be adjectives or adverbs depending on the context where they occur. Hence, I have manually tagged them assigning them the right grammatical class for the context in

which each token appears. In addition, certain uses of modal auxiliary verbs which have not been included in my framework of modality have been manually excluded (e.g. WOULD when used to express usuality in the past).

To do this, UAM CorpusTool offers an extremely useful application, which is the Auto-code facility –see Figure 2 below. It allows the user to perform a concordance search, and create segments on the modality layer corresponding to the search hits. For example, the following creates an Autocode rule to create modal tokens tagged as "lexical-verb", wherever the program finds a 1st person singular pronoun followed by a token defined as a mental-verb in its dictionary.

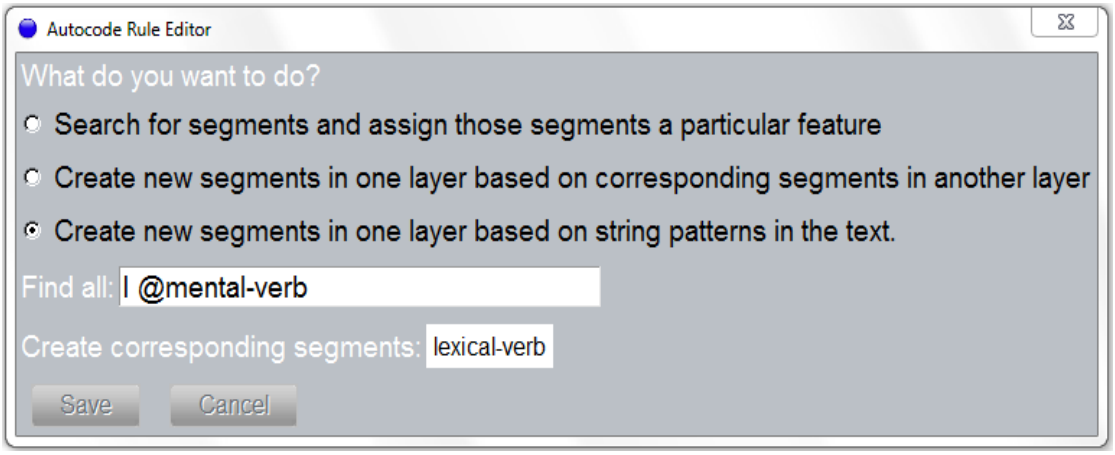


Figure 2: The Autocode facility

The program then displays all hits. The user can deselect all instances which are false matches. Then, pressing "Code selected" creates segments for each remaining match, and assigns them the features: modal-marker>verbal-modality>lexical-verb. Figure 3 below shows all the matches of *think* and how clicking on the magnifying glass the software retrieves the token in its original context.

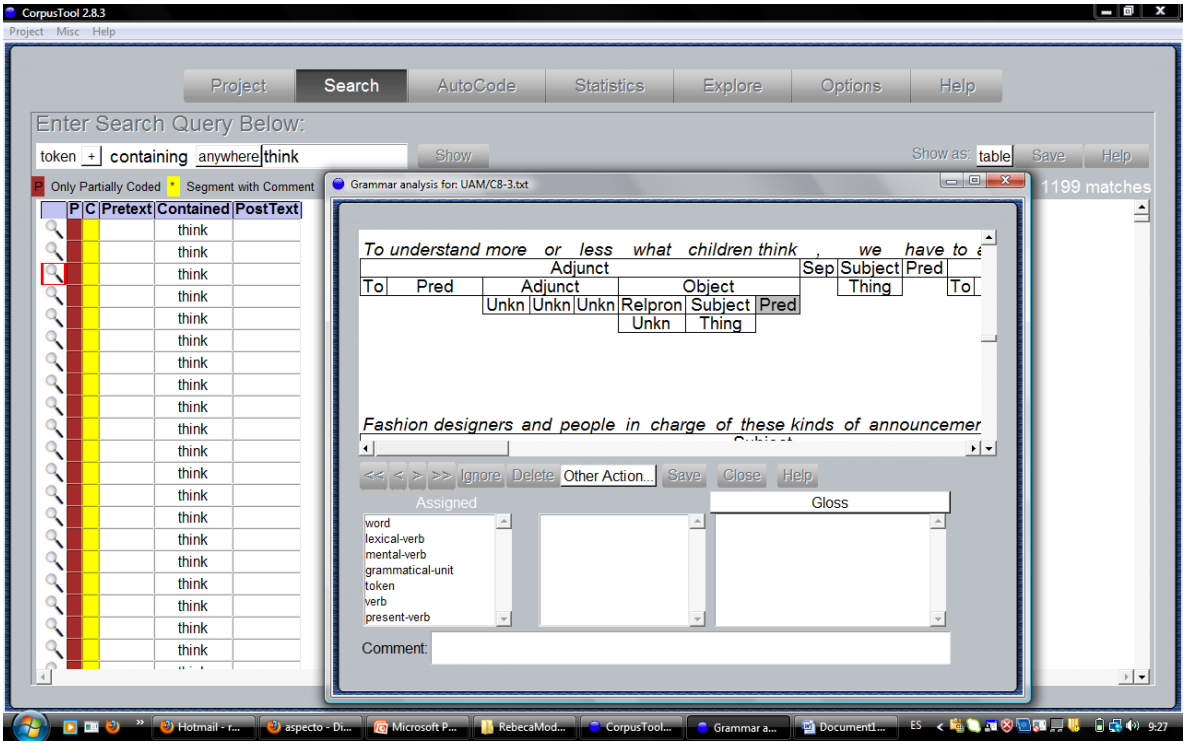


Figure 3 Uam CorpusTool

Regarding statistics, I cannot apply any conventional statistical method, such as t-test or chi-squared test, since I am not comparing two datasets, as is more common in research studies (in this case I am comparing five sets of data, i.e. the usage of modality at the five proficiency levels the students display). In order to establish the statistical significance of the movement over the five levels of proficiency some form of ANOVA would seem to be appropriate, but more work is needed to establish exactly what type of significance test would be appropriate for the current study. The time-scale of the DEA did not allow sufficient time to explore this issue fully.

3.3. Analyses

After having described the corpus used as well as the procedures, I will now describe the main principles guiding my selection and analysis of data.

- **The semantics of modality:**

Drawing on the literature on modality, summarised and discussed in the literature review section (section 2 above), I will briefly put forward the decisions taken for the design of this study. To begin with, it is worth highlighting that, following the majority of authors, I consider modality as a semantic category, whose defining element is, as Nuyts (2001) claims, the speaker's evaluation of the likelihood of the event expressed in the proposition –namely, epistemic modality; of the necessity, moral acceptability or desirability of a given state of affairs –that is to say, deontic modality; or of someone's ability, intention or willingness to do something -dynamic modality.

In turn, within epistemic modality I will further distinguish between certainty, probability or possibility, depending on the speaker's degree of commitment to the truth value of the proposition. Regarding deontic modality, three main subclasses can also be established, i.e. obligation –with two subtypes, which are inescapable obligation, on the one hand, and unfulfilled obligation or advisability, on the other-; necessity and its counterpart, lack of necessity; and permission. Finally, in my view, dynamic modality covers three main semantic subtypes, ability, willingness and intention.

The following scheme (see table 7 below) tries to summarize the main modal classes and subclasses I will consider from a semantic point of view.

At this point it should be mentioned that some notional categories, traditionally included under the umbrella of modality, will not be considered here for not involving an evaluation on the part of the speaker. In particular, the domain of usuality, traditionally included within modality (see my summary of Halliday's model in section 2.4.1), will be excluded here since it does not fit in my theoretical model, briefly presented above.

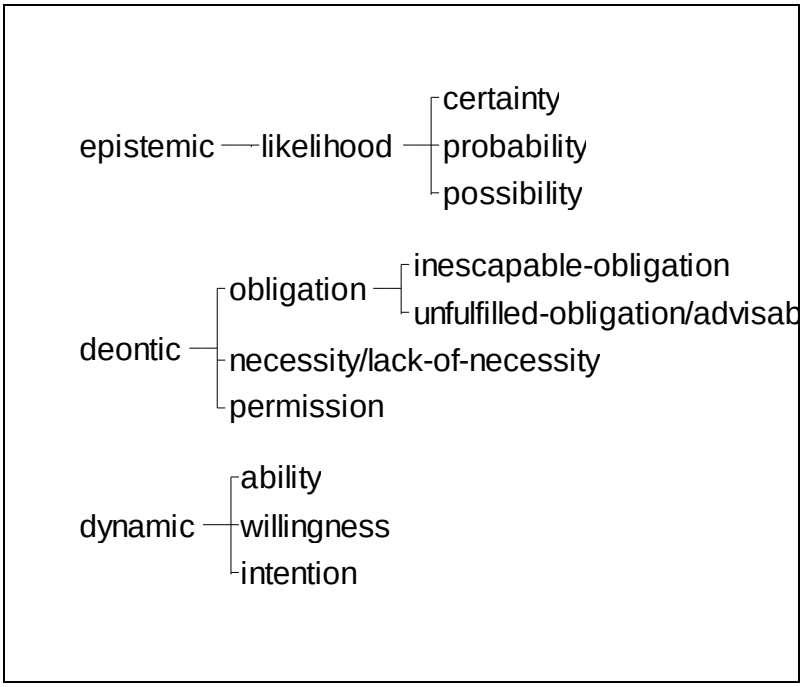


Table 7 Semantic types of modality included in this study

Also, the semantic dimension of evidentiality, regarded as a subtype of epistemic modality by some authors (among others, Palmer, 1979; 1986; 2001 and 2003), will not be taken into account here because it lacks the aforementioned defining element of an evaluation by the speaker. To this respect, I will follow Nuyts (2001) in his claim that "evidentiality" merely concerns "the speaker's indication of the nature (the type and quality) of the evidence invoked for the state of affairs expressed in the utterance", rather than the speaker's evaluation on the basis of evidence.

After having presented the main semantic types and subtypes of modality I will consider in the study, I should emphasize that I have not tagged every modal expression in the corpus as belonging to one particular semantic class/subclass or the other. As I have pointed out in the method section, this is a grammatical account of modality. Besides this, the large amount of data used prevents us from manually tagging every single modal expression according to its semantic type, since to do so the context where a modal marker appears has to be taken into account.

Therefore, the value of establishing my own model of the semantics of modality lies in the fact that it has informed my decisions when selecting the grammatical elements to be analyzed. In the remaining section I will provide the taxonomy of grammatical categories expressing modality I have designed and used in my analysis of data.

- **Grammatical elements expressing modality: verbal and non-verbal modal expressions**

As for the grammatical means used to express modality, I will distinguish between two main grammatical categories, verbal and non-verbal modal expressions. Within the former, besides the prototypical realization of modality by means of modal auxiliaries, other verbal resources that I will take into consideration are lexico-modal auxiliary verbs, certain lexical verbs and modal idioms. Regarding non-verbal modal elements, I will mainly focus on adverbials and adjective phrases. The following table, table 8, offers the taxonomy I have designed.

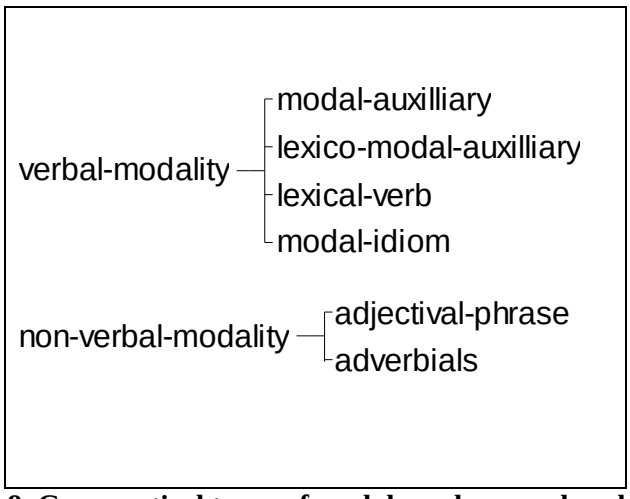


Table 8 Grammatical types of modal markers analyzed

Thus, in general, I will follow Halliday's model closely, since it constitutes a satisfactory account of modality (see section 2.4.1) but adapting it to suit my own taxonomy regarding the classes and subclasses of modality (my full network, with all the semantic and grammatical

categories considered is presented in Appendix 1). Furthermore, I will take into account some important points made by Nuyts in his analysis of modality (already summarized in section 2.4.2), which will be developed below.

- **Verbal modality:**

This section will present the various verbal modal categories, and the elements included within each, I have established as well as the reasons for their inclusion.

a) Modal auxiliaries

Concerning the "prototypical" category to encode modality in English, i.e. modal auxiliary verbs (Halliday, 1985; 1995; Greenbaum *et al*., 1985; Palmer; 2001; 2003; Aijmer, 2002; Halliday and Matthiessen, 2004), I will include what Greenbaum *et al*. (1985) label as "central modals" (see section 2.2.1. above), which I will here rename as *modal auxiliaries*. Hence, I will analyze the use of the modal verbs CAN, COULD, MAY, MIGHT, WILL, WOULD, SHALL, SHOULD and MUST. Even though I will only use their bare forms in capital letters for ease of presentation, I have included in my analysis not only their positive but also their negative polarity forms as well as their contracted forms.

In addition, I will include in this category the modal auxiliary OUGHT TO, which Greenbaum *et al* (1985) consider within "marginal modals". The reason is that it passes all the modal auxiliary tests discussed in section 2.2.1., that is to say, it does not take the –s form in the 3rd person singular of the present simple; it does not have non-finite forms such as gerund or infinitive; and, finally, it cannot co-occur with other members of the same class, since "*he will ought to tell me*" is clearly ungrammatical.

Regarding Nuyts' reanalysis of subjective/objective epistemic modality (Lyons, 1977) as subjective/intersubjective (Nuyts, 2001), (which I have explained in section 2.4.2.), I will not consider it here because it far exceeds our current goals. Nevertheless, I feel the clarification is

interesting and necessary, at least, on theoretical grounds, even though, for the time being, it will be of no consequence in my analysis of data.

b) Lexico-modal auxiliaries

Another element of the "marginal modals" class put forward by Greenbaum *et al* (in the same section 2.2.1.), which I will also include in my study, is the verb NEED. In particular, Greenbaum *et al*. (1985) talk about two different uses of the verb. One, as a modal auxiliary but whose use is "rare in both American English and British English" and, another, as a main verb, which is more common.

Moreover, they claim that, as a main verb, it is followed by to-infinitive and with inflected –s, -ing and past forms, whereas, as a modal auxiliary, it is followed by the bare infinitive and never appears in the inflected forms.

In this case, I will not distinguish between a modal use or a main verb use of NEED, because learners may make mistakes with the right complement or with the inflected forms. Therefore, I will include in my analysis all the occurrences of NEED as a verb, whenever it is followed by a clause, not by a noun phrase. And this is the reason why I consider it neither a modal auxiliary, nor a lexical verb, but something in between.

This means that instances such as *(1)* and *(2)* below:

(1) "I personally believe that persons ***need*** *more attention in childhood than in adulthood" (UAM/C7-1)*

(2) "Homosexual couples ***need*** *a law which regulates their situation" (UAM/C61-2)*

will not be considered as modal since they do not involve an evaluation over a proposition on the part of the speaker. Hence, I have manually eliminated all these uses from the modal elements automatically annotated by UAM CorpusTool as well as the occurrence of NEED as a noun.

On the contrary, examples like *(3)* and *(4)* have been kept for the same reasons:

(3) "adult people ***need*** *to know more about this and take care about this"* (UAM/ A196-1)

*(4) "they **need** to work as prostitutes to support a drug habit"* (UAM/C40-2)

Also, I have included some of the so-called *semi-auxiliary verbs* (Greenbaum *et al*, 1985) in the category of *lexico-modal auxiliaries* (whereas the remaining elements will appear in the class of *adjectival phrases* as I will explain below). The main reason to do this is that, given that my approach in this study is grammatical, I will group the various elements capable of expressing modality according to the grammatical class which they originally belong to, rather than to their functional category.

In this way, verbs or set verbal phrases, regarded as semi-auxiliaries by Greenbaum *et al* (1985), such as HAVE (GOT) TO, BE GOING TO, BE SUPPOSED TO, BE OBLIGED TO, BE REQUIRED TO, BE BOUND TO and BE ALLOWED TO, besides the above mentioned NEED, will be considered here as *lexico-modal auxiliaries*.

The criterion informing this decision is that all of them, except for NEED, are made up of at least one auxiliary verb, BE or HAVE, and, in most cases, also by a lexical verb which clearly expresses a modal meaning, regardless of the context where they occur.

Therefore, from a grammatical and semantic point of view, they are neither modal auxiliaries, since they do not fulfil the modal auxiliary criteria mentioned above (namely, not allowing –s forms, non-finite forms and co-occurrence) nor full lexical verbs but something in-between. Again, I will only present here their bare forms in capital letters although in my analysis of data I will include all their variations, such as contractions, negative and inflected forms (tense and person inflections) and so on. By the same token, all their occurrences outside the grammatical domain where they express modality (e.g. active voice in the case of the passive predicators) have been manually excluded.

As regards the remaining elements included by Greenbaum *et al.* (1985) within the semi-auxiliary class, such as BE ABLE TO, BE SURE TO, BE CERTAIN TO or BE LIKELY TO, they will be here placed in the adjectival phrase category. The principle underlying this decision

is, as I have already mentioned, grouping the elements capable of expressing modality according to the grammatical nature of their main constituent.

Thus, in this case the main element expressing modality is an adjective *–able, sure, certain* and *likely,* respectively- which, therefore, typically appears in copulative sentences. Moreover, since the main goal of my study is to examine the evolution in the quantity and variety of grammatical structures students use when encoding modality, the borders between the various grammatical categories must be clearly defined and delimited.

c) Certain lexical verbs: Mental State Predicates

Some lexical verbs will also be included within verbal modal expressions. As was explained in the literature review section when dealing with some relevant theoretical considerations suggested by Nuyts (2001), I will include in my analysis certain uses of a number of lexical verbs, more specifically, some mental state predicates.

I will consider as modal the following verbs, only when used in the first person singular of the present tense (following Nuyts, 2001): I THINK, I BELIEVE, I SUPPOSE, I GUESS, I FEEL, I FIND, I EXPECT, I KNOW, I RECKON, I CONSIDER, I INTEND and I DOUBT. Again, only those instances where the above mentioned verbs are followed by clauses, and which, hence, involve an evaluation of a proposition by the speaker, have been included.

Furthermore, Nuyts (2001) establishes a distinction between performative and descriptive epistemic expressions (see section 2.4.2. for a full account). The main difference between them lies in the fact that the performative class involves an evaluation and the speaker's actual commitment to it whereas the descriptive class entails the reporting of a qualification, either of the speaker or someone else, which does not imply the speaker's commitment at the time of speaking.

If we turn to our corpus, the following examples will clarify this dichotomy:

a) *"**I consider** this argument reasonable because **I think** that it is important to make the children understand..." (UAM/C8-3)*

b) *"Secondly, personally speaking, **I believe** that television is another factor which increases violence in children..." (UAM/C7-1)*

c) *"To me those classes were interesting but **most of people thought** they were unreliable..." (UAM/C62-3)other way" (UAM/A94-2)*

d) *"...under this BMI **the World Health Organization believes** it is underweight.." (UAM/C91-1),*

As should be evident, (a) and (b) differ from (c) and (d) in that the first two involve an evaluation on the part of the speaker as well as his current commitment to it whereas the last two merely report on someone else's thoughts or beliefs, without any kind of judgment or qualification on the speaker's part. Consequently, only performative modal expressions, such as a) and b), and their negation will be examined in this study. Mental state predicates will be excluded when the subject is not the first-person-singular pronoun and the tense is not the present tense.

Another point that Nuyts (2001) very convincingly makes refers to the ambiguous nature of mental state predicates (especially, if compared with adverbial or adjective phrases) because of their conflation of an epistemic and an evidential meaning. He offers a detailed explanation of the exact meaning of the main mental state predicates above mentioned, distinguishing two main groups. On the one hand, those mental state predicates with a primary epistemic meaning, such as "think", "believe" or "doubt" and, on the other, those where the evidential meaning is more prominent, like "know", "suppose", or "guess".

However, as I have suggested above, Nuyts ends up his discussion by concluding that "this combination of an evidential and an epistemic meaning is one of the most central characteristics of mental state predicates" (2001). Therefore, I have decided to include all the

mental state predicates mentioned above since, besides the evidential, the epistemic meaning is, to some degree and without doubt, present at the core of all of them[4].

As for those verbs such as "hear" and "see", or "appear" and "seem", which express a qualification on the part of the speaker clearly distinct from the epistemic evaluation, they have been excluded from my examination. All of these primarily and predominantly convey an evidential meaning, a semantic area briefly outlined in section 2.3.a) (see Chafe's schema), which lies beyond our object of analysis.

Equally, verbs such as "acknowledge", "concede" or "agree", which involve an implication of truth but not a judgment about the truth value of the proposition, will not be considered. Additionally, mental verbs such as "want", "wish", "hope", "understand", "wonder", "recommend", "support", "realize" will not be taken into consideration nor will reporting verbs such as "say", "claim", "suggest", "show", "argue" and so on. Evidential uses of verbs such as FEEL or FIND will also be excluded.

Finally, Nuyts' distinction between a qualificational and non-qualificational meaning of mental state predicates will also be followed here. Hence, only mental state predicates in grammatical environments expressing a qualificational meaning will be taken as modal. In particular, and as mentioned above, only mental state predicates followed by a that-clause complement, as in (a) below, or a reduced variant of it, or when used parenthetically, like in (b), will be included in my analysis:

a) *"**I think** that the number one of the books is "Harry Potter"." (UAM/C47-1)*

b) *"In conclusion, we have seen the four problems that **I think** are the most worrying these days" (UAM/A170-1).*

[4] A different matter is that a cline from more epistemic to more evidential mental state predicates could be established but, due to the nature and goals of this study, this task goes beyond our main concerns here.

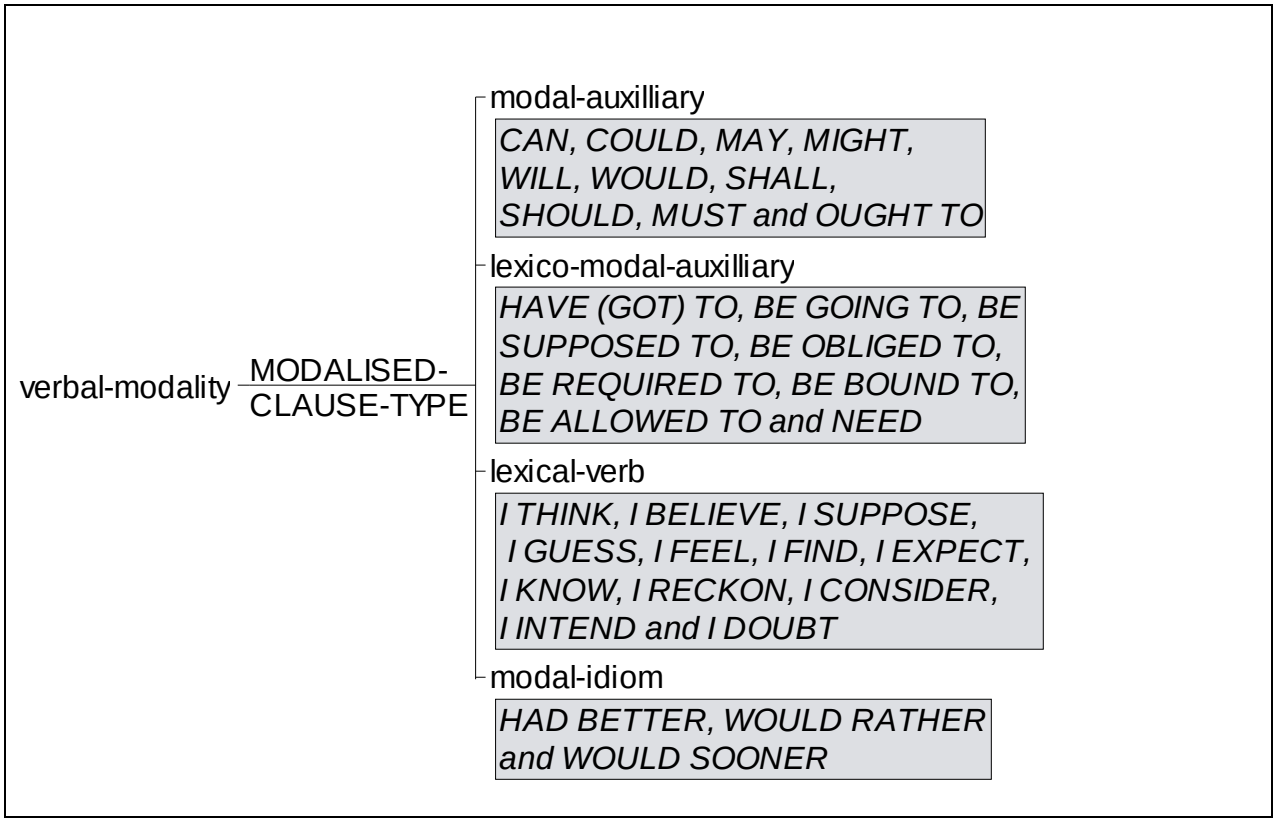
In contrast, other grammatical structures of mental state predicates, such as when followed by fixed prepositions, as in (c), have been excluded from my examination of data:

c) *"If I think about all the people who is true to Catholic Church, I could say that.." (UAM/A48-1)*

d) Modal idioms

Finally, another class mentioned by Greenbaum *et al* (1985) which will also be looked at is modal idioms. The prototypical elements within this category are HAD BETTER, WOULD RATHER and WOULD SOONER. As has been mentioned above, contracted forms have also been included in the analysis although, for ease of presentation, only the bare forms in capital letters are here reproduced.

In order to illustrate and summarize all the verbal elements expressing modality which will be part of this study, the following scheme has been designed:



➢ Non-verbal modality

Having explained the verbal categories expressing modality I will examine, I will now deal with the non-verbal elements I will analyze (i.e. adverbials and adjective phrases). One of the grammatical categories most frequently mentioned in the literature as capable of expressing modality, i.e. noun phrases, is problematic in my analysis of modality. In my approach, the crucial element to draw the borderline between modal and non-modal expressions is the speaker's evaluation of a given state of affairs. However, noun phrases do not appear to convey such an evaluation.

The following examples from WriCLE will illustrate this:

*a) "(…) parents could have the **possibility** of revising the contents of the educational programs (…)" (C45-2)*

*b) "everyone should have the **chance** to choose" (UAM/C22-2)*

*c) "they feel the **necessity** to be fully aware of what sex is" (UAM/C52-2)*

As should be evident, the nouns in italics do not express the speaker's evaluation, either of the likelihood of the event expressed by the proposition in (a) or (b), or of the necessity of the state of affairs conveyed by the proposition in (c). These three nouns ("possibility", "chance" and "necessity") refer to something much more definite and tangible than an evaluative judgment on the part of the speaker. Nominalized modality becomes, thus, something else which does not fit in my definition of modality.

Furthermore, there is a clear difference in meaning between (a) or (c) and (d) or (e),

*d) "..public transport is **possibly** safer because…" (A47-1)*

*e) "..since jurors do not **necessarily** possess a wide knowledge.." (C44-1)*

since the latter set of examples clearly implies the judgment or evaluation on the part of the speaker, which lends the modal meaning. This is the reason why, given the difficulty to clearly

assess the semantic nature of nominalized modality, I have excluded the grammatical category of noun phrases from my analysis.

e) Adverbials

In order to decide which adverbials should be treated as expressions of modality, I have basically followed Greenbaum (1969) and Greenbaum *et al.* (1985). Even though, when dealing with lexico-modal auxiliaries, I grouped tokens according to the grammatical nature of their heads, in the case of adverbials, a functional analysis appears to be the most useful. However, it is important to point out that "when adverbs, prepositional phrases and other structures are functioning as part of an element in the sentence structure, they cannot be regarded as adverbials in terms of the sentence in question" (Greenbaum *et al*., 1985:501).

Greenbaum *et al.* (1985) mention four grammatical functions that adverbials can perform in the sentence, which are adjuncts, subjuncts, disjuncts and conjuncts. In particular, the main features they provide to distinguish between them are the following:

First, adjuncts and subjuncts are "relatively integrated within the structure of the clause". On the contrary, disjuncts and conjuncts "have a more peripheral relation in the sentence". Second, from a semantic point of view, disjuncts "express an evaluation of what is being said either with respect to the form of the communication or to its meaning". Conjuncts differ in that they express "the speaker's assessment of the relation between two linguistic units" (1985:631).

Therefore, the adverbials I will consider here are some of those Greenbaum *et al.* include within the broad functional categories of subjuncts and disjuncts since some of their subclasses imply either an evaluation on the part of the speaker or the comment that what is being said is true. This class of adverbials, roughly speaking, corresponds with Halliday's notion of adjuncts.

Concerning subjuncts, Greenbaum *et al*. define them as "adverbials which have, to a greater or lesser degree, a subordinate role in comparison with other clause elements" (1985:566-67). Moreover, they provide a number of subclasses, depending on whether this "subordinate role" applies to the whole clause, labelled as "wide orientation", or to an individual clause element, known as "narrow orientation" (1985:567).

In particular, I will focus on a specific subtype of "narrow orientation" subjuncts, namely, emphasizers, which, in turn, can be broken down into two main types:

a) GROUP A: expresses the comment that what is being said is true,

actually, **certainly**, clearly, **definitely**, **indeed**, obviously, plainly, really, **surely**, **for certain, for sure, of course**[5].

b) GROUP B: conveys the speaker's assertion that his words are the unvarnished truth,

frankly, honestly, literally, simply, fairly (BrE), just (1985:583).

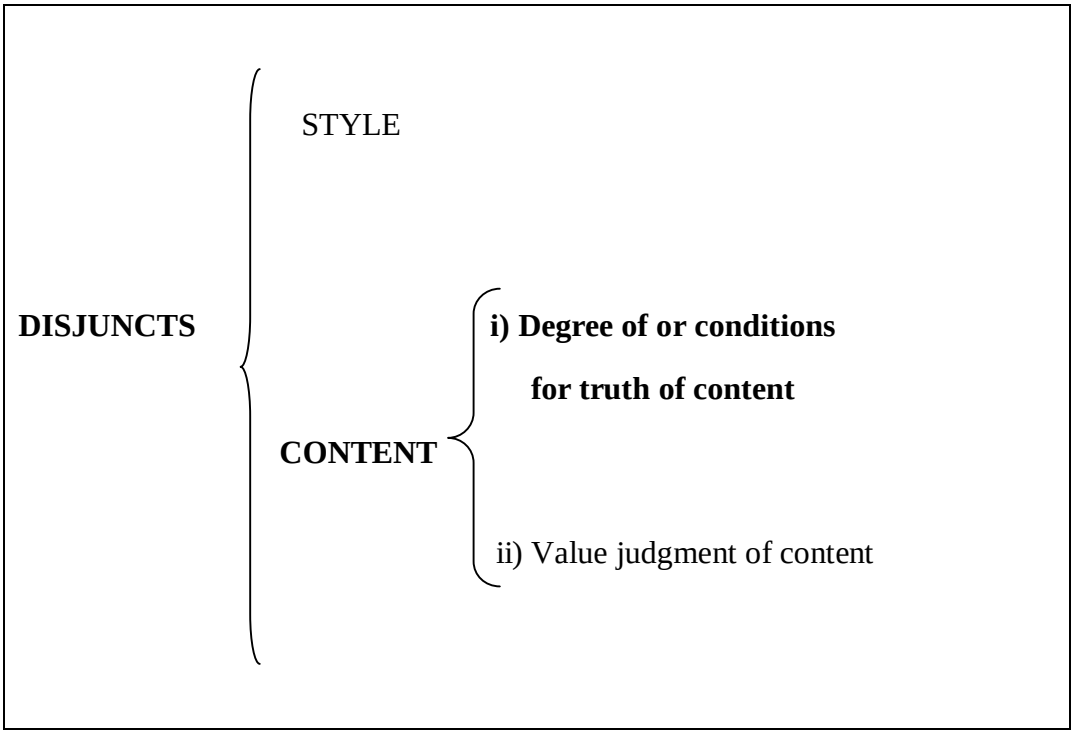
I will not include this last group since they express a semantic notion different from modality. As for those adverbials included within group A, I will only analyze the ones that I have highlighted in bold because they fit in our notion of modality, while the others express evidentiality. To the list of included items, I have also added **without doubt**.

Regarding disjuncts, Greenbaum *et al* (1985) distinguish between two main classes, *style* disjuncts, on the one hand, and *content* disjuncts, on the other. The former convey "the speaker's comment on the style and form of what he is saying, defining in some way under what conditions he is speaking as the "authority" for the utterance". On the contrary, content disjuncts

[5]In bold I have highlighted the only elements or categories I will consider in my analysis.

make “observations on the actual content of the utterance and its truth conditions” (1985:615). Therefore, content disjuncts will be our target here.

In turn, and depending on the type of observation or judgment on the state of affairs the speaker makes, content disjuncts can be of two main types: (i) **degree of or conditions for truth of content** and (ii) value judgment of content. The following scheme reflects and summarizes the different types of disjuncts distinguished by Greenbaum *et al.* (1985:615):



Main classes of disjuncts (adapted from Greenbaum *et al*, 1985:615)

Within “degree of or conditions for truth of content” disjuncts, Greenbaum *et al.* point out that adverbs falling in this type constitute almost a closed class and establish three main groups (1985:620):

a) Adverbs that express ***conviction***, either as a direct claim (eg. *undeniably*) or as an appeal to general perception (eg. *evidently*):

Admittedly, *assuredly (rare, formal)*, *avowedly (formal)*, **certainly,** *decidedly (rare, formal)*, **definitely,** *incontestably (rare, formal)*, *incontrovertibly (formal)*, **indeed**, *indisputably (formal)*, *indubitably (formal)*, **surely,** *unarguably (formal)*, *undeniably*, **undoubtedly**, *unquestionably*, *clearly*, *evidently*, *manifestly (formal)*, *obviously*, *patently (formal)*, *plainly*

Among them, only those adverbs expressing conviction as a direct claim will be analyzed, since the other type –those encoding conviction as an appeal to general perception- pertains to the domain of evidentiality and, therefore, will not be considered in my analysis. To make it clear, I have highlighted in green and italics the category expressing evidentiality whereas the ones to be examined here, appear in black bold letters.

Also, I have excluded those adverbs which are rare or pertain to a formal register –information which is offered in brackets-, after making sure there are no instances of them in WriCLE.

b) Adverbs that express some ***degree of doubt***:

Allegedly, **arguably**, *apparently*, **conceivably, doubtless, doubtlessly, likely, maybe, perhaps, possibly, presumably,** *purportedly (formal)*, *reportedly*, *reputedly (formal)*, *seemingly (formal)*, **supposedly**

c) Adverbs that state the sense in which the speaker judges what he says to be true or false, e.g. a*ctually, really, factually, ideally, theoretically, basically, essentially, fundamentally, etc.*

As should be evident, none of the adverbs in this last subtype will be included in the study, since, strictly speaking, they do not involve the speaker's evaluation of the likelihood of the event, neither of the necessity, moral acceptability or advisability of a certain state of affairs, nor of the ability, intention or willingness of someone to do something. This subclass conveys a different type of judgment or evaluation, which does not fit in our definition of modality.

Another point that should be mentioned is that Greenbaum *et al.* (1985) include in this category the classes of *hypothetical clauses* "on which closely reasoned discourse depends, such as concessive, conditional, reason, and other adverbial clauses" (1985:622). However, I will not take into account the category of clauses for technical reasons, since the parser of UAM CorpusTool cannot annotate as modal such long strings of words.

To summarize, within the grammatical category of adverbials I have included the following elements:

MAYBE, PERHAPS, POSSIBLY, DEFINITELY, CERTAINLY, SUPPOSEDLY, SURELY, UNDOUBTEDLY, DOUBTLESS, DOUBTLESSLY, LIKELY, PRESUMABLY, ARGUABLY, CONCEIVABLY, INDEED, FOR SURE, FOR CERTAIN, OF COURSE and WITHOUT DOUBT, PROBABLY and NECESSARILY.

An important aspect I should mention is that, whenever problems have arisen in the process of establishing my inclusion and exclusion criteria, the bottom line has been the distinction between the realis and irrealis domains. This is why adverbs that clearly pertain to realis, such as *actually* or *really,* have been discarded.

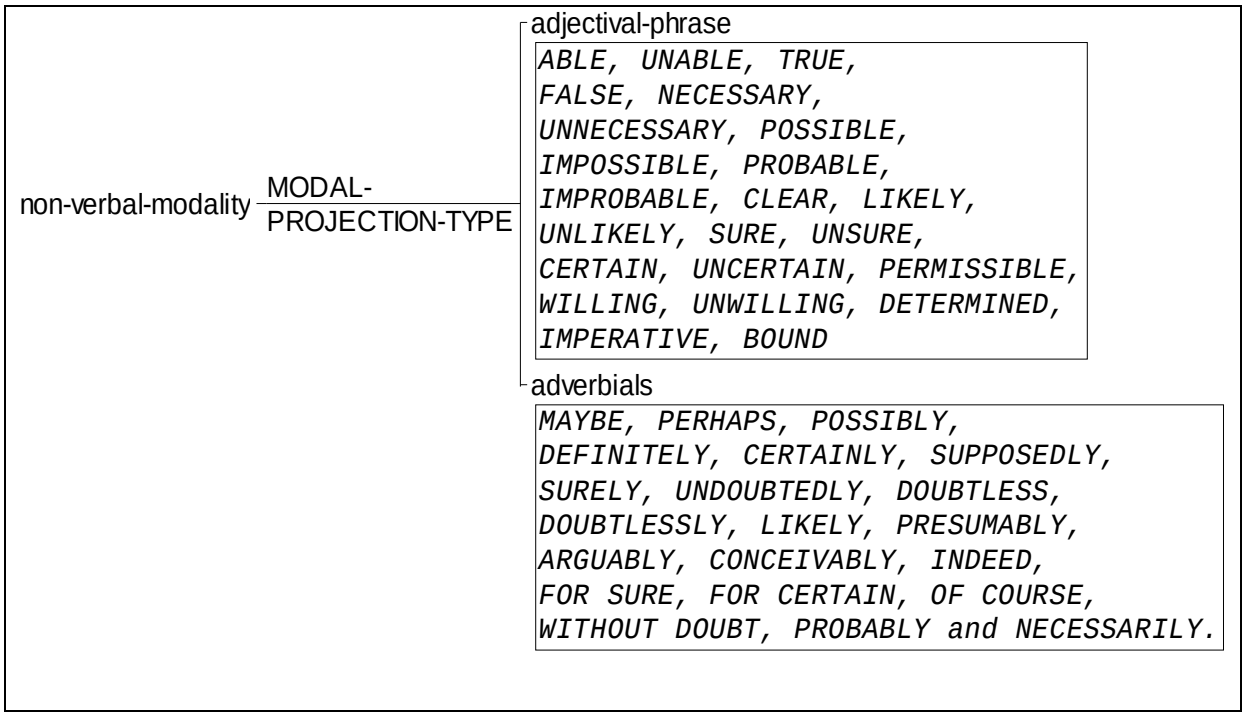
f) Adjective phrases:

As for adjectives expressing modality I have included the following items:

ABLE, UNABLE, TRUE, FALSE, NECESSARY, UNNECESSARY, POSSIBLE, IMPOSSIBLE, PROBABLE, IMPROBABLE, CLEAR, LIKELY, UNLIKELY, SURE, UNSURE, CERTAIN, UNCERTAIN, PERMISSIBLE, WILLING, UNWILLING, DETERMINED, IMPERATIVE, BOUND.

As I mentioned when dealing with modal auxiliary verbs, and can be deduced from the list above, all the negative counterparts have been included where possible, either with the negative prefix -eg. improbable, unnecessary, etc.- or in the form NOT + positive adjective. In order to do this, I have manually eliminated or included the target elements just mentioned by means of UAM CorpusTool (O´Donnell, 2008), whose possibilities and mechanics I have described in the previous section.

To summarize and illustrate the elements I will analyze within non-verbal modality, the following scheme has been created:



4. RESULTS

Before putting forward the data obtained by means of UAM CorpusTool, I should mention that the unit of analysis I have considered in my examination of the data is the clause. In my

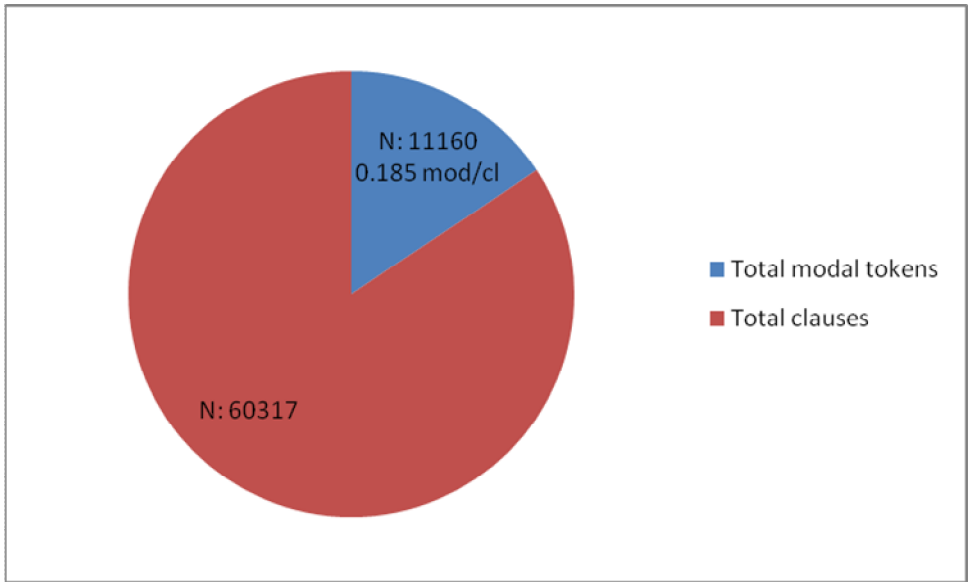
view, this offers a more reliable insight into the proportion of modality used across proficiency levels for various reasons.

First, students tend to write more as their linguistic competence increases, which in practical terms means either producing longer clauses or a larger number of them. In this way, if we looked at the proportion of modality in relation to the total number of words written per clause, we would face the risk of obtaining biased results, since the longer the clause, the lower the proportion of modality employed. Second, in my theoretical framework, modality is expressed at the level of the clause, not the sentence or the phrase, as modal markers have a scope, or rather, provide the speaker's evaluation over the proposition expressed in the clause.

However, I should point out that, after providing the proportions of modality per clause, I will also offer the proportions per total modal markers used. That is, what proportion of all modal markers in the essay is made up by each particular grammatical class across proficiency levels. This approach ignores the relative high or low use of modality per clause, but does reveal the relative importance of each modal type to students at the different proficiency levels. This approach is more useful when looking at the breakdown of the occurrence of the main classes –namely, verbal and non-verbal- and subclasses of modal markers –modal auxiliaries, lexico-modal verbs and certain lexical verbs within verbal modality, on the one hand, and the non-verbal categories of adjectival phrases and adverbials, on the other.

The main reason underlying this is that it allows us to observe which particular grammatical structures students prefer when expressing modality at each level of proficiency and also, to compare and contrast their evolution across proficiency levels. Thus, each indicator –percentages per clause and percentages per total modal markers used- serves different functions and, consequently, both will be included here and taken into consideration in my analysis of data.

To begin with, it should be mentioned that the section of WriCLE used in this study is made up of a total number of 60,317 clauses (finite or nonfinite). The corpus shows 11,160 elements belonging to any of the grammatical categories established in this study as expressing modality, which means that we have on average 0.185 modal markers per clause. If we were to assume there is only one modal marker per clause, this would mean 18.5% of the clauses were modalised. In fact, some clauses contain more than one modal token, so the actual percentage is closer to 17.9%. This means that there is an average of almost one modal marker per five clauses.



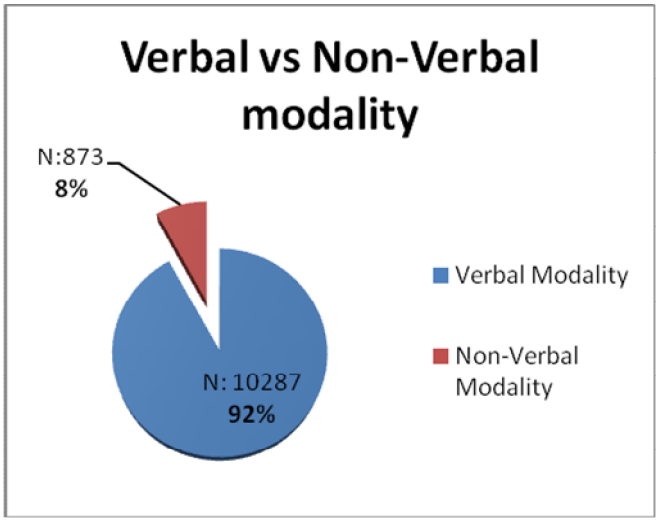
Graph 4: Total modal tokens, total number of clauses and amount of modality per clause

Concerning the two main types of grammatical elements that can be used to express modality, namely, verbal and non-verbal modal markers, the students overwhelmingly use verbal resources to express modality. If we look at raw numbers, we can distinguish 10,287 verbal modal elements as opposed to only 873 non-verbal modal tokens, in total. Table 9 below details the total number of occurrences of each verbal and non-verbal class in the section of WriCLE analyzed.

MODAL MARKERS	N
Verbal-modality	10287
Modal-auxilliary	7929
Lexico-modal-auxilliary	996
Lexical-verb	1362
Non-verbal-modality	873
Adjectival-phrase	504
Adverbial-phrase	369
TOTAL	11160

Table 9 Raw numbers of total modal markers

This means that verbal elements expressing modality represent 92% of the total number of modal markers employed whereas non-verbal resources merely constitute 8%, as Graph 5 below illustrates.



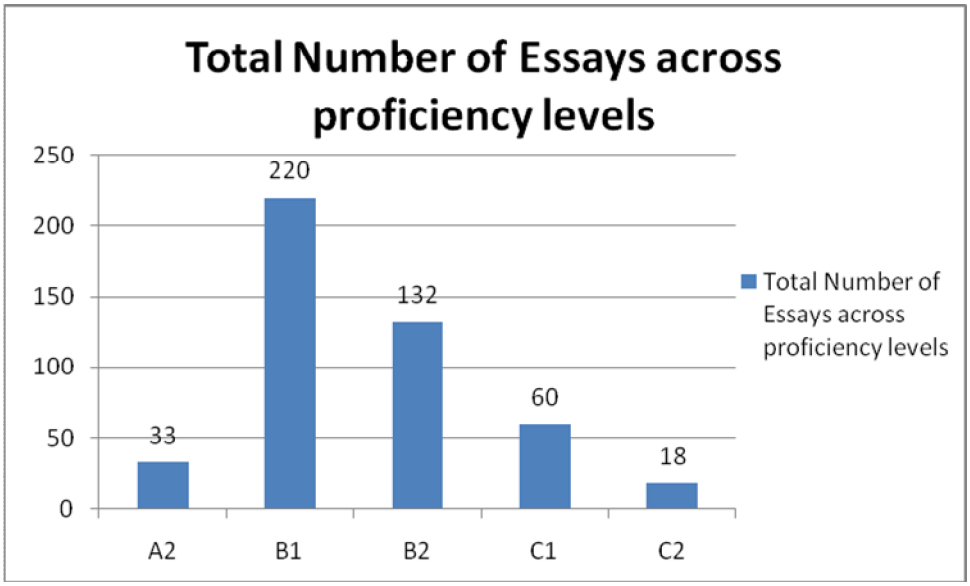
Graph 5: Verbal vs. Non-verbal modal elements used

Regarding the total number of modal markers used across proficiency levels, A2 students produced 640 modal tokens in total; B1 employed 4,982 modal markers; B2 used 3,384 modal elements; C1 displayed 1,697 modal resources; and C2 utilized 457 grammatical elements expressing modality (see Table 10 below for details). However, it must be noted that the section of WriCLE I have used has not been randomly selected, so the number of essays across proficiency levels included in this study is, by no means, even.

	A2	B1	B2	C1	C2
verbal-modality	570 (89%)	4653 (93%)	3112 (92%)	1546 (91%)	406 (89%)
lexico-modal-auxilliary	58 (9%)	464 (9%)	333 (10%)	111 (7%)	30 (7%)
modal-auxilliary	441 (69%)	3580 (72%)	2333 (69%)	1228 (72%)	347 (76%)
lexical-verb	71 (11%)	609 (12%)	446 (13%)	207 (12%)	29 (6%)
non-verbal-modality	70 (11%)	329 (7%)	272 (8%)	151 (9%)	51 (11%)
adjectival-phrase	54 (8%)	192 (4%)	153 (5%)	83 (5%)	22 (5%)
adverbial-phrase	16 (3%)	137 (3%)	119 (4%)	68 (4%)	29 (6%)
Total	1280	9964	6768	3394	914

Table 10: General distribution of grammatical categories of modality across proficiency levels

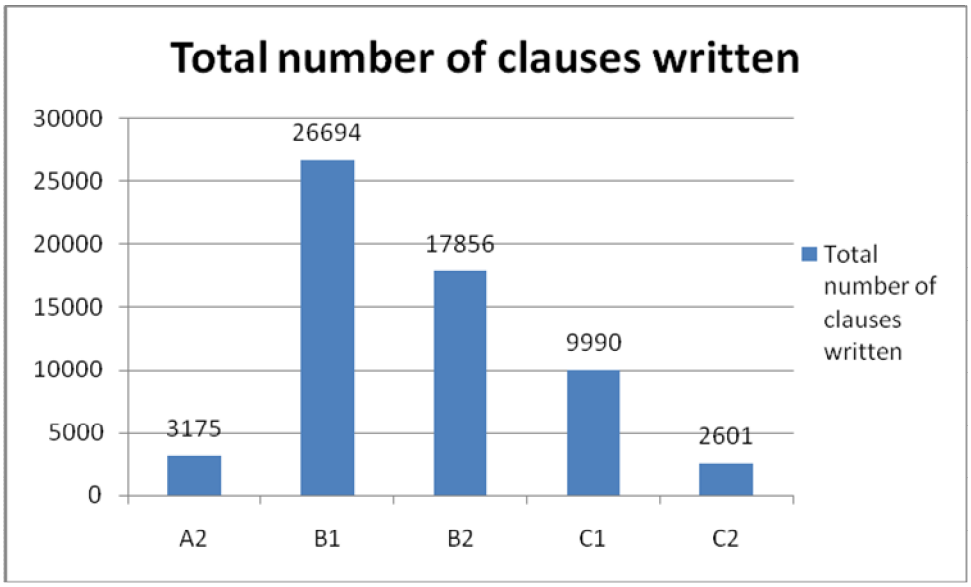
The following graph (see Graph 6 below) reflects the ample difference in the number of essays studied across proficiency levels in raw numbers.



Graph 6: Total number of essays analyzed across proficiency levels

Moreover, besides the difference in the number of essays across proficiency levels included in the analysis, another important distinction to take into account is the number of clauses written at each stage of proficiency, since this is the unit of analysis used. Thus, A2 students wrote a total number of 3,175 clauses; B1 learners produced 26,694 clauses; B2 employed 17,857 clauses; C1 made use of 9,990 clauses; and, finally, C2 students composed 2,601 clauses. Graph 7 below visually represents these differences.

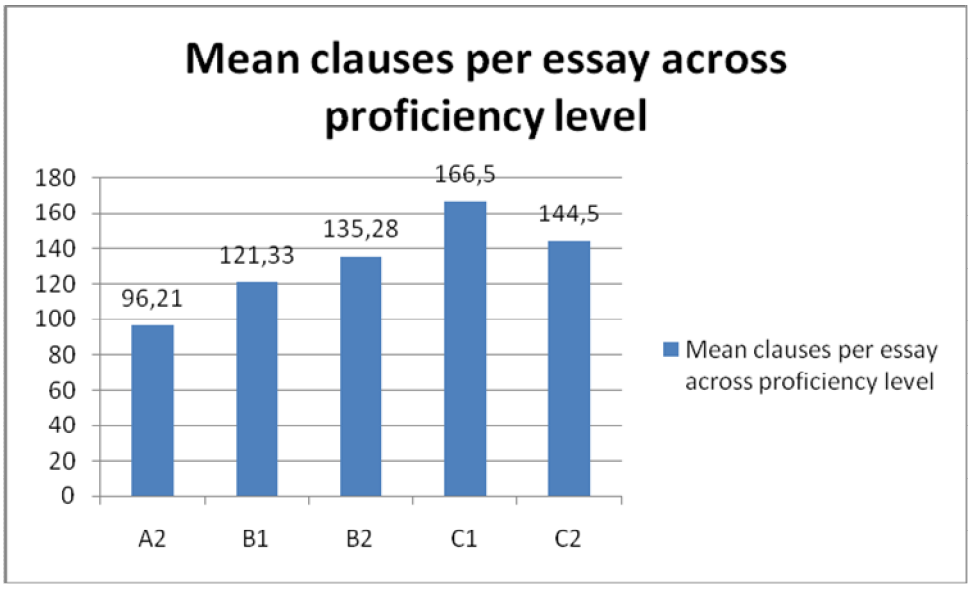
Obviously, the main reason behind this wide difference in the number of clauses is the inclusion of a very divergent number of essays across proficiency levels. These two elements, number of clauses and number of essays are inextricably related, as the almost identical appearance of Graph 6 above, with the number of essays, and Graph 7 below, with the number of clauses across proficiency levels, evinces. However, another element to take into consideration is the already mentioned tendency among learners to produce more clauses, that is to say, to write more as they improve their linguistic competence.



Graph 7: Total number of clauses written across proficiency levels

Thus, if we look at the mean number of clauses per essay produced by the students at each level of proficiency (see Graph 8 below), we will realize that, on average, the higher proficiency level the students have, the more clauses per essay they write and, conversely, the lower their linguistic level, the fewer clauses produced. In this way, the more advanced students, namely, C1 and C2, produced the largest mean number of clauses per essay, with 166.5 and 144.5 clauses respectively. It is remarkable that C1 students wrote more clauses than the more advanced C2 learners.

As could be expected, C1 and C2 were closely followed by B2 students, who produced a mean number of 135.28 clauses per essay. Obviously, the students who wrote a smaller number of clauses were the ones with the lowest proficiency levels, that is to say, B1 students, with an average of 121.33 clauses per essay, and A2 students, with merely 96.21 clauses per essay on average.

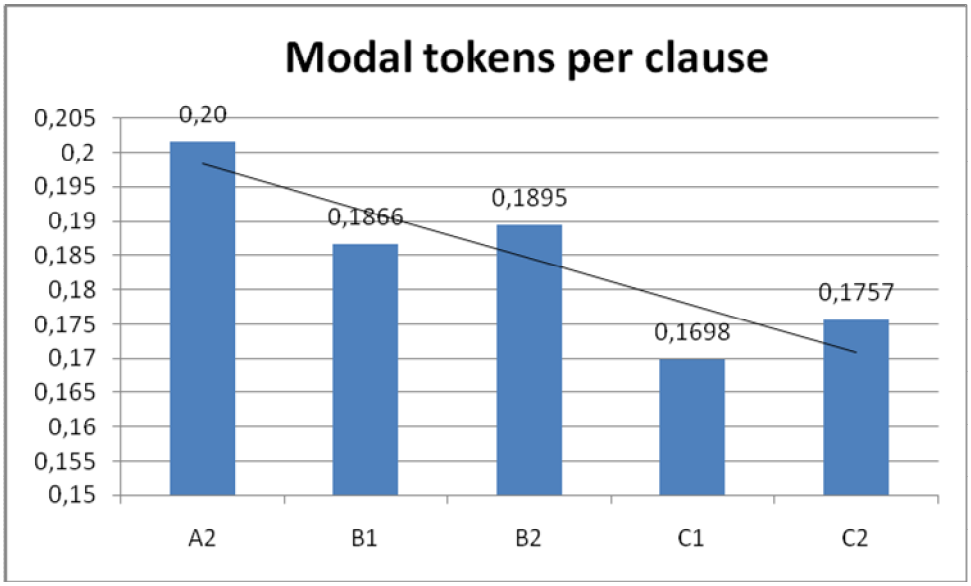


Graph 8: Mean number of clauses per essay written across proficiency levels

4.1. Modality per clause across proficiency levels

In order to have an accurate idea of the degree of modality used by students across proficiency levels, instead of looking at raw numbers, I will consider the number of modal tokens employed per clause. As Graph 9 below indicates, A2 students display the highest proportion of modality, with 0.20 modal tokens per clause. These learners are closely followed by B2 and B1 students, with 0.1895 and 0.1866 respectively. Finally, contrary to my expectations, the more advanced students, namely, C2 and C1, show the lowest occurrence of

modal tokens per clause, with 0.1757 and 0.1698 respectively. Thus, we can distinguish a declining use of modal markers as the students' linguistic competence improves.



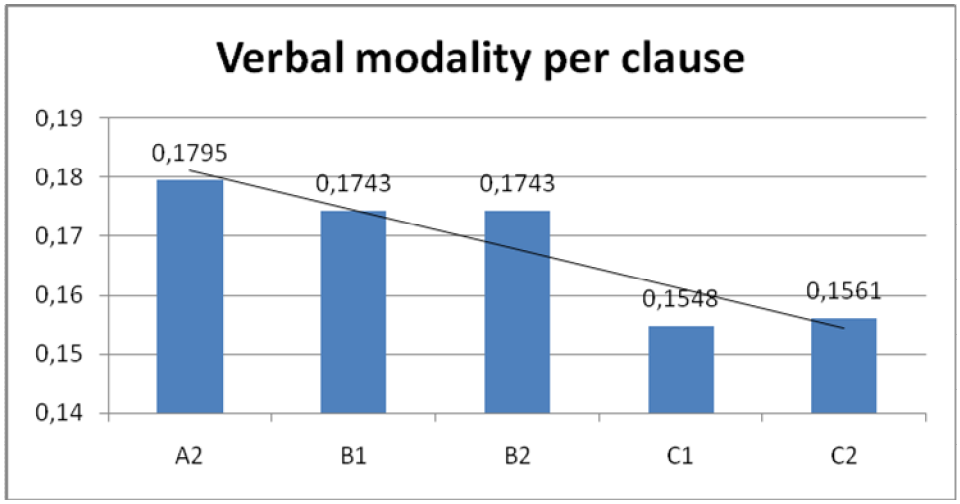
Graph 9: Proportion of total modal tokens per clause across proficiency levels

4.2. Verbal and non-verbal modality per clause across proficiency levels

I stated above that if we look at the dichotomy verbal vs. non-verbal modal markers, the learners, regardless of their proficiency level, prefer verbal modality by far, which represents 92% of the total modal markers employed. Regarding their use across proficiency levels, I will first examine and review the use of verbal modal markers and then, non-verbal modal resources.

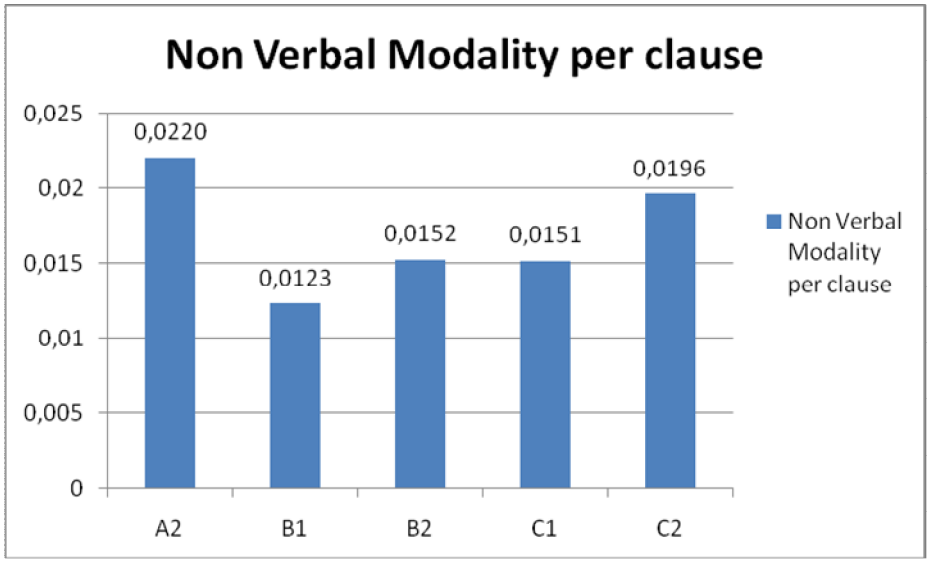
If we look at the proportion of verbal modal elements per clause displayed across proficiency levels (see Graph 10 below), A2 students show the highest figure, with 0.1795 instances of verbal modality per clause, which is very similar to the usage observed in B1 and B2 essays, 0.1743 in both cases. These figures clearly contrast with the lower modality usage among advanced students, with almost identical proportions (0.1561 in the case of C2 students and

0.1548 for C1 students). Thus, there is a decreasing use of verbal modal elements as the students' linguistic competence improves (which is visually reflected by the line in Graph 10).



Graph 10: Proportion of verbal modal elements per clause across proficiency levels

Regarding non-verbal modality, in spite of the low percentages, there is a tendency to increase its use as the students' level of proficiency rises, with the striking exception of A2 students (see Graph 11 below). Students at this proficiency level show 0.0220 of non-verbal modal elements per clause, which, contrary to my expectations, constitutes the highest figure among all proficiency levels. The rest of proficiency levels display a clear rising pattern, after a fall at B1, which displays the lowest usage, 0.0123. It is followed by B2 and C1 learners, with an almost identical figure (0.0152 and 0.0151, respectively). C2 students yield the highest percentage (0.0196).



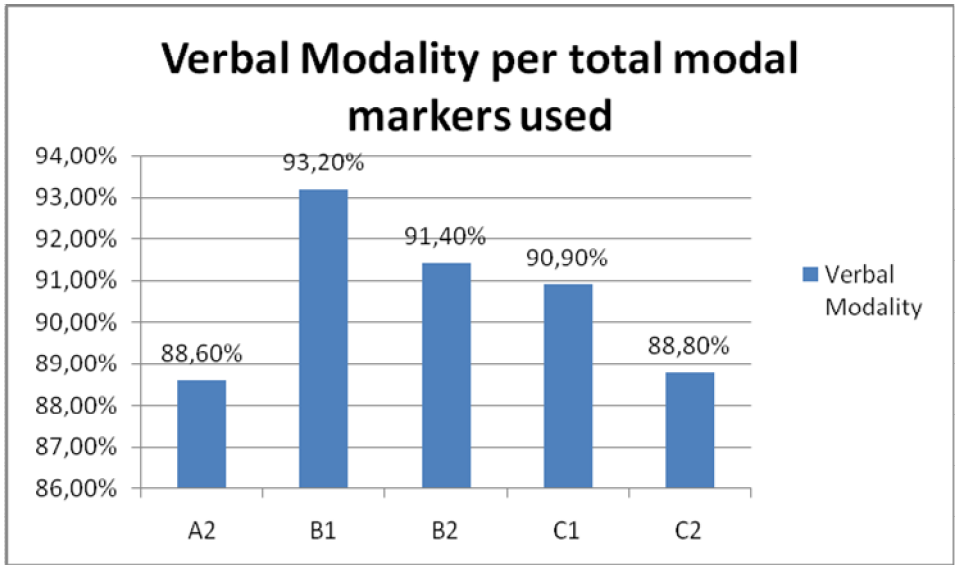
Graph 11: Proportion of non-verbal modality per clause across proficiency levels

As a result of this, we can safely state that, on average, low-level-proficiency students tend to use more verbal modality than students with a higher level of English. Conversely, the more advanced students make use of more non-verbal modal elements than lower-proficient learners, with the notable exception of A2 students. Nevertheless, it is worth mentioning the ample difference between the proportions of verbal and non-verbal modal markers displayed at any proficiency level, which can be observed by comparing Graphs 10 and 11.

4.3. Verbal and non-verbal modality per total modal markers used across proficiency levels

As the amount of modality per clause widely differs across proficiency levels, I will also offer the proportion of verbal and non-verbal modality used at each language proficiency level in relation to the total amount of modal markers employed. This is important to see which type of modal markers students prefer at the different stages of their learning process. Also, it will help to confirm or discard the tendencies mentioned just above.

Graph 12 shows the changing proportion of verbal modality markers out of all modality markers as proficiency rises. It shows a declining tendency, with the only exception of A2 students, who clearly show a lower percentage (88.60%) especially in comparison to B1 learners (with 93.20% they represent the highest figure). Thus, we can observe the more evident descending pattern in the use of verbal means to express modality in the light of the total modal markers employed.

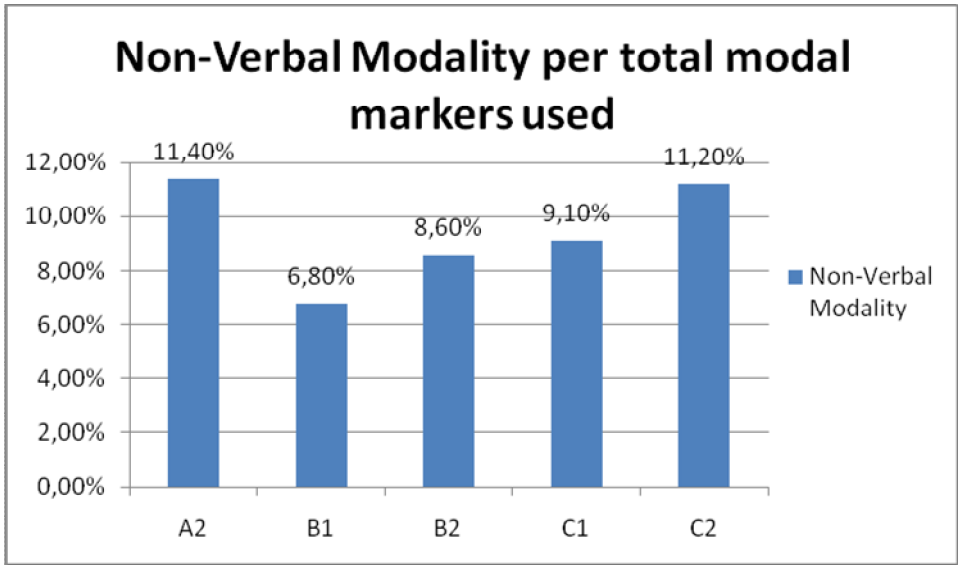


Graph 12: Verbal modal elements per total modal markers used

As for non-verbal means to express modality, the same falling-rising pattern observed when looking at non-verbal modality per clause (Graph 11) can be seen when looking at their proportion in the light of the total number of modal markers used (see Graph 13 below). As Graph 13 illustrates, there is a clear fall in the use of non-verbal modal elements among B1 students, since there is a difference of almost 5% in its usage between A2 (with an average of 11.40%) and B1 students (6.80%).

However, by B1 the usage of non-verbal modality starts rising, as it is clear from the figures: 8.60% at B2, 9.10% at C1, and 11.20% at C2 (an almost identical proportion to A2 students,

11.40%). Therefore, the rising and falling patterns observed respectively when looking at the amount of verbal and non-verbal modal means per clause are corroborated by their proportions in relation to the total number of modal markers used. A2 students constitute in both cases a notable exception to the rule, which could be due to experimental error, since this is one of the least representative proficiency levels in the study, as it merely contains 33 essays, and these are generally on the short side.



Graph 13: Non-verbal modal elements per total modal markers used

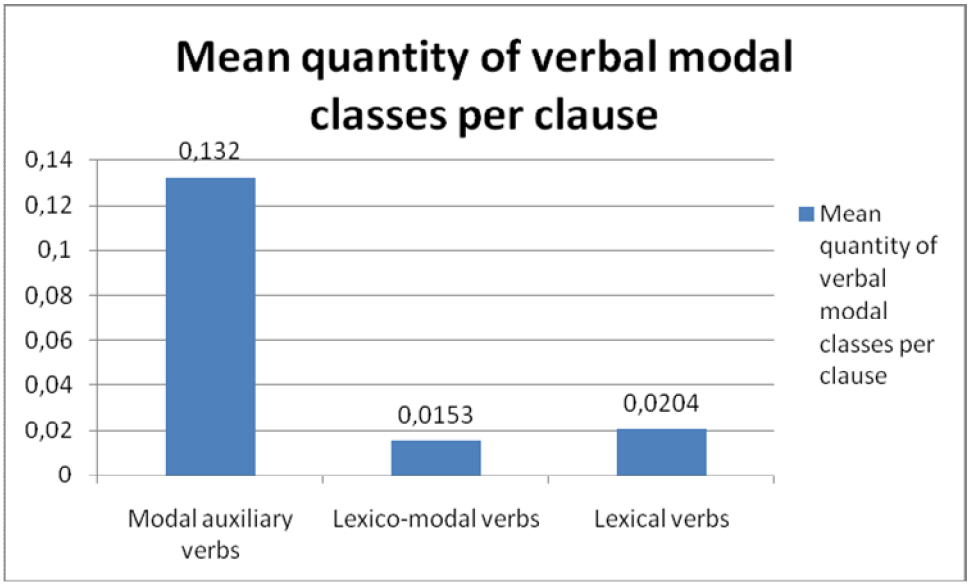
After having provided the raw numbers and percentages of modal markers used per clause, in total and across proficiency levels, and having also distinguished between verbal and non-verbal modal elements used per clause and per total modal markers employed, at the different stages of the learning process, I will now provide a detailed account of the proportion and distribution of the different subtypes of verbal and non-verbal modal markers employed by students at their various levels of linguistic competence.

4.4. Proportion of each modal category per clause:

▪ Verbal categories:

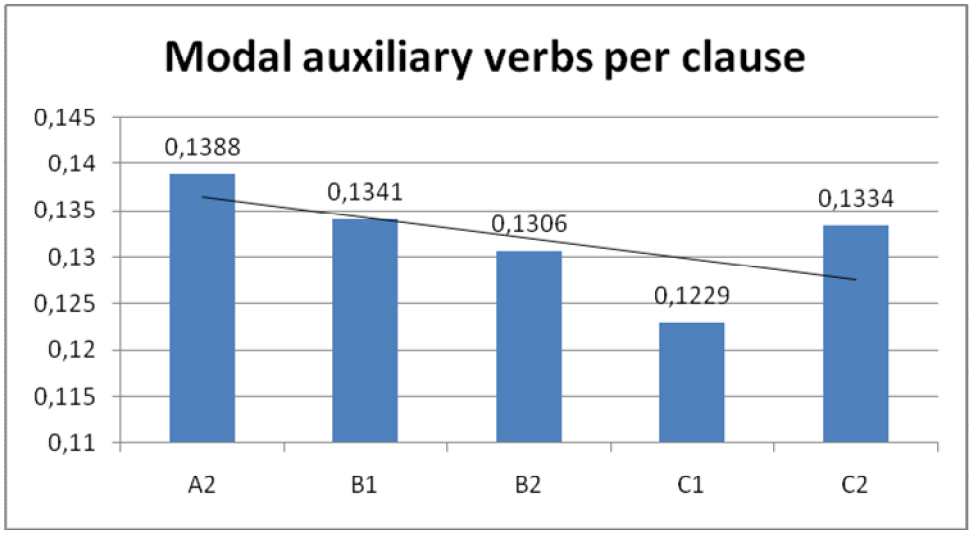
As I explained in the method section (section 3.2.) within the verbal means capable of expressing modality, I have included the following categories in my examination of the data: modal auxiliaries, lexico-modal verbs, certain uses of some lexical verbs and modal idioms. In this section I will offer the proportion of all these verbal classes per clause as well as the type of elements most and least frequently used within each verbal category –in general and at the different stages of the learning process- in order to establish some patterns of usage and general tendencies.

As can be observed in Graph 14 below, the most widely used verbal category, regardless of the students' linguistic competence, is modal auxiliaries –the so-called "prototypical realization of modality" in the literature. On average, there are 0.1320 modal auxiliary verbs per clause in all the essays examined. The other two verbal classes used by students are lexico-modal verbs and lexical verbs, since there is no instance of modal idioms in the whole corpus. However, I should highlight the great quantitative difference between the use of modal auxiliaries and the other two verbal classes. If regarding lexical verbs, the learners display such a meagre mean quantity of 0.0204 per clause, when it comes to lexico-modal verbs, they even show an inferior mean usage of 0.0153 per clause.



Graph 14: Mean quantity of verbal modal classes per clause in all the essays

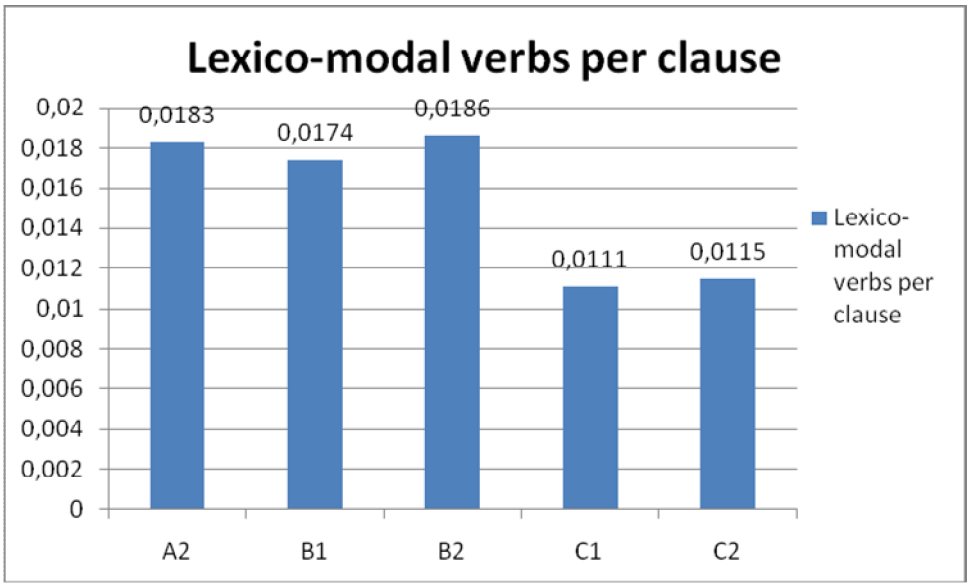
As regards to modal auxiliaries employed per clause across proficiency levels, there is a decline in their usage, except for C2 students, who constitute a clear exception to the general tendency. As can be seen in Graph 15, this declining tendency ranges from 0.1388 modal auxiliary verbs per clause at A2 level to 0.1229 at C1 level.



Graph 15: Modal auxiliary verbs per clause across proficiency levels

However, as I have mentioned above, C2 students show an increase in the use of modal auxiliary verbs, with 0.1334, almost identical to B1 students. As I explained above in relation to A2 students, this linguistic level is the least representative of all, as merely 18 essays within the corpus analyzed belong to C2 students. Therefore, these results could be due to experimental error as well.

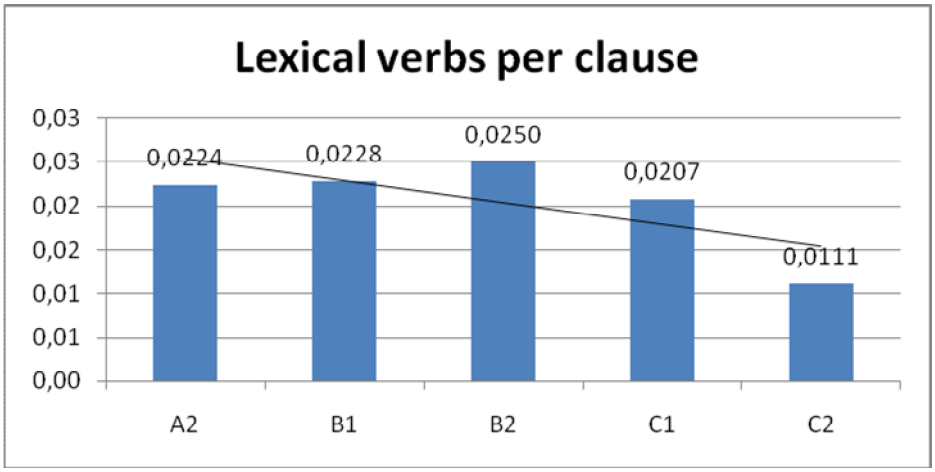
Regarding lexico-modal verbs (see Graph 16 below), there is also a decreasing tendency in their use as students' proficiency level approaches the more advanced levels, although in the intermediate levels, B1 and B2, their use is quite similar to A2. If we look at the figures, A2 students employ 0.0183 of lexico-modal verbs per clause, B1 and B2 students display similar figures, 0.0174 and 0.0186, respectively, while C1 and C2 students' usage of lexico-modals drop down to 0.0111 and 0.0115, respectively.



Graph 16: Amount of lexico-modal verbs per clause across proficiency levels

Concerning the lexical verbs which I have included in my analysis, there is a slight increase from 0.0224 per clause, displayed by A2 students, to 0.0250 employed by B2 learners, through 0.0228 used by B1 students. But the most proficient learners, that is to say, C1 and C2,

show a clear decrease in the use of lexical verbs to express modality per clause, with 0.0207 and 0.0111, respectively, as is shown in Graph 17 below.



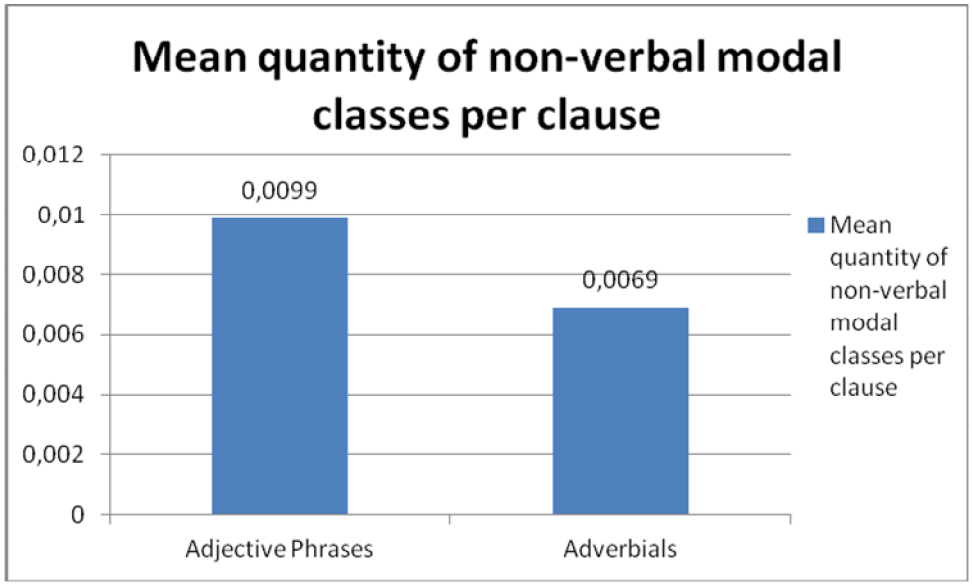
Graph 17: Amount of lexical verbs per clause across proficiency levels

To sum up, within the verbal categories, there is a slightly decreasing use of modal auxiliary verbs as the students' proficiency level rises, with the notable exception of C2 students, whose use of this verbal modal class experiences an increase. As for lexico-modal elements and lexical verbs, both categories display a similar pattern of use, that is to say, a slight increase, reaching the highest peak at B2, and then a fall in their use by C1 and C2 students.

Also, it is noticeable that the use of modal auxiliaries is quite stable across proficiency levels, since the biggest difference between the highest and the lowest figure only represents 0.0160, taking into account that on average they use 0.1320 per clause. Regarding the other two verbal categories, the difference between the highest and the lowest figures is even smaller but, given the fact that their total amount is much lower, they imply a more relevant difference. For instance, in the case of lexical verbs, the difference between the highest usage, 0.025, and the lowest, 0.0111, involves a drop of more than half the highest quantity, which is, thus, much more remarkable.

- **Non-verbal categories:**

With regard to non-verbal modality, the main categories I have included in my analysis are adjective phrases and adverbials (see section 3.2 for a full account). Graph 18 below displays the average usage of both non-verbal modal categories per clause in the whole section of WriCLE analyzed. As the figures show any of them is much less frequent, in quantitative terms, than any of the verbal classes. Students, regardless of their level of proficiency, only use 0.0099 adjective phrases per clause, while the amount of adverbials per clause is even inferior, 0.0069 per clause. I will now look at the usage of these non-verbal classes per clause across proficiency levels to see their evolving patterns.

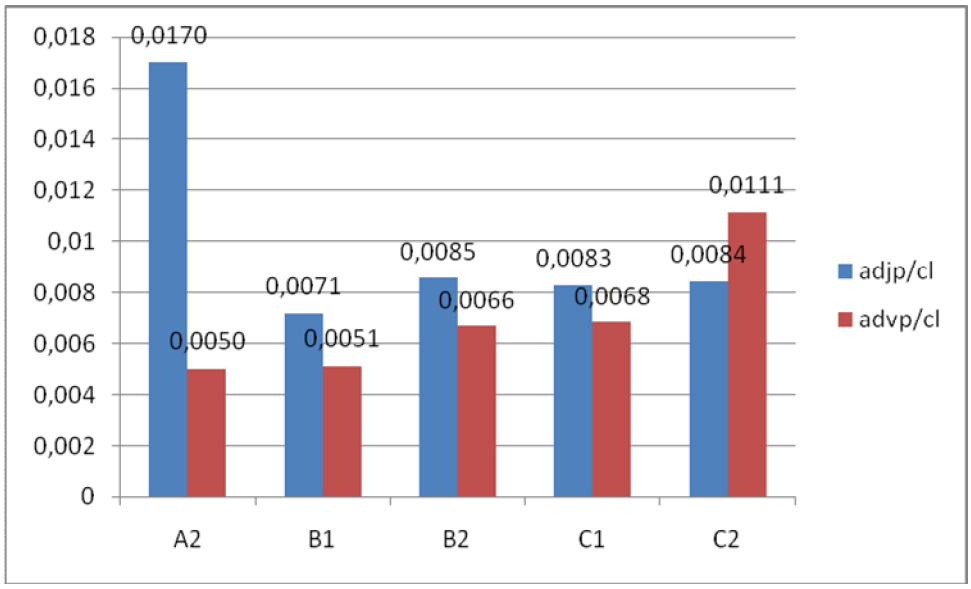


Graph 18 Mean quantity of non-verbal modal classes per clause

Graph 19 below shows the amount of non-verbal modal markers per clause across proficiency levels. As can be observed, there is a wide difference in the usage of adjective phrases shown by the lowest proficiency level (A2 students, with 0.0170 per clause) and all the others. The quantities employed by the rest of the proficiency levels are all very similar, ranging from 0.0071 in B1 essays, 0.0085 in B2 essays, 0.0083 in C1 and 0.0084 in C2.

Hence, roughly speaking, the proportions displayed by all the levels over B1 represent a little less than half of the total amount of adjective phrases per clause used by A2 students. If we had to establish a pattern, we could say there is a falling-rising pattern, since by the severe fall at B1, the use of adjective phrases experiences a slight increase.

As for adverbials, the figures suggest an increase in their use as the linguistic competence of students improves. Hence, adverbials represent 0.0050 and 0.0051 in A2 and B1 essays, respectively, whereas at C2 there are 0.0111, which is the double. In between we have 0.0066 and 0.0068 in B2 and C1, respectively.



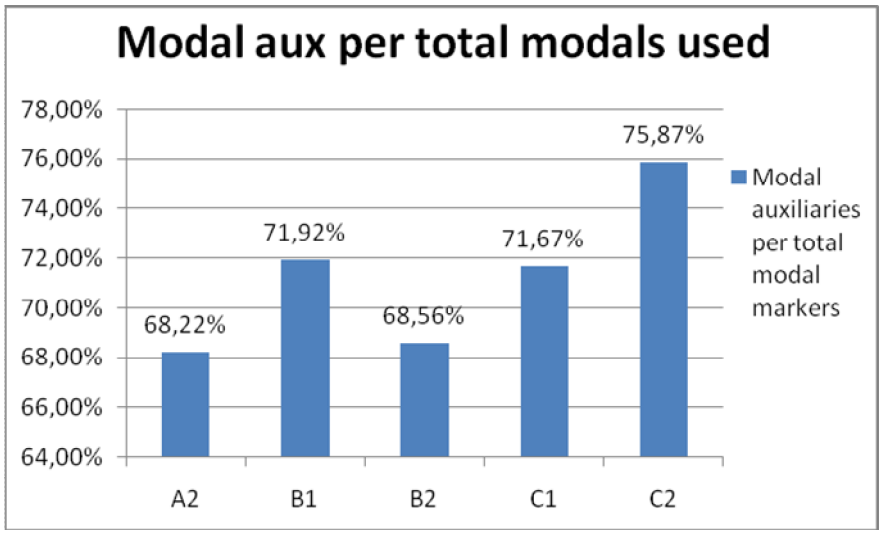
Graph 19: Amount of non-verbal categories per clause across proficiency levels

4.5. Proportion of verbal and non-verbal modal categories per total number of modal markers used

Having examined the usage of the different grammatical classes expressing modality in absolute terms, in general and across proficiency levels, I will now look at the proportion of each verbal and non-verbal modal category in relation to the total number of modal markers used at

each level of proficiency, to see whether the tendencies shown are similar to the ones observed when looking at the quantities per clause above mentioned, and also to see the choices students make across proficiency levels and their evolution.

In regards to modal auxiliaries, there seems to be a clear rising pattern, as Graph 20 below indicates. If we take the amount of modal auxiliaries per total modal markers used by A2 students, 68.22%, and we compare it with the proportions of the rest of proficiency levels –displayed in Graph 20-, we can conclude that there is an increasing preference for this type of modal markers as the students' linguistic level rises.

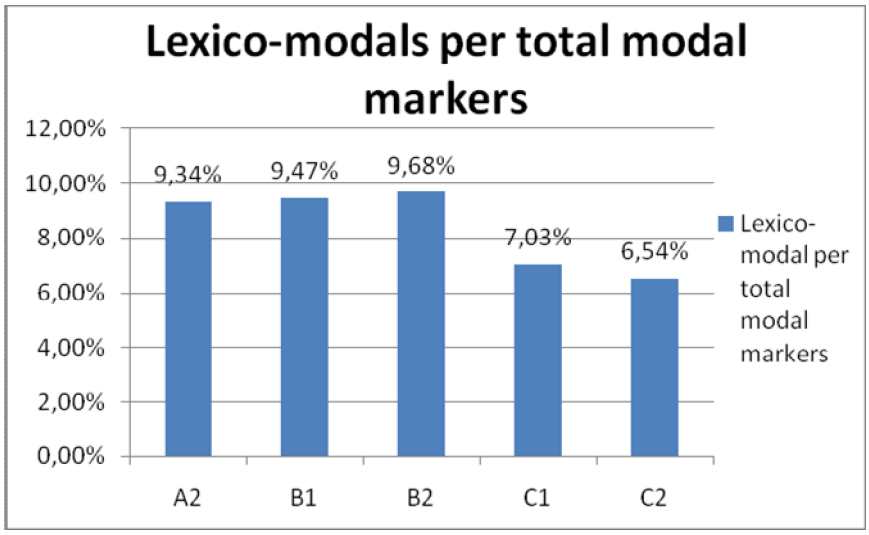


Graph 20: Modal auxiliary verbs per total modal markers used across proficiency levels

As it follows, these results are quite different from the tendency observed in the usage of modal auxiliaries per clause, presented in the previous section, which pointed to a declining tendency across proficiency levels (with the only exception of C2 students). This last trend would have to do with the general declining use of modality as the proficiency level rises.

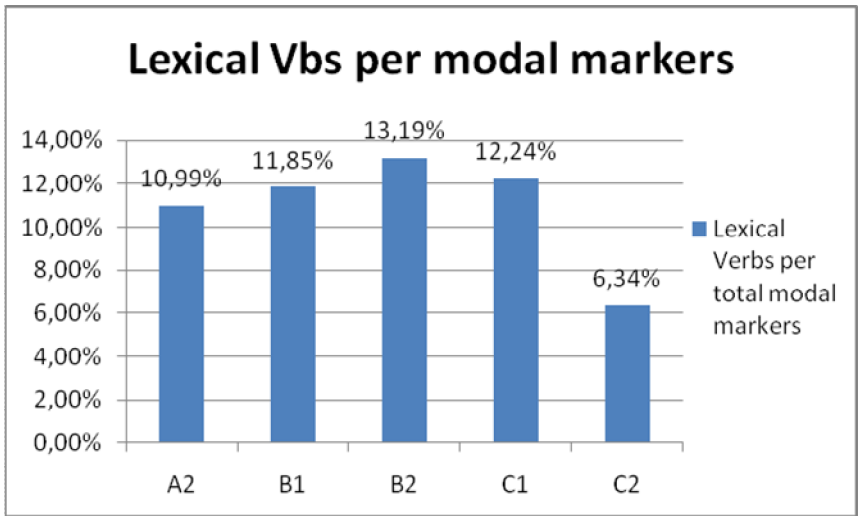
As for lexico-modal verbs, the percentages show a slight increase until B2 (see Graph 21 for the exact figures). Then, their use falls in 2.65% by C1 (with 7.03%) and even 0.5% more

over the following level (with 6.54%). Hence, the tendency distinguished here is identical to the one observed in the proportions per clause.



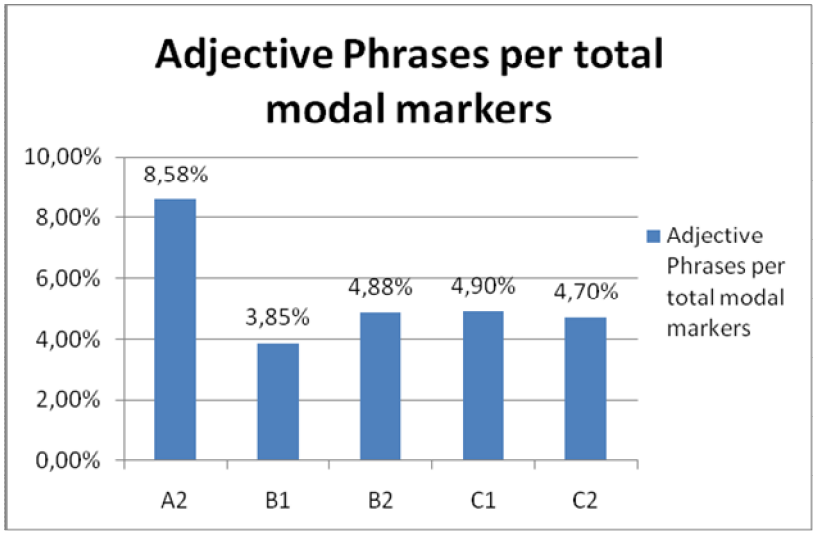
Graph 21: Lexico-modal verbs per total modal markers used ac ross proficiency levels

Regarding the lexical verbs, the pattern that can be observed in Graph 22 is a rising-falling one, where the amount of lexical verbs per total modal markers used increases from A2, with 10.99%, to B2, with 13.19%, through B1, with 11.85%. It, then, falls to 12.24% by C1 and by C2 it drops to half this figure, 6.34%. Consequently, the tendency here is exactly the same as I noticed when looking at proportions per clauses.



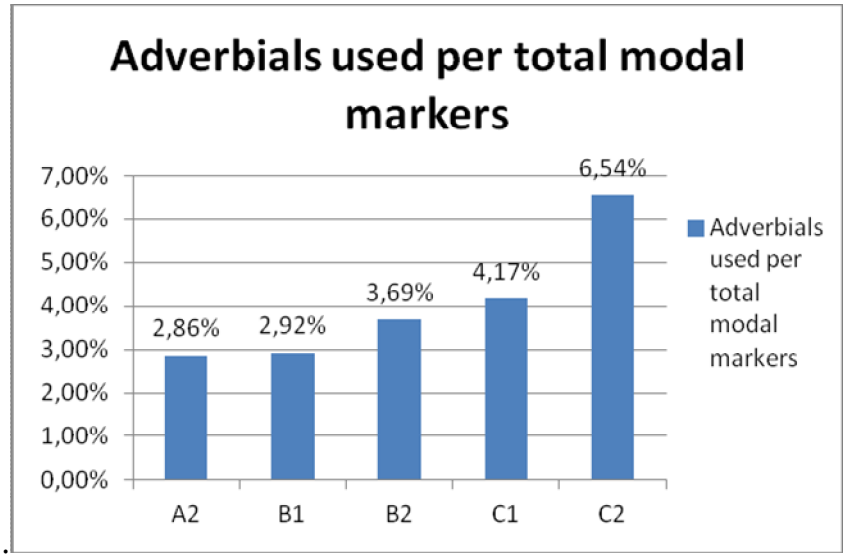
Graph 22: Lexical verbs per total modal markers used across proficiency levels

Concerning non-verbal modal means, the proportions of adjective phrases per total modal markers used across proficiency levels yield a quite similar pattern of usage to the one distinguished when dealing with proportions per clause. Thus, there is also a high peak at A2 level, a severe fall over B1 (see Graph 23 below for the exact percentages), and then, a rise of approximately 1%, which continues up to C2, with similar proportions for B2, C1 and C2, as Graph 23 shows:



Graph 23: Adjective phrases per total modal markers across proficiency levels

Finally, the amount of adverbials per total modal markers also shows a similar rising tendency as their proportions per clause, mentioned above. Obviously, it is the actual proportion per total modal markers that differs, which is displayed in Graph 24 below



Graph 24: Adverbials per total modal markers across proficiency levels

In conclusion, all the patterns I noticed when dealing with the proportions of each grammatical category expressing modality per clause are identical to the ones obtained when looking at proportions per total modal markers used, with the only exception of modal auxiliaries. If looking at modal auxiliaries per clause, this verbal class shows a clear contraction over proficiency levels except for C2 students, in their proportion per total modal markers their tendency is exactly the opposite, since their use increases across proficiency levels, only to varying degrees.

4.6. Elements realizing each modal class

As important as the quantity of modality used is the actual elements realizing each modal class that students choose and their proportions. I will, thus, now examine the actual tokens that comprise each category, both in general and across levels of proficiency, to see their frequencies.

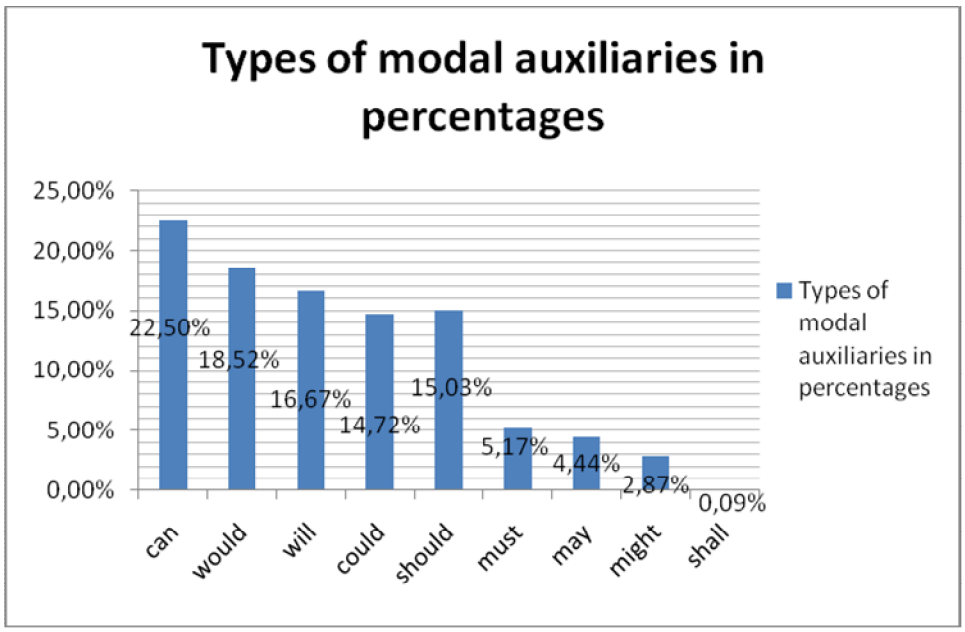
- ### Verbal categories:

I will first look at the realizations of each verbal class produced by the students as a whole, and then, I will look into each proficiency level.

After this, I will examine the elements selected by the learners within the non-verbal modal categories.

a) Modal Auxiliary verbs

Within the class of modal auxiliary verbs the most frequently used token is *can*, followed by *would*, *will*, *should*, and *could* which represent approximately the same percentage, as shown in Graph 25 below –which displays the exact proportions. The remaining four modals: *must*, *may*, *might* and *shall* are scarcely employed, since their proportions are inferior to 6%. *Shall* is the least frequently used modal auxiliary.



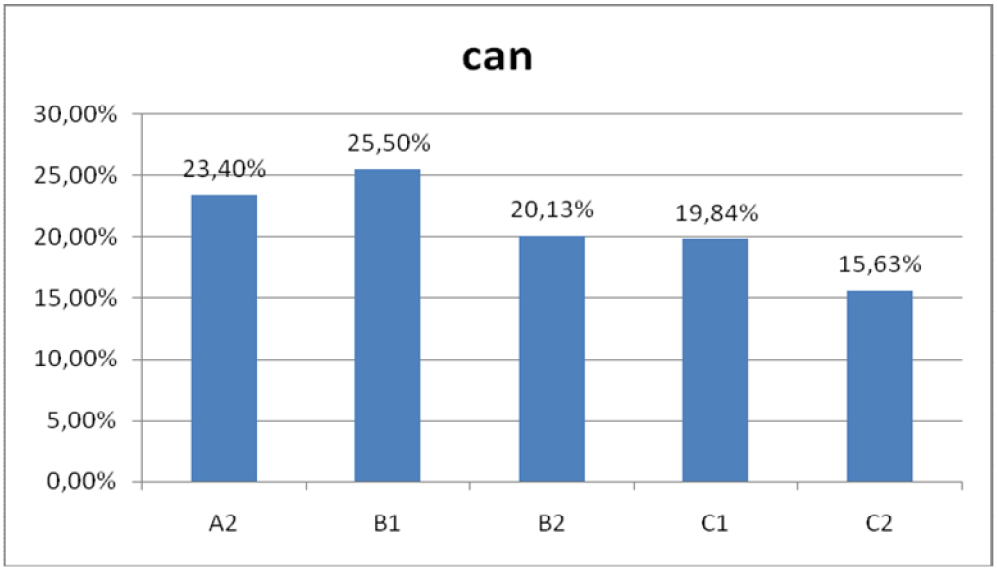
Graph 25: Elements realizing the modal auxiliary verb category in all the essays

It is worth mentioning again that the count of each particular realization of the category, such as CAN, includes its negated and contracted forms. However, for ease of presentation, I choose to use their bare forms in the discussion of the data. Since my analysis is grammatical, the inclusion of the positive and negative forms of the same modal auxiliary verb –which may

express a totally different semantic type of modality- does not affect the results (e.g. *can* and *cannot* express different semantic types of modality).

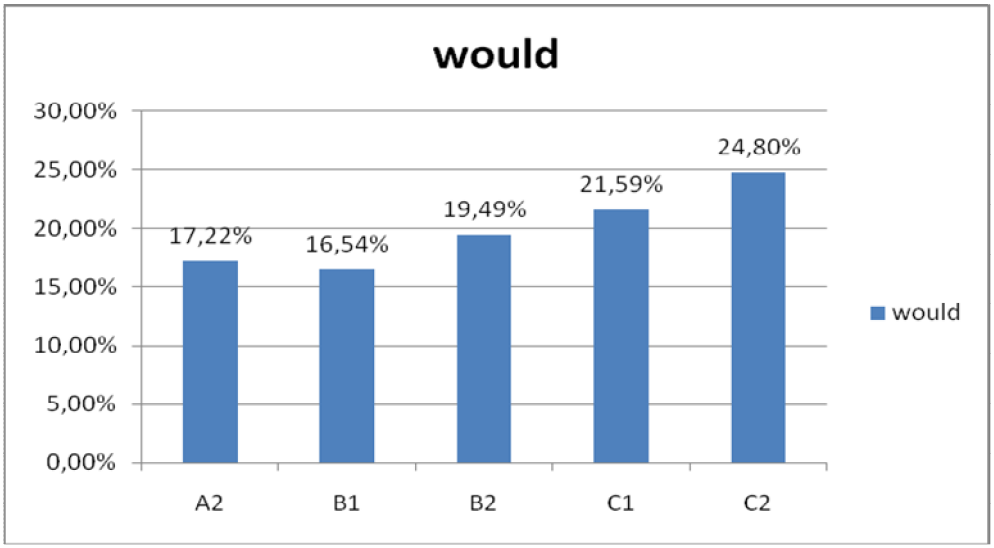
At this point, instead of looking at all the elements realizing the category of modal auxiliaries across proficiency levels, I opt for showing the patterns of use of the five most frequently used modal auxiliary verbs (i.e. CAN, WOULD, WILL, COULD and SHOULD, respectively) for the sake of clarity in my exposition and for space reasons. However, I will include the graphs which offer a detailed account of the types and percentages of modal verbs used across proficiency levels in **Appendix 1**.

From most to least frequently used modal auxiliary verbs, I will start with the evolution of the usage of CAN across proficiency levels. As Graph 26 below illustrates, there is a fall in the use of this modal verb as student's proficiency level rises.



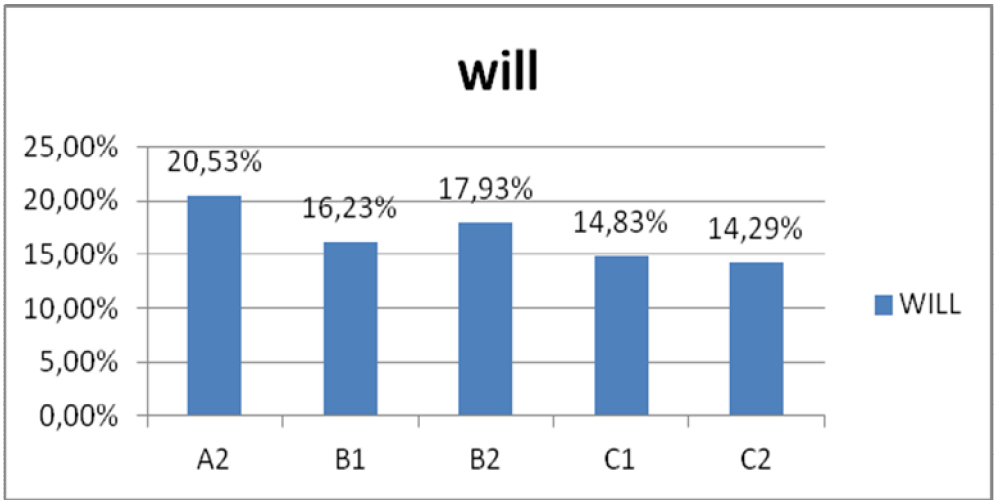
Graph 26: Use of CAN across proficiency levels

Conversely, the use of WOULD shows exactly the opposite tendency (shown in Graph 27 below) with a rising pattern as proficiency levels increase.



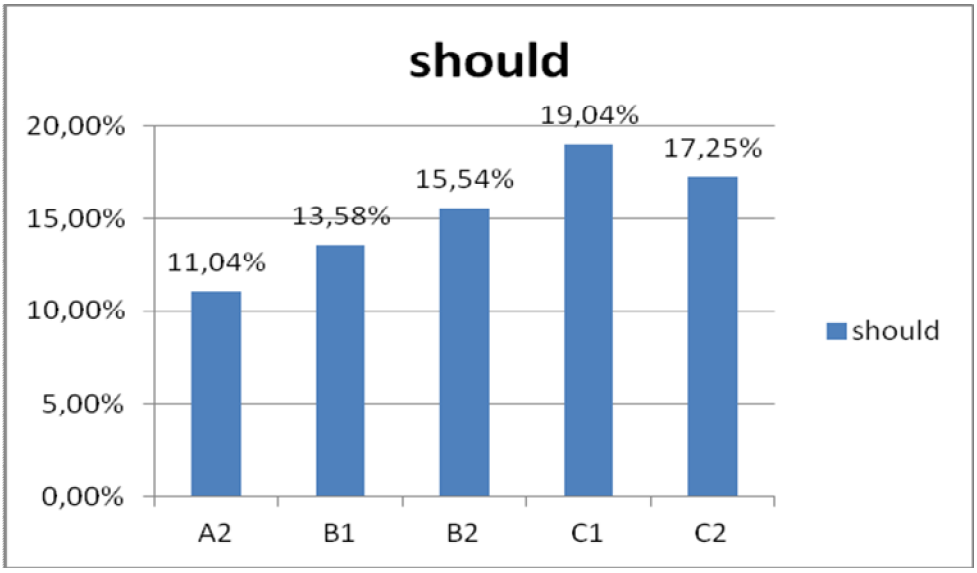
Graph 27: Use of WOULD across proficiency levels

Regarding the third most frequent modal auxiliary in the corpus, WILL, its usage experiences a similar falling pattern than CAN as the learners' linguistic competence improves. This pattern is visually displayed by Graph 28.



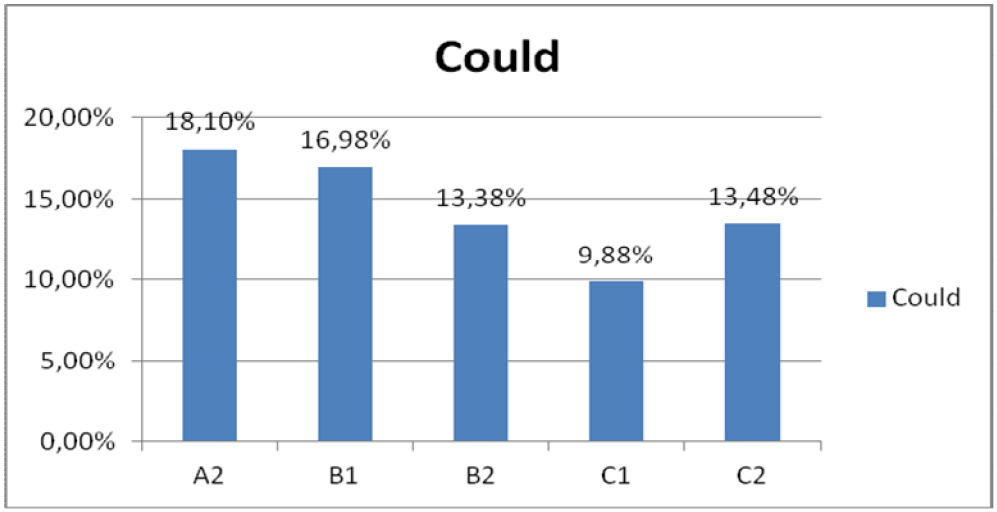
Graph 28: Use of WILL across proficiency levels

As for SHOULD, it shows an increasing presence in the writing of the students as their proficiency level rises, with the only exception of C2 students, whose percentage falls a bit in comparison with C1 students, as can be seen in Graph 29.



Graph 29: Use of SHOULD across proficiency levels

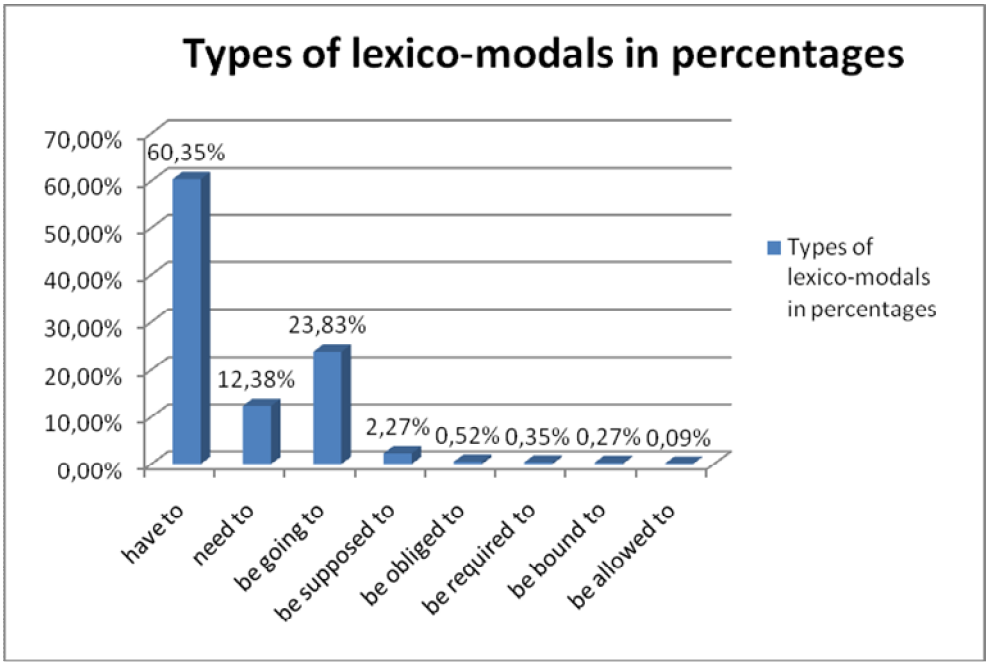
Finally, COULD shows a decreasing pattern of usage, as Graph 30 makes clear. Again, the exception is represented by C2 students, whose percentage rises over C1 and, also, B2 students.



Graph 30: Use of COULD across proficiency levels

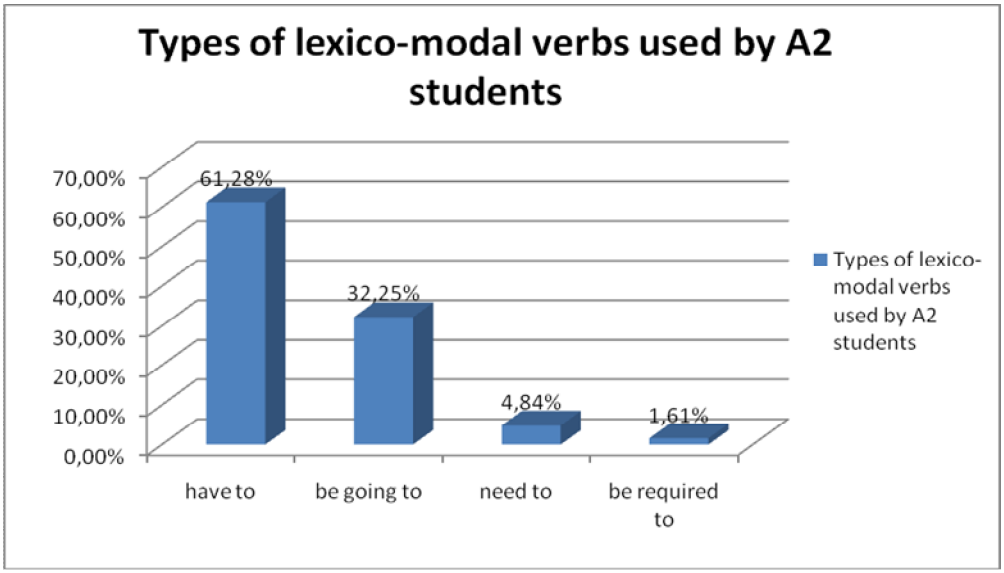
b) Lexico-modal verbs

I will now closely look at another category of verbal elements that can convey modality, that is, lexico-modal verbs. First, regarding the actual elements used within this category in all the texts I have examined, the most common element is *have to* (which represents 60.35% of the total occurrence of this category). This is followed, by far, by *be going to* (with 23.83%) and *need* (with 12.38%). The other tokens used are almost inexistent with less than 1%. This is the case of elements such as *be obliged to, be required to, be bound to* or *be allowed to*. The only exception is *be supposed to* with a slightly higher percentage, 2.27%, as can be seen in Graph 31.



Graph 31: Elements realizing the lexico-modal verb category in all the essays

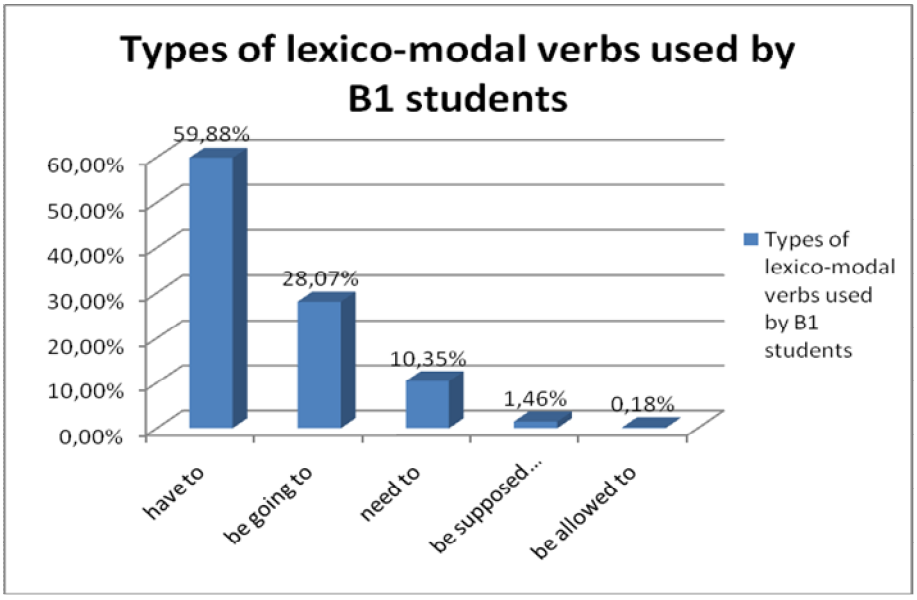
As opposed to modal auxiliary verbs, I will here provide the proportions of all the tokens students use across proficiency levels, with the aim of showing the wider variety of lexico-modal elements employed as students' proficiency level rises. As Graph 32 illustrates, A2 essays follow the general pattern presented above (see Graph 31 above), since the most common realization of this class is *have to*, which almost doubles the following one, *be going to* (see Graph 32 below for the exact figures). Only two other tokens are employed at this level, *need to* and *be required to*, but with very low percentages.



Graph 32: Elements realizing the lexico-modal verb category in A2 essays

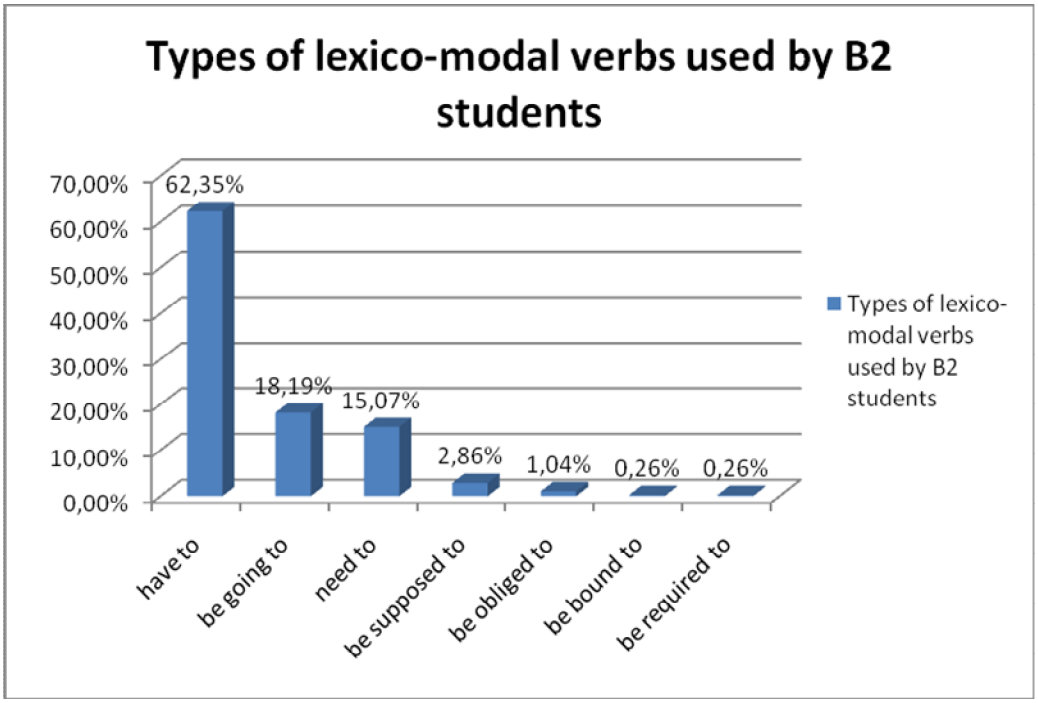
As a result of this, it can be safely stated that the scarce variety of lexico-modal tokens is the most salient feature at this level as well as the predominance of *have to*.

As for the lexico-modal tokens displayed by B1 students, they are roughly the same as the ones used by A2 students with a difference in their percentages, as Graph 33 illustrates. Again, the most common verb is *have to*. Also, *be going to* and *need to* occupy the second and third positions but with much lower percentages. Finally, *be supposed to* and *be allowed to* constitute the new categories B1 students start using, even though in very small quantities.



Graph 33: Elements realizing the lexico-modal verb category in B1 essays

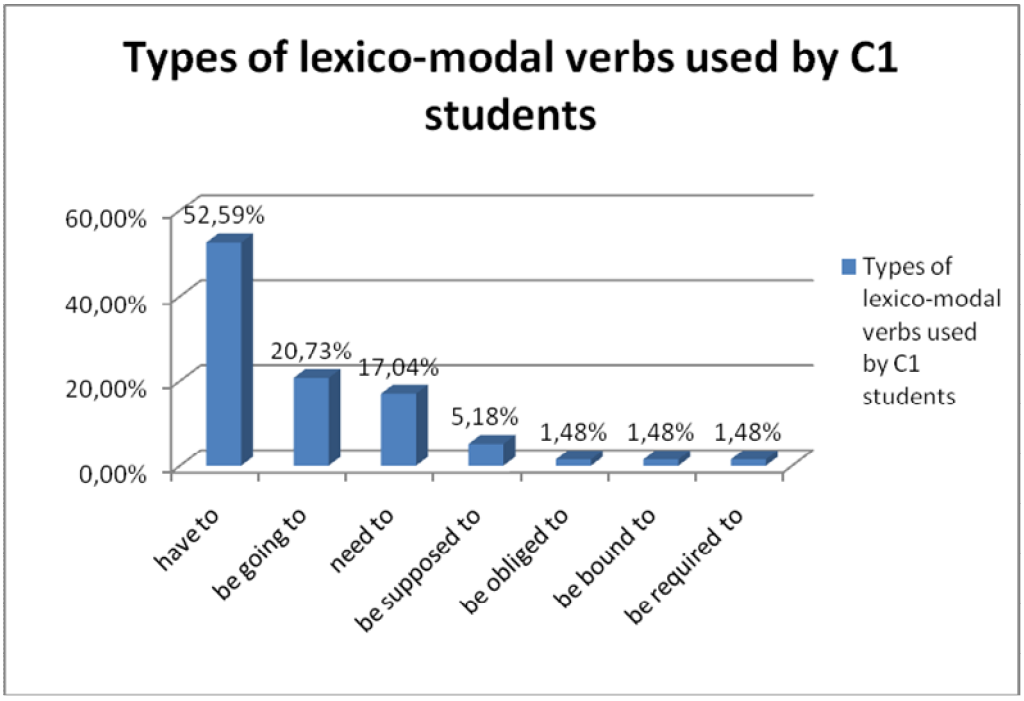
B2 students also prefer *have to* within this category, which constitutes an overwhelming proportion of 62.35%, as Graph 34 represents. It is again followed by *be going to* and *need to,* with lower proportions. It should be mentioned that if the former experiences a clear contraction in its presence, the latter shows an opposite evolution to rise. The lowest percentages correspond in this case to *be supposed to, be obliged to, be bound to* and *be required to*, which are displayed in Graph 34 below. The most characteristic trait at this level is, thus, the increasing variety of elements within this grammatical category to express modality.



Graph 34: Elements realizing the lexico-modal verb category in B2 essays

By the same token, C1 students' data clearly show an increasing variety of elements within lexico-modal verbs. If we look at the figures, almost half of the times C1 students employ a lexico-modal verb they choose *have to*. Again, it is followed by *be going to* and *need to*, in much lower proportions, as illustrated by Graph 35. Thus, the latter continues its growth whereas the former does not display here its lowering tendency. Nevertheless, its proportion at this level is not so different from B2 essays.

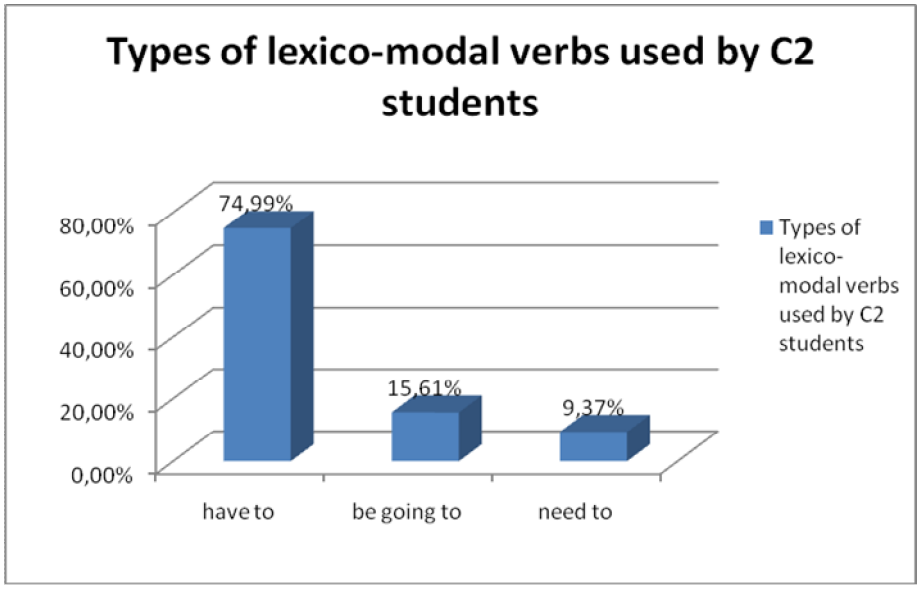
Another interesting aspect of C1 essays is the higher frequency of the, so far, least frequently employed lexico-modal verbs. A proportion of 5.18% is obtained for *be supposed to* whereas 1.48% is the figure for *be bound to, be obliged to* and *be required to,* as Graph 35 shows.



Graph 35: Elements realizing the lexico-modal verb category in C1 essays

C2 students display the least variety of lexico-modal verbs with only three different realizations, which is represented by Graph 36 below. However, they stick to the general pattern noticed across proficiency levels in that these elements are the recurrent lexico-modals *have to, be going to* and *need to*. *Have to* is the dominant token within the category with 74.99%, very far from *be going to* and *need to*, which constitute 15.61% and 9.37% of the total class respectively.

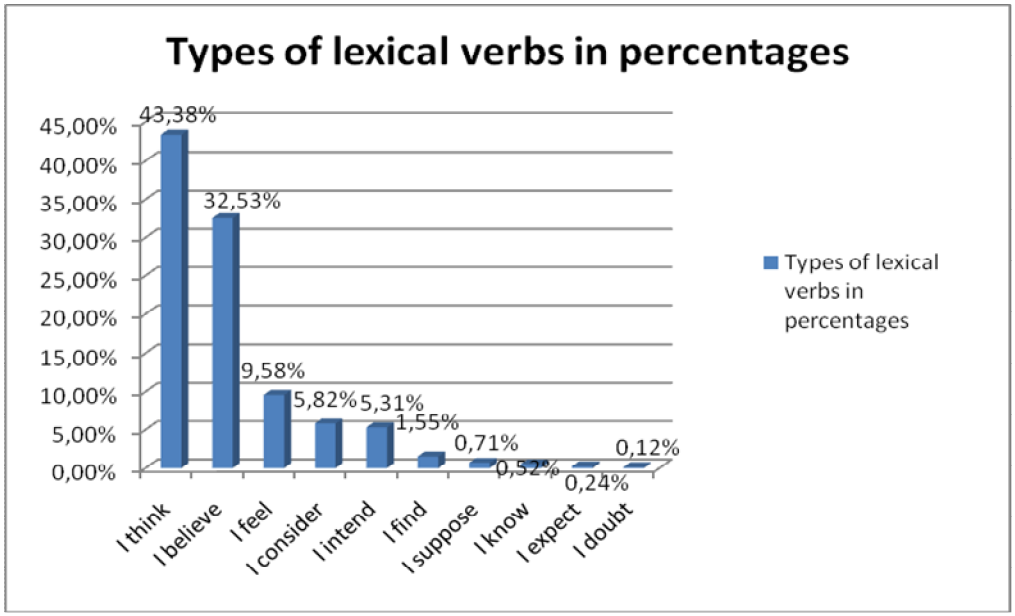
Consequently, *need to* does not show its tendency to rise here, although *be going to* does follow the decreasing trend observed across proficiency levels, as Graph 36 below illustrates. In any case, it should be mentioned that the most advanced level, C2, is the least representative in this study, therefore, caution should guide our observations regarding their patterns of use.



Graph 36: Elements realizing the lexico-modal verb category in C2 essays

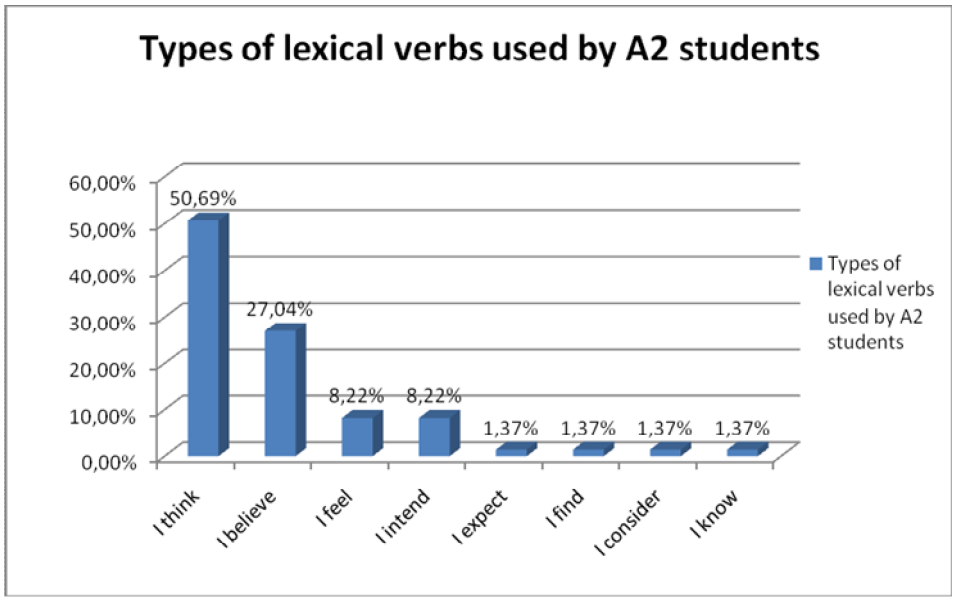
c) Lexical Verbs

As for lexical verbs, the learners as a whole prefer *I think* and *I believe*, which together represent three quarters of the total elements used within this category (with 43.38% and 32.53% respectively). The following elements are used in a proportion inferior to 10% each. These are *I feel* (with 9.58%), *I consider* (with 5.82%), and *I intend* (with 5.31%). Finally, the least frequently used lexical verbs are *I find*, *I suppose*, *I know*, *I expect* and *I doubt* (see Graph 37 below for exact figures).



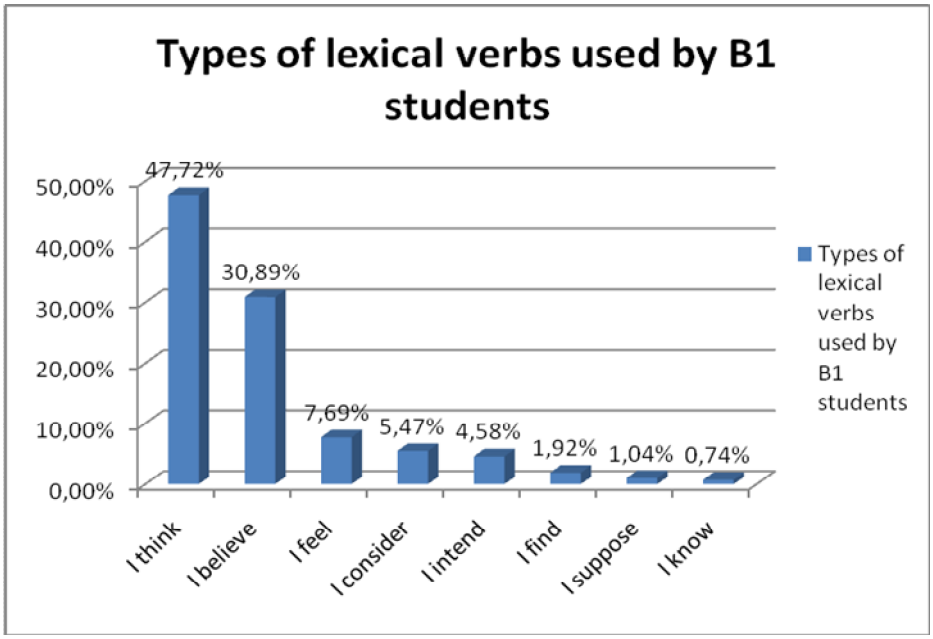
Graph 37: Elements realizing the lexical verb category in all the essays

Across proficiency levels and focusing on A2 students, the data gathered suggest a similar tendency to the general pattern mentioned above, since *I think* and *I believe* constitute a little bit more than 75% of the total class of lexico-modals. Far from these figures, and as Graph 38 below shows, we can find the verbs *I feel* and *I intend*. The least common elements within this category are *I expect*, *I find*, *I consider* and *I know.*



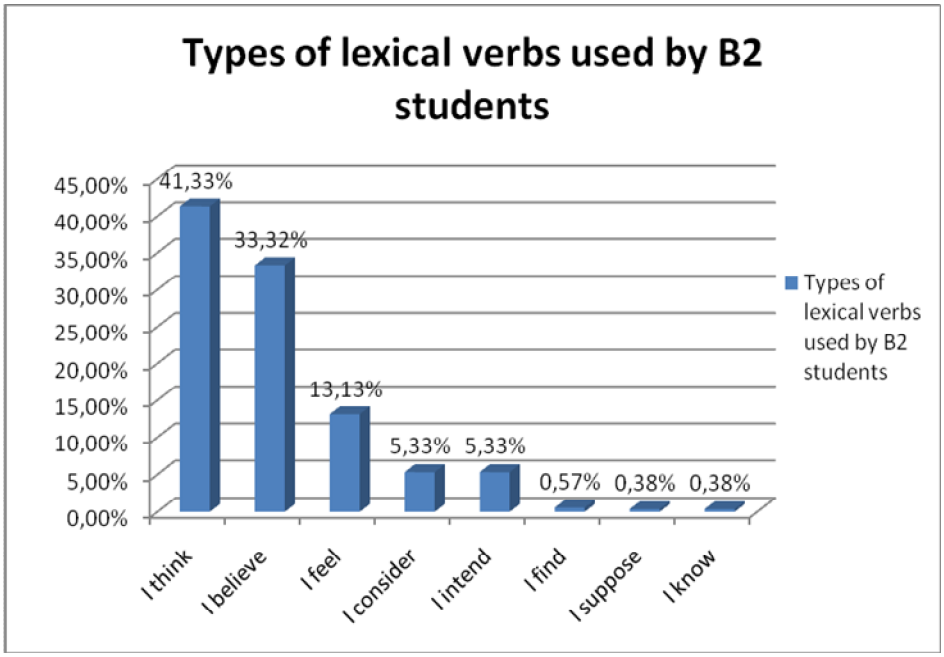
Graph 38: Elements realizing the lexical verb category in A2 essays

Again, a similar pattern can be observed among B1 students, represented in Graph 39. They also employ the lexical verbs *I think* and *I believe* with a modal meaning to a great extent, representing a bit more than 75% of the total class (47.72% and 30.89% respectively). Three verbs follow them in the distance, with an occurrence inferior to 10%, which are *I feel* with 7.69%, *I consider* with 5.47%, and *I intend*, 4.58%. The least common verbs present are *I find*, with 1.92%; *I suppose* with 1.04%; and, finally, *I know* with only 0.74%.



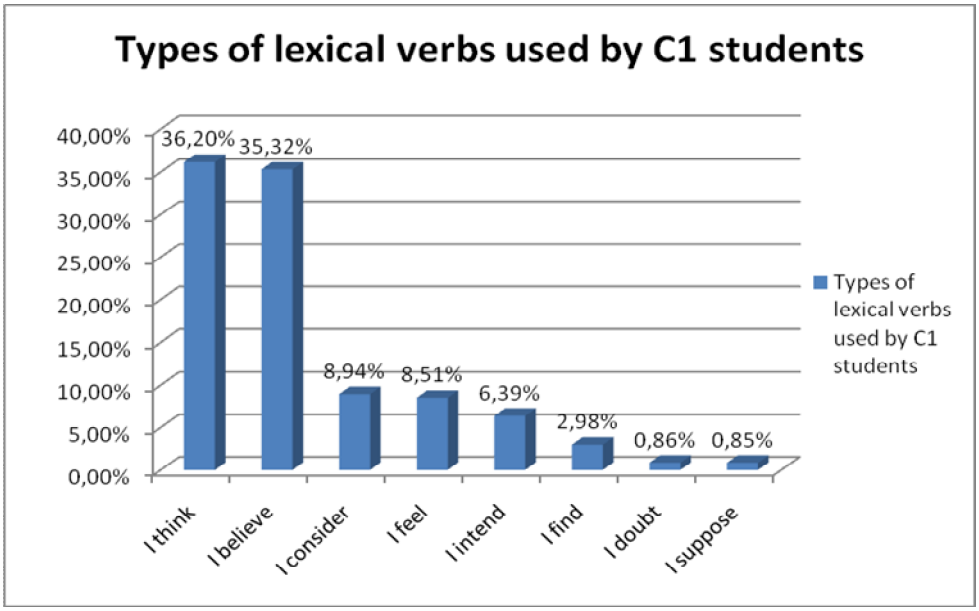
Graph 39: Elements realizing the lexical verb category in B1 essays

Similarly, B2 students widely use *I think* and *I believe*, which again constitute almost 75% of the total elements used within this class (41.33% and 33.32% respectively), as Graph 40 shows. With a much lower percentage, 13.13%, *I feel* represents the third most common verb B2 students make use of to convey modality. Even less representative, with merely 5.33%, are *I consider* and *I intend*. And, with less than 1% there appear *I find, I suppose* and *I know.*



Graph 40: Elements realizing the lexical verb category in B2 essays

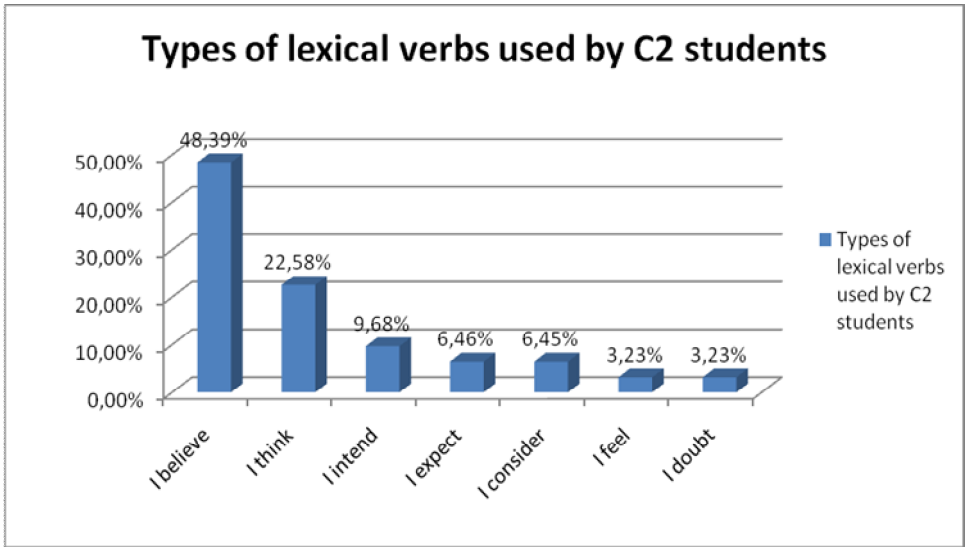
Likewise, C1 students display the same tendency observed at the lower linguistic levels, since *I think* and *I believe* also represent the highest percentages, 36.20% and 35.32% each and a bit more than 70% both together. The following most frequently used verbs are *I consider* with 8.94%, *I feel* with 8.51%, *I intend* with 6.39%, and *I find* with 2.98%, but with a far less presence in their writing as Graph 41 shows. Finally, *I doubt* and *I suppose* represent less than 1% of the total elements used in this verbal category.



Graph 41: Elements realizing the lexical verb category in C1 essays

Conversely, C2 students constitute the exception here since they are the only ones who prefer by far the verb *I believe (*with 48.39%) to *I think* (with merely 22.58%). Again, I should mention the few C2 essays collected in WriCLE and, as a result of it, the need for more data to confirm or discard the tendencies observed.

However, it is worth mentioning that they still stick to the above mentioned pattern of employing both *I believe* and *I think* pervasively –both together represent a bit more than 70%. The following choices C2 students make are *I intend* with 9.68%, *I expect* and *I consider*, with 6.46% and 6.45% respectively. Again, there is a long distance between the two most frequently used verbs and all the others, which is displayed by Graph 42. Finally, with an identical percentage, 3.23%, the least frequently used verbs are *I feel* and *I doubt*.



Graph 42: Elements realizing the lexical verb category in C2 essays

d) Modal Idioms

As it has already been mentioned, no instance of modal idioms has been found in the corpus regardless of the level of proficiency.

▪ NON-VERBAL MODAL CATEGORIES:

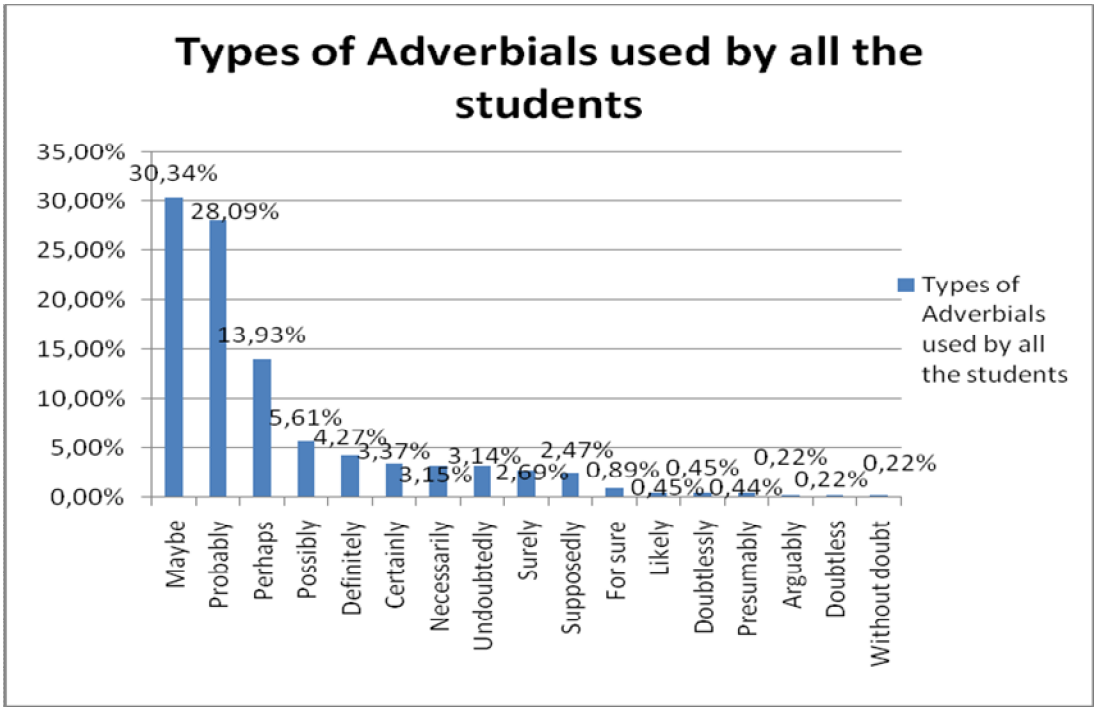
As regards to non-verbal modality, the main categories to be analyzed are adverbials and adjective phrases. I will start by examining the adverbials employed, offering an overview of the specific adverbials used by all the students as a whole and then, I will provide a detailed analysis of the adverbial realizations found across proficiency levels.

e) Adverbials

To begin with, the learners as a group show a general tendency to prefer *maybe* and *probably*, which together represent almost 60% of the total class of adverbials (see Graph 43

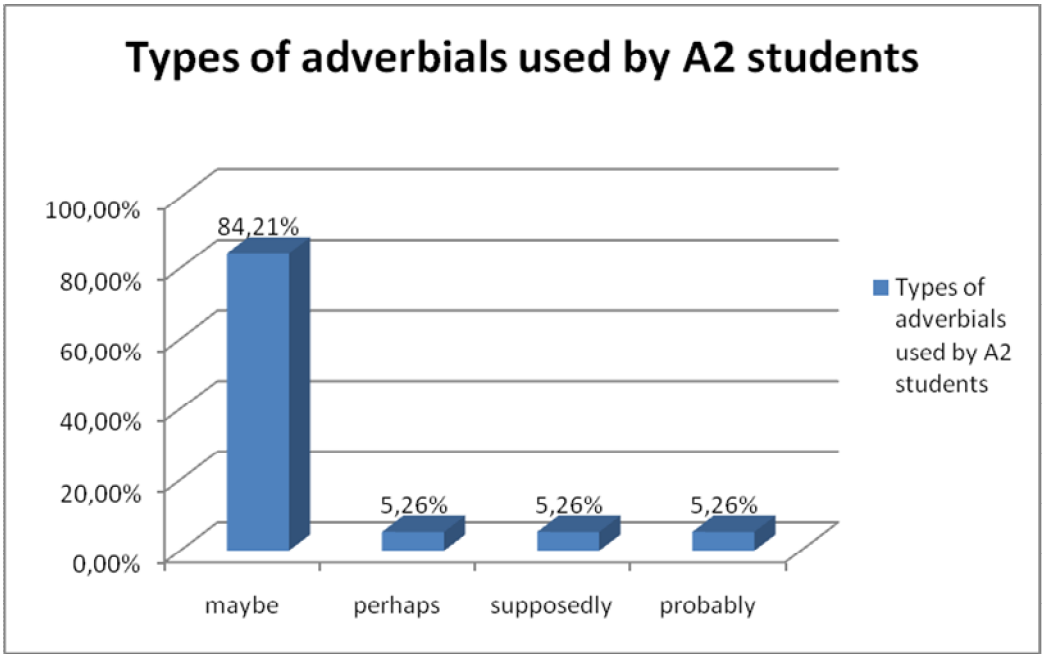
below for exact figures). The third most frequent adverbial is *perhaps* with 13.93%. This wide difference is visually captured by Graph 43 below.

Far less representative, with less than 6%, are *possibly*, 5.61%; *definitely*, 4.27%; *certainly*, with 3.37%; *necessarily*, 3.15%; *undoubtedly*, 3.14%; *surely*, 2.69%; and, *supposedly*, 2.47%. The least representative elements of this class, with less than 1%, are *for sure, likely, doubtlessly, presumably, arguably, doubtless* and *without doubt*.



Graph 43: Elements realizing the adverbial category in all the essays

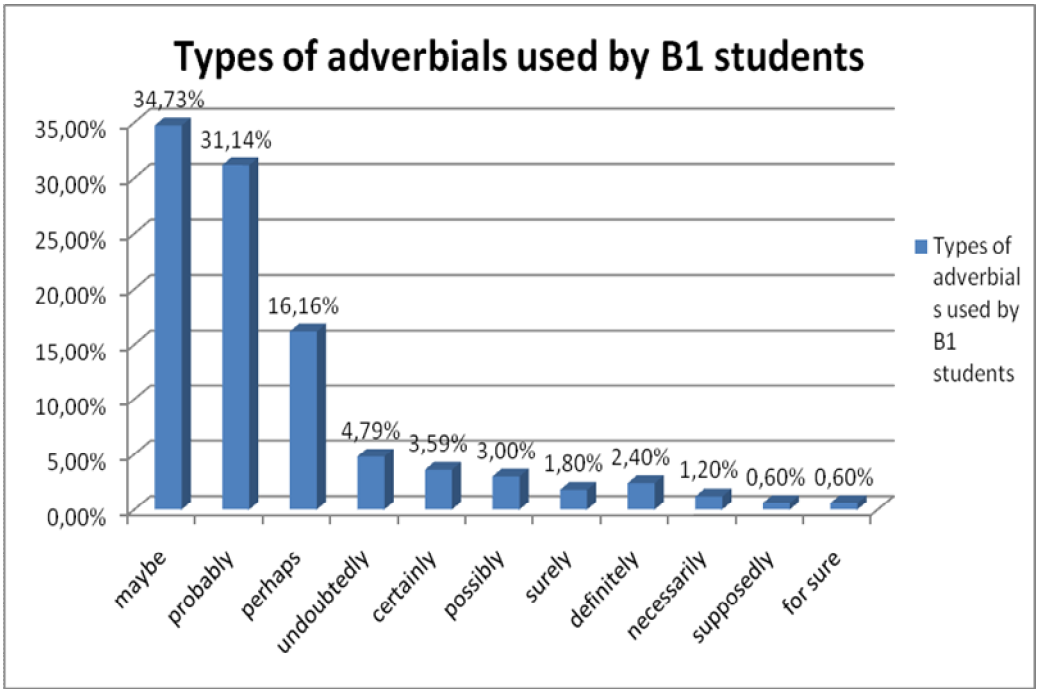
When looking at the type of adverbials learners use at each level, we can see a clear preference for *maybe* among A2 students (which represents almost 85% of the total class of adverbials), represented in Graph 44 below. Only three more elements are used within this category, *perhaps*, *supposedly* and *probably*, which are evenly distributed, 5.26% each.



Graph 44: Elements realizing the adverbial category in A2 essays

Also, B1 students use *maybe* to the largest degree, 34.73%, but obviously, its presence is not so dominant as in the case of A2 students, where it constitutes 84.21% of the total. Another difference with respect to the lowest proficiency level present in the corpus is the ample presence of *probably*, which represents 31.14% of all the adverbials. The third choice students make at this level is *perhaps* with 16.16%. Graph 45 below visually illustrates these percentages.

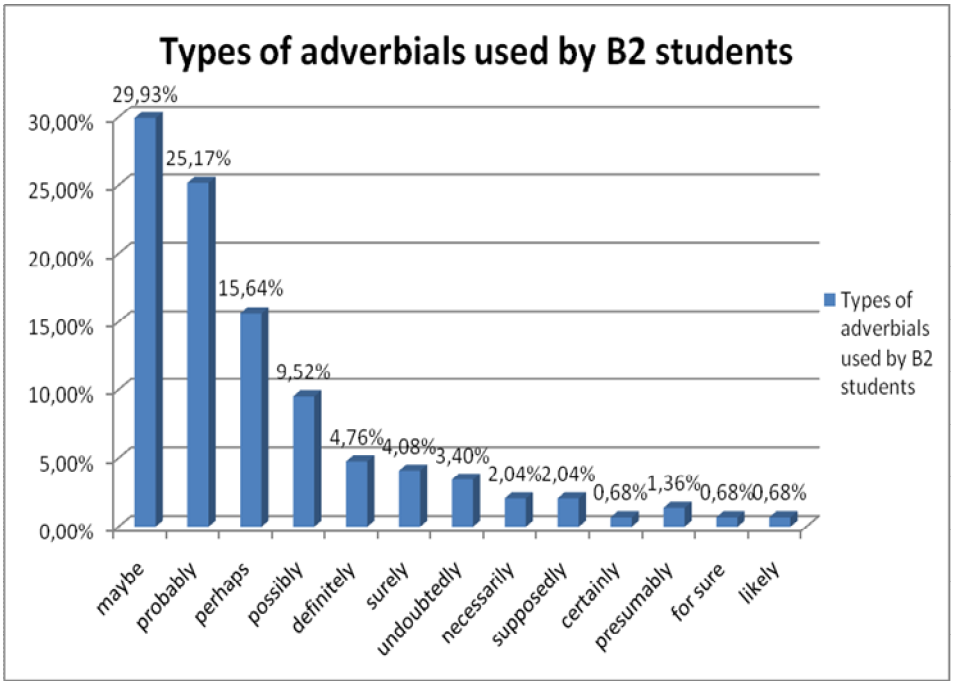
With less than 5% of presence in their writing, there appears a large number of adverbials, such as *undoubtedly*, *certainly*, *possibly*, *definitely*, *surely* and *necessarily*. This wider variety of elements also constitutes a significant difference in relation to A2 students, whose range of tokens within this class proved very limited. Finally, *supposedly* and *for sure* represent less than 1% of the class (see Graph 45 for details).



Graph 45: Elements realizing the adverbial category in B1 essays

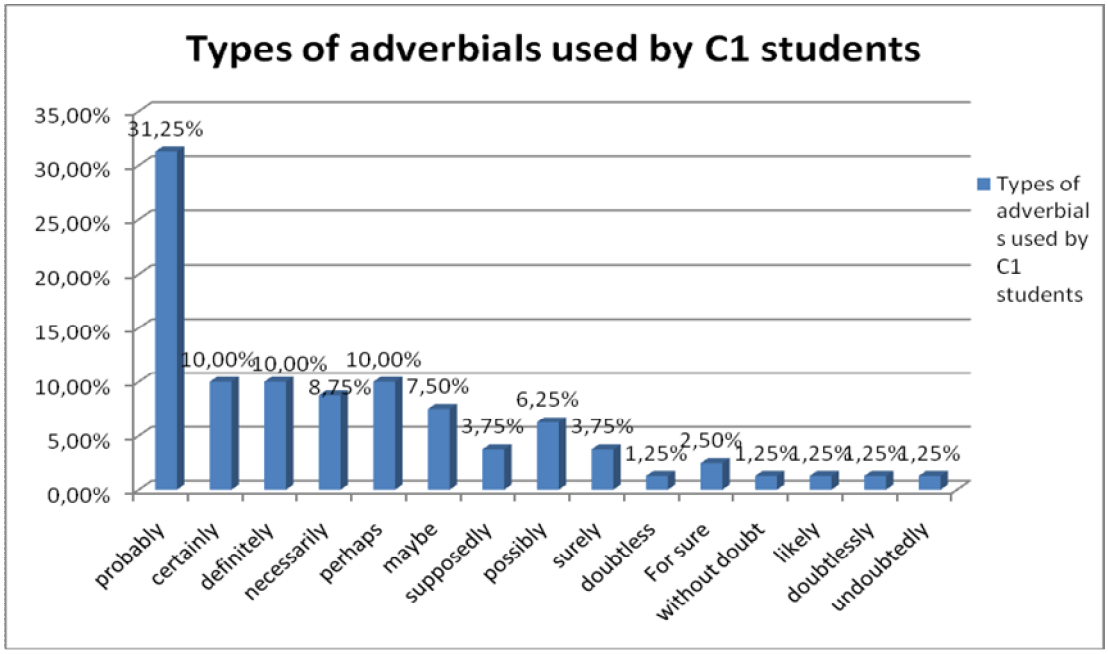
A very similar pattern is displayed by B2 students. They also use pervasively *maybe*, 29.93%, and *probably*, 25.17%, which by themselves represent more than half the total class. Again, *perhaps* occupies the third position with a lower figure, 15.64%. As for the variety observed among B1 students in the less representative adverbial tokens, we can conclude that it is also present in B2 essays, as can be observed in Graph 46 below.

As a way of illustration, we can mention *possibly* with 9.52%; *definitely*, 4.76%; *surely*, 4.08%; *undoubtedly*, 3.40%; *necessarily*, 2.04%; *supposedly*, 2.04%; and *presumably*, 1.36%. Finally, with an occurrence inferior to 1%, there appear *certainly*, *for sure*, and *likely*.



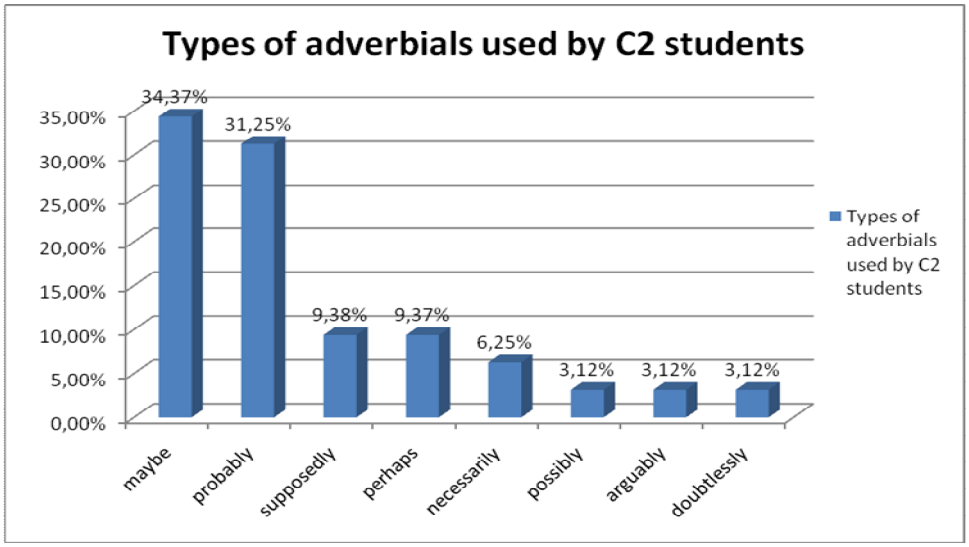
Graph 46: Elements realizing the adverbial category in B2 essays

Contrariwise, C1 students prefer *probably*, with 31.25% of the total, to *maybe*, which, among these learners, merely represents 7.50%. Three elements, *certainly*, *definitely* and *perhaps* occur evenly with quite a significant presence, 10%. A bit lower is the percentage of *necessarily*, 8.75%, and *possibly*, 6.25%. With less than 4% we can find *supposedly* and *surely*, both with 3.75%; *for sure* with 2.50%; and with 1.25% *doubtless*, *without doubt*, *likely*, *doubtlessly* and *undoubtedly*. Graph 47 below visually represents the distribution of this wide variety of elements.



Graph 47: Elements realizing the adverbial category in C1 essays

C2 students do not show as much variety of adverbials as C1, B2 or B1 students, as Graph 48 below shows. However, they follow the general tendency observed among the other proficiency levels –with the only exception of C1 students- of preferring *maybe*, which represents 34.37%, and *probably*, 31.25%. The three elements which show a less significant occurrence, but still represent more than 5% of the total, are *supposedly*, *perhaps* and *necessarily*, with 9.38%, 9.37% and 6.25%, respectively. The least frequent adverbials used by advanced learners are *possibly*, *arguably* and *doubtlessly*, with an identical proportion, 3.12%.

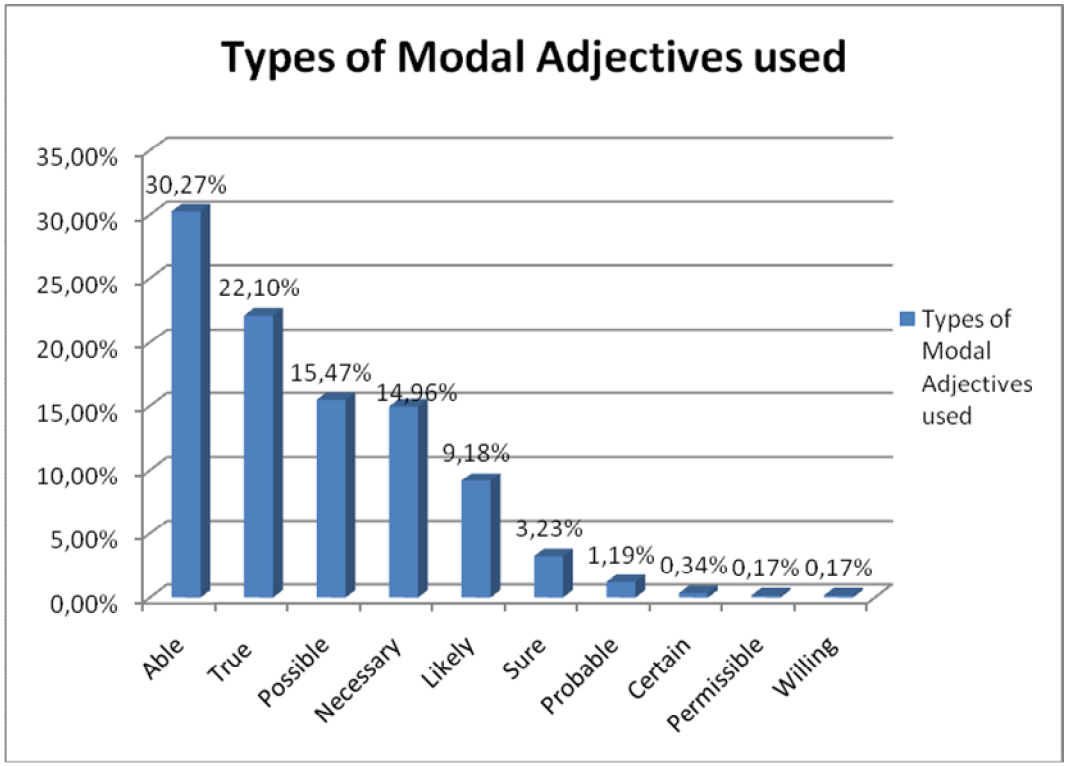


Graph 48: Elements realizing the adverbial category in C2 essays

f) Adjective Phrases

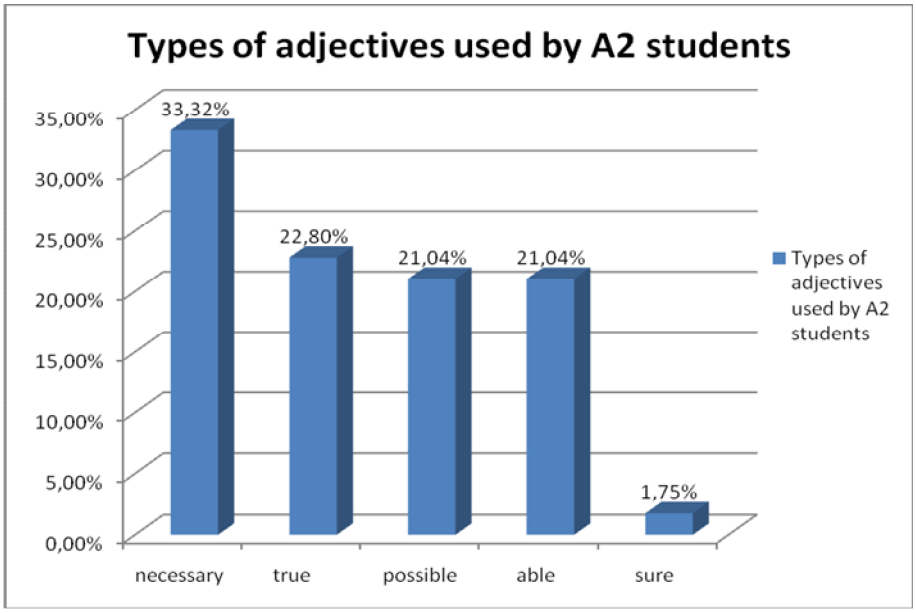
Finally, concerning the other class of non-verbal means to express modality, that is to say, adjective phrases, I will first summarize the elements belonging to this category that students as a whole prefer, and then, I will look at each level of proficiency in detail to try to raise some conclusions.

As regards the bulk of students, the most frequent adjective used with a modal meaning is *able*, which represents 30.27%. The second token in percentage is *true* with 22.10%, followed by *possible* and *necessary*, with 15.47% and 14.96%. With less than 10% of occurrence we can mention *likely*, with 9.18%; *sure*, with 3.23%; *probable*, with 1.19%; and *certain*, *permissible* and *willing* with less than 1%. Graph 49 below offers a visual representation of these percentages.



Graph 49: Elements realizing the adjective phrase category in all the essays

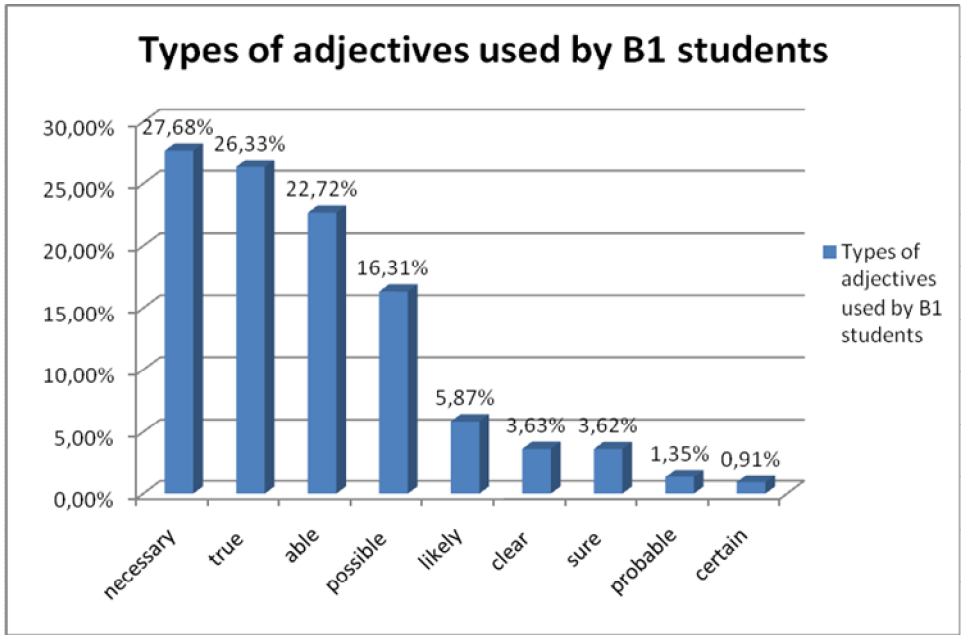
If we look at the usage of adjective phrases at the different proficiency levels, we observe that A2 students prefer *necessary*, with 33.32%. Three other adjectives appear in a similar proportion, *true*, *possible* and *able*, with 22.80%, 21.04% and 21.04%, respectively. The least commonly used element within this category is *sure*, with just 1.75%. As Graph 50 below shows, what is remarkable at this proficiency level is the scarce variety of adjectives employed by students.



Graph 50: Elements realizing the adjective phrase category in A2 essays

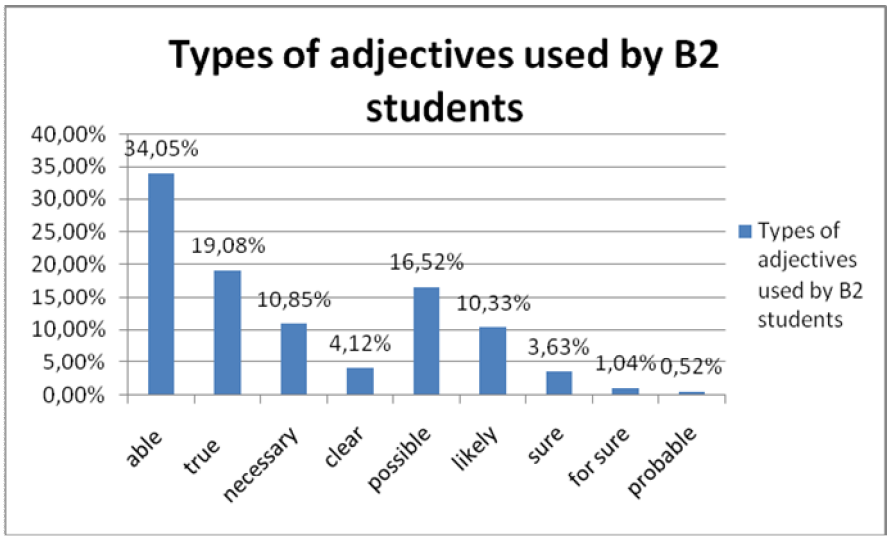
Conversely, B1 students display a wide variety of elements within the class of adjective phrases expressing modality. Three adjectives occur in a proportion higher than 20%, which are *necessary*, *true*, and *able*. Another adjective quite frequent, although lower in percentage, is *possible* (see Graph 51 below for exact figures). Under 6% we can distinguish *likely*, *clear*, *sure*, *probable* and *certain*.

Interestingly enough, the four most frequently used adjectives at this stage, namely, *necessary*, *true*, *able* and *possible*, coincide with the four most common elements used by A2 students.



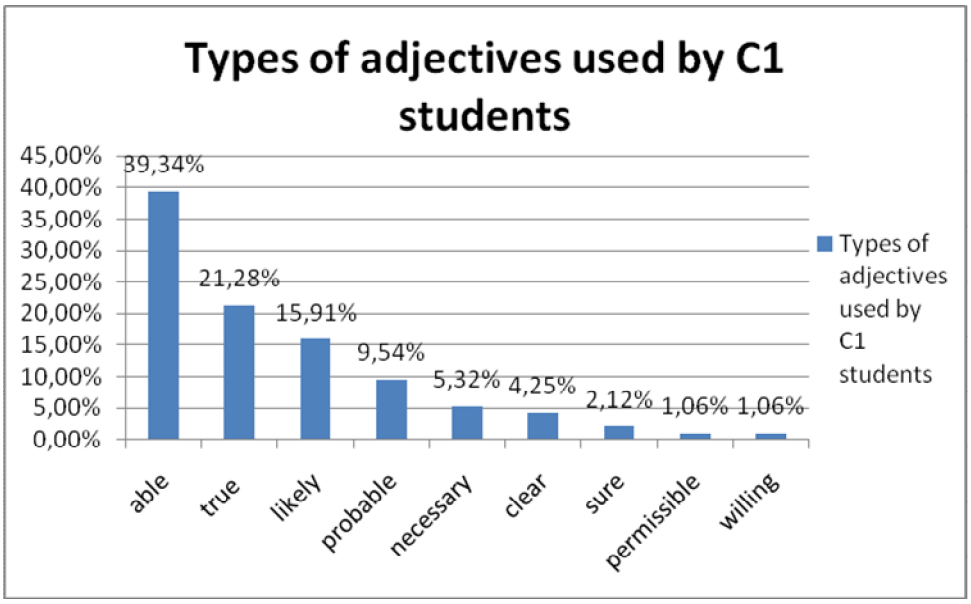
Graph 51: Elements realizing the adjective phrase category in B1 essays

Likewise, these four adjectives constitute the most frequent choice made by B2 students, since *able* represents 34.05% of all the class; *true*, 19.08%; *possible*, 16.52%; and, *necessary*, 10.85. To these we should add *likely*, which is used in a proportion of 10.33%. With percentages below 5% and, therefore, the least frequent choices made by B2 students are *clear*, *sure*, *for sure* and *probable*. Graph 52 offers the percentages in detail.



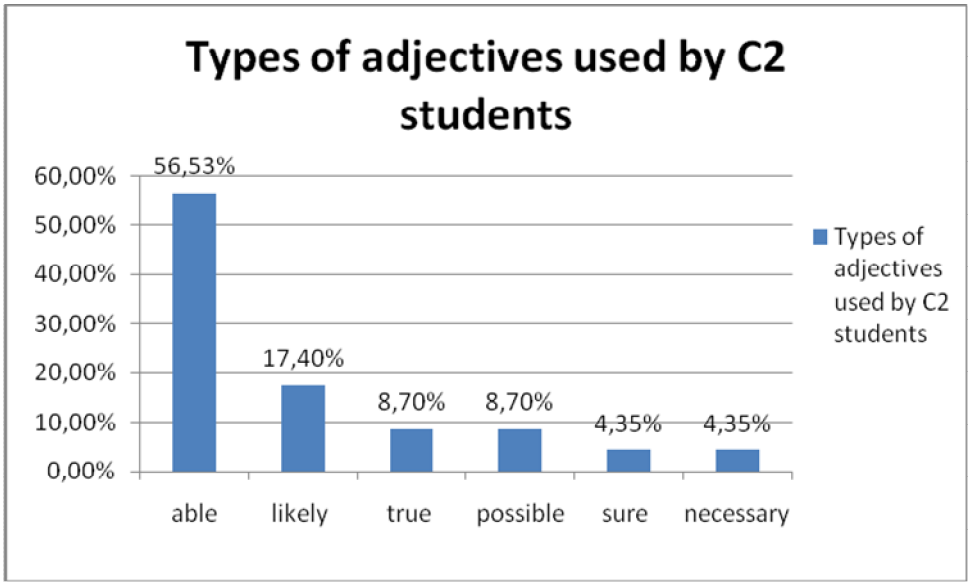
Graph 52: Elements realizing the adjective phrase category in B2 essays

Regarding C1 students, they also follow the general tendency observed in the lower proficiency levels, since they employ *able* in the largest proportion, with 39.34%. Also, *true* is very frequently used, with 21.28%. It is closely followed by *likely* with 15.91%. With less than 10% of occurrence we can find *probable*, with 9.54%; *necessary*, 5.32%; *clear*, 4.25%; *sure*, 2.12%; *permissible*, 1.06%; and *willing*, 1.06%, as Graph 53 shows:



Graph 53: Elements realizing the adjective phrase category in C1 essays

Even the most advanced learners, C2, show the same tendency to use *able* predominantly. In this case, it represents 56.53% of the total class of adjectives. By far it is followed by *likely* with 17,40%; *true*, with 8.70%; and *possible*, 8.70%. *Sure* and *necessary* constitute the least representative tokens of this category, both with 4.35%. Graph 54 below displays all these figures.



Graph 54: Elements realizing the adjective phrase category in C2 essays

5. DISCUSSION

As it has been mentioned above, if we look at the total amount of modal markers used in general, regardless of their grammatical nature, the figures show a decreasing use across proficiency levels. That is to say, as the students' linguistic competence improves, they use fewer modal expressions. Moreover, this falling usage across proficiency levels is quite constant over the main levels –intermediate, B1 and B2, on the one hand, and advanced, C1 and C2, on the other-, which experience a similar fall of around 2% or 1.5% with respect to the proportion employed at the immediate lower level.

Several reasons might be offered to explain this unexpected trend. First, it could be argued that as learners become more proficient in English, they get more certain in their claims or, at least, they show a higher degree of confidence in their argumentation. However, as I will explain below in detail, the type of elements most frequently used by the students at the different proficiency stages examined lead us to discard this possibility.

I will here briefly refer to one of the most salient categories, modal auxiliaries, where the most recurrent element used across proficiency levels varies from CAN –which represents the highest percentage at A2, B1 and B2- to WOULD –the preferred modal auxiliary in the advanced levels, C1 and C2. This, besides the remarkable increase in the use of SHOULD, suggests that students become more tentative and indirect in their assertions as their proficiency in English rises. This is logical if we take into account that English is a more indirect language than Spanish. Thus, the students' linguistic improvement involves not only a progressive mastery of the English grammar but also of its pragmatics.

Second, we could consider two variables to account for the decreasing use of modal markers as students' linguistic competence improves, which are the notions of genre and topic (Martin and Rose, 2008; Eggins, 2004). As I have mentioned in the method section, all the essays collected in the WriCLE corpus are argumentative essays and, in particular, they belong to the well-established genre of academic writing. Within it, we can further distinguish between personal opinion and discussion essays.

Hence, if as students' level of proficiency increased, they wrote more discussion essays, it would have an impact on the amount of modality used, since discussion seemingly involves less evaluation on the part of the speaker –and more of other people's views- than providing their personal opinion about a given topic. Likewise, the topic chosen or the kind of question used to elicit students' answer might also influence the quantity and type of modal markers employed, like many studies suggest (Hinkel, E., 1995; Thompson, P., 2000).

However, a close examination of these two variables in my data indicates that the proportion of discussion and personal opinion essays is identical across proficiency levels, which is evenly distributed, and the topics chosen or the questions students answer are exactly the same at any level of proficiency. Thus, these two variables do not play any role here.

Third, another argument to be considered is the apparently reverse relationship between the mean number of clauses per essay written across proficiency levels and the proportion of modal tokens employed at the different proficiency levels. Thus, those levels with fewer clauses per essay (from the lowest to the highest mean numbers, A2, B1, B2) display the highest proportions of modal tokens per clause, with A2 on top, closely followed by B2 and, then, by B1.

By the same token, C1 and C2 -which have the highest number of clauses per essay on average- yield the lowest proportions of modal tokens per clause. Moreover, within the advanced levels, C1, which has the highest mean number of clauses, shows the lowest percentage of modal tokens. It seems to me that these opposing but related tendencies are not accidental. On the contrary, I strongly believe that, in order to account for the decreasing proportion of modal markers used as students' linguistic competence improves, the average number of clauses written per essay at the different stages should be taken into consideration.

And this is due to the fact that the proportion of modal markers employed by the students is diluted as the number of clauses written increases. It should be taken into account that the more complex and sophisticated the argumentation, the more clauses it entails. However, the amount of modality used might not increase in the same proportion, since the speaker's evaluation on the arguments they put forward –which might be expressed by a variable number of clauses across proficiency levels- may not vary as much in quantitative terms. In this way, if the advanced students employ the same number of modal markers than the rest, in absolute terms, but they write more clauses to support their arguments, their proportion of modal markers will be lower.

The usage of non-finite clauses could also have an influence in the general decreasing use of modal markers. Modality is closely connected to the finite element of the clause, therefore, if

students preferred non-finite to finite clauses as their linguistic competence improves, it would have an impact on the total amount of modality they use. As I will explain below in detail, the students do show a higher usage of non-finite clauses as their proficiency level rises, which, thus, can help to explain this general fall in the total amount of modality students employ as their level of English improves.

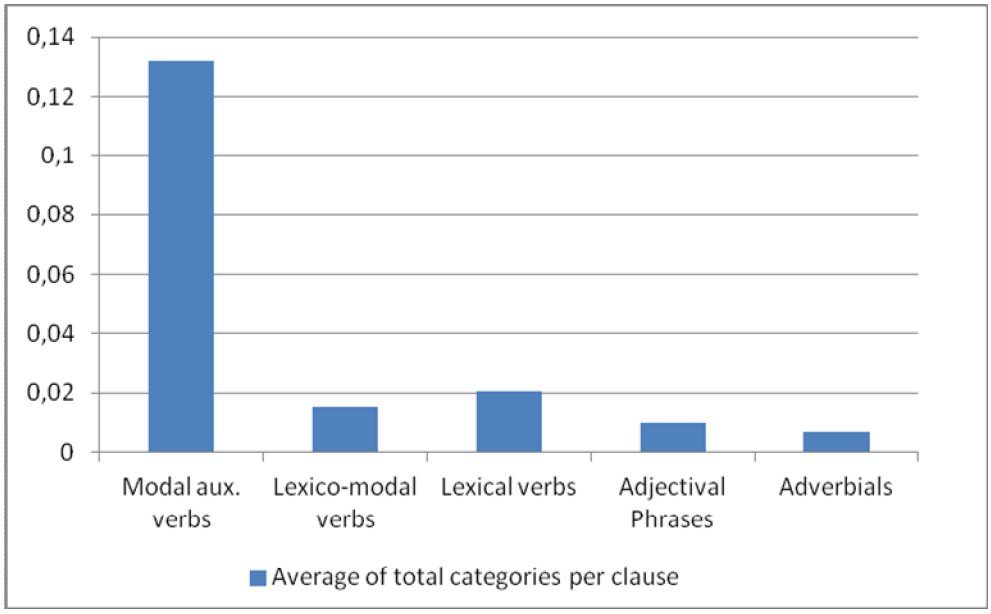
Closely related to this, another element to be mentioned here is that the grammatical elements capable of expressing modality that I have included in my analysis are the most straightforward ones. As it follows, it could well be the case that, as students' level of proficiency rises, they might use more subtle forms to express modality, which are not captured by my model.

For instance, they could be moving from the verbal and non-verbal categories I have analyzed to noun phrases, through the more complex grammatical process of nominalization. Or they could be using more reporting verbs to present other people's viewpoints on a particular topic, or resorting to mental state predicates different from the ones I have considered here to offer their own views. Also, they could be making use of some of the disjuncts and subjuncts I have mentioned in the method section but have not included in my analysis. Finally, they could be resorting to expressions pertaining to the domains of usuality or evidentiality, which lie beyond the notion of modality that guides this study (see section 2).

Besides the decreasing use of total modal markers -as students improve their linguistic competence, another tendency to be highlighted is the overwhelming preference for verbal modal markers at any proficiency level. As a glance at the breakdown of verbal categories suggests, this is mainly due to the learners' remarkable preference for modal auxiliary verbs. With the mere purpose of illustrating the enormous difference between the usage of modal auxiliary verbs and all the other categories, either verbal or non-verbal, I will provide again their average proportion

per clause, regardless of proficiency levels, and a graph that visually represents this enormous difference (see Graph 55 below).

Modal auxiliary verbs constitute an average of 0.1320 per clause, whereas all the other categories are well below, for instance, lexical verbs on average represent 0.0204 per clause; lexico-modal verbs, 0.0153; adjective phrases, 0.0099; and, finally, adverbials, 0.0069. Consequently, learners' patterns of use comply with the views on modality that authors such as Halliday have and that I have summarized in section 2.



Graph 55 Average proportion of modal categories per clause regardless of proficiency levels

Nevertheless, as was noted in section 4, along with this undeniable predominance of modal auxiliary verbs at any level of proficiency, two opposing tendencies that can be observed are a decreasing usage of verbal modality as the students' level of English improves and an increasing presence of non-verbal modal markers. These trends can be distinguished by looking at both their proportions per clause and, also, per total modal markers used.

The only remarkable exception here is A2 students who do not follow the general tendency regarding the use of non-verbal modal resources, since they show the highest usage, instead of the lowest, as would be expected. As has already been mentioned, the low number of A2 essays contained in WriCLE could be the most likely explanation for this unexpected result. But what is important here is that, generally speaking, as students' proficiency level rises, there is a movement away from verbal to non-verbal elements to express modality.

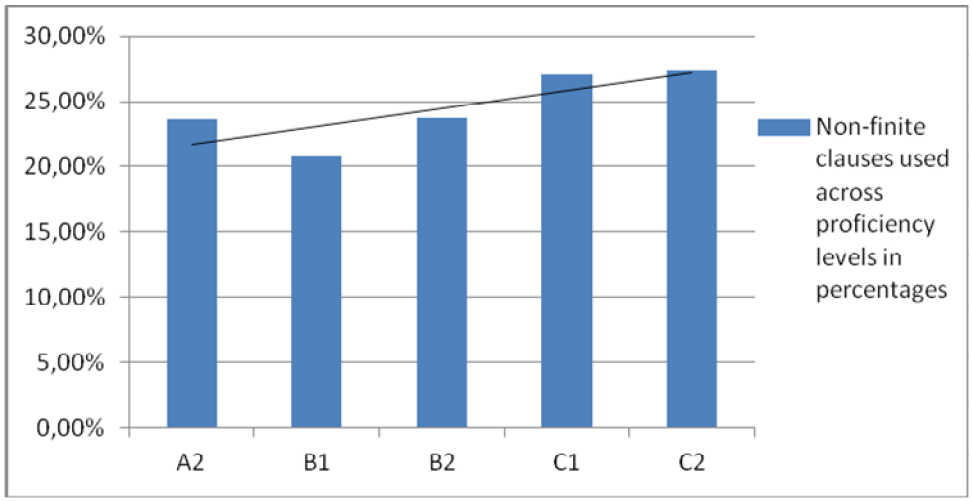
Within the verbal categories established in this analysis as expressing modality, the proportions per clause and per total modal markers of both lexico-modal and lexical verbs display similar patterns of use across proficiency levels. Their usage experiences a slight increase over the lowest levels –A2, B1, B2-, with the highest amount used at B2, and then, their use falls quite drastically at C1, especially in the case of lexical verbs.

As for modal auxiliary verbs, the patterns obtained in their usage per clause and per total modal markers yield opposing tendencies, since the former point at a clear decreasing usage of modal auxiliaries from A2 to C1 –the tendency is not followed by C2 learners-, which would be related to the general fall in the use of modal markers as the proficiency level rises. In turn, the figures per total modal markers used show that among all modal markers students employ, there is an increasing preference for modal auxiliary verbs as they become more proficient in English, even if their actual usage has fallen.

Nevertheless, the fall of modal auxiliaries per clause is very low, since the difference of usage between the highest and the lowest quantities used, that is, between A2 and C1, merely represents 0.0160 of an average use of 0.1320 modal auxiliary verbs per clause.

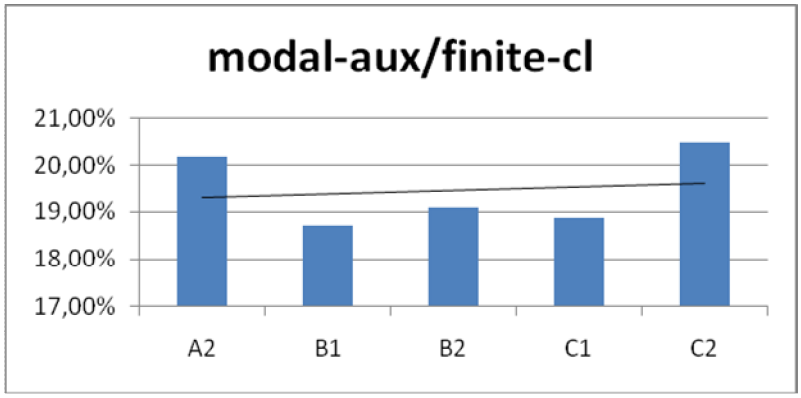
In any case, it should be mentioned that the number of modal auxiliaries per clause may fall as their proficiency rises due to students' increasing preference for non-finite clauses, which do not allow a modal auxiliary. If we look at the figures, the proportions of finite and non-finite

clauses used across proficiency levels show a decrease and an increase, respectively. As a way of illustration, and just focusing on non-finite clauses, the data, from A2 to C2, yield the following results: 23.59%, 20.80%, 23.75%, 27.02% and 27.34%. As can be seen in Graph 56 below, with the only exception of B1 essays, all the other levels show a rising use of non-finite clauses and, consequently, a lower presence of finite-clauses.



Graph 56: Percentage of non-finite clauses used across proficiency levels

To avoid this problem, we can just look at the use of modal-auxiliaries in finite clauses. As Graph 57 below illustrates, the amount of modal auxiliaries employed per finite clauses across linguistic levels shows a minor increasing trend, if we ignore A2. Thus, it suggests the same trend as the proportion per total modal markers used, namely, an increasing, although quite stable, presence of modal auxiliaries across proficiency levels.



Graph 57: Modal auxiliary verbs per finite clauses across proficiency levels

As was noted in the previous section, this verbal class of modal auxiliaries is, by far, the most widely used modal marker at any proficiency level, which fits in with most authors' view that modal auxiliary verbs are the "prototypical realization of English modality" (Coates, 1983; Halliday, 1985; Greenbaum *et al*, 1985; Nuyts, 2001; Palmer, 2001). As should be obvious from Graph 57 above, the only exception here is represented by A2 students, which might be due to experimental error, because of the low number of essays at this level of proficiency included in the corpus.

Concerning non-verbal categories, two clear patterns can be distinguished. On the one hand, adjectival modality shows a severe fall between A2 and B1, no matter whether we look at their usage per clause or per total modal markers. Its use experiences, then, a slight increase and, by B2, its proportion remains quite constant until C2, although it never reaches the percentage displayed by A2 students. On the other hand, regarding adverbial expressions, there is a clear rising pattern in their use, noticeable both in their quantity per clause and per total modal markers used.

Consequently, and in the light of all these patterns of use, we can safely state that, as students become more proficient, there is an increasing, even if mild, use of non-verbal elements and a corresponding fall of verbal modal resources. However, regarding Halliday's dichotomy of

implicit (modal elements forming part of the clause, *eg*. modal auxiliary verbs, lexico-modal verbs and adverbials) vs. explicit modality (modal elements projected as a main clause, different from the clause containing the proposition they qualify, *eg*. lexical verbs and adjective phrases), a different trend can be established (Halliday, 1985; Halliday, 1995; Halliday and Matthiessen, 2004).

The clear fall of lexical verbs among advanced students and the almost identical use of adjectival phrases in the intermediate and advanced levels suggest that there is no preference for explicit means of modal expression as students' proficiency level rises. Conversely, implicit modal resources, such as modal auxiliary verbs and adverbials, do show an increase across proficiency levels. Hence, students display a higher usage of implicit means of modal expression as they progress in their learning process and a lower, although quite stable, usage of explicit modal resources.

If we take into account the tokens employed in each modal class across proficiency levels, we may reach conclusions that could possibly account for the figures just commented. Starting with modal auxiliary verbs, besides an increasing use, we can see a clear evolving pattern from the predominant use of modal auxiliaries such as *can* and *will* at A2 to the more indirect and cautious *would* and *should* at C2, which I have already mentioned.

In the same vein, the results obtained by the SPICLE team[6] in their examination of the use of modal verbs in the writing of Spanish university students (and comparison with native-speakers and non-native learners with a L1 different from Spanish (Neff, J., 2004)) show similar tendencies. Neff *et at.* claim that Spanish students overuse *can* whereas they underuse *could*, *may* and *might*.

[6] A review of their corpus-based studies on modal verbs has been offered in the literature review section above.

However, because they do not take into account the students' level of proficiency as an independent variable, it is not clear whether their students' use of these particular elements varies as their English level increases or not. Even though our purposes and methods are different, it is worth noticing the similar trends put forward by the SPICLE team and me – the Spanish students they analyze overuse *can* in comparison with native speakers while the low-proficient students of WriCLE display the highest quantity of the same modal auxiliary, which decreases as the proficiency level rises (Neff *et al*., 2004).

This pattern shown by the type of modal verbs employed –*would* and *should* overtaking *can* and *will* as the most frequently used modal auxiliary verbs in the advanced levels- along with the increasing use of adverbials as students' level of English improves and, in particular, of epistemic adverbials, suggest a more native-like qualification in students' assertions as they rise in proficiency.

Besides this, another element to be considered is grammatical accuracy. It is evident that learners make mistakes, which are essential to the learning process. As students improve their linguistic competence, it is logical to expect more grammatical accuracy and, therefore, fewer mistakes. Thus, this increasing mastery of the language could account for the decreasing use of modal markers such as lexico-modal verbs.

If we analyze the elements used within this class, we can see that the use of *be going to* falls from 32.35% at A2 to 15.61% at C2, that is to say, it experiences a fall of half its initial proportion. Having a close look at its occurrence in the essays, I realize that in many cases it is wrongly used when stating the purpose of the essay in the thesis statement, instead of the more native-like choice of *will*. Therefore, the behavior of elements like *be going to* would prove that a higher use of modal markers does not necessarily mean a better expression of modality –or rather, more quantity does not necessarily mean more quality.

Also, it is worth noticing the wider variety of tokens within each grammatical class used by students as their linguistic competence improves. In particular, this tendency can be observed in the class of lexico-modal verbs, where along this reduction in the usage of *be going to* as the proficiency level increases, there is also a rising usage of other elements such as *need,* and the increasing presence of a wider range of tokens, like be *supposed to, be obliged to, be bound to, or be required to,* which reaches its peak at C1. The notable exception here is C2 students, who merely use three elements, namely, *have to, be going to* and *need to*. This is perhaps due to experimental error, since this is the least representative proficiency level in my study, with only 18 essays.

This same pattern of an increasingly wider range of elements as students' English improves can also be seen in the class of adverbials. From merely four elements at A2, students display a richer repertoire at B1 with 11 different adverbials, even more at B2, with 13, and, again, the widest range of elements appear at C1, where 15 adverbials can be distinguished. Once again, C2 students break the rising tendency, reducing the variety of adverbials to 8 elements. In the light of these results, it is obvious that as students' proficiency level increases, the variety of elements displayed, at least in certain categories, also grows, regardless of the total proportion of the category. The other categories remain quite stable across proficiency levels.

Finally, within the category of adverbials, the decreasing use of *maybe* is remarkable and, conversely, there is an increasing occurrence of *probably* as the level of proficiency rises. To account for this I should resort to the crucial notion of register (Eggins, 2004) since, as Greenbaum *et al* (1985) point out, *maybe* pertains to an informal register whereas *probably* does not. Hence, the rising use of *probably* and the falling use of *maybe* as the students' linguistic competence improves would involve an increasing awareness and acquisition of English pragmatics. Also, the notion of genre (Martin and Rose, 2001; Eggins, 2004), inextricably

related to register, plays an important role here, since advanced students seem to have a more accurate idea of the conventions of academic writing, in general, and of argumentative essays, in particular.

7. Conclusions

To sum up, a corpus-based analysis of the grammatical expression of modality in the writing of Spanish university students across proficiency levels reveals that, contrary to my expectations, an increasing level of proficiency does not involve the usage of a larger number of modal markers. However, a close examination of the grammatical classes of modal markers I have considered in this study, their proportion and evolution across proficiency levels suggest that more quantity does not necessarily mean an improvement in the expression of modality.

On the contrary, sometimes grammatical accuracy helps us to account for the decreasing use of certain elements, such as certain lexico-modal verbs, as students' linguistic competence improves. As has been mentioned, the number of clauses written by students could also have an influence on the results since the use of the same number of modal markers in a higher number of clauses results in a smaller proportion. Obviously, the advanced students produce more clauses, therefore, this could explain their lower number of total modal markers.

What the data do confirm is my assumption that the higher the level of proficiency, the more ample the repertoire of elements employed in most categories, such as lexico-modal verbs or adverbials. Besides this, as student's proficiency level rises, an increasing acquisition of pragmatics can be observed by looking at the type of elements used in each category.

In particular, the evolution of the type of modal auxiliary verbs (from *can* and *will* at A2 to *would* and *should* at C2) or adverbials (the rising and falling patterns of *probably* and *maybe*, respectively) students prefer across proficiency levels suggests a progressive mastery of English pragmatics, discourse conventions, register and genre, along with a more grammatical accuracy.

In addition, the "prototypical realization" of modality, namely, modal auxiliary verbs (Halliday, 1984; Halliday and Matthiessen, 2004) are the dominant means of modal expression at any linguistic level examined. This, obviously, has an influence in the higher figures obtained for verbal modal means as opposed to non-verbal modal resources at any proficiency level. Nonetheless, across proficiency levels we can distinguish a slight fall of verbal means and a mild rising presence of non-verbal modal expressions, which could be, in part, explained by the increasing occurrence of non-finite clauses as students' proficiency level improves. Consequently, this hypothesis –a movement away from verbal to non-verbal modality- has been confirmed by the data.

A contrastive interlanguage analysis could clarify whether this trend is due to the students' avoidance because of the inherent difficulty of non-verbal modal markers for Spanish learners of English, or due to L1 transfer (as Neff *et al*. (2003) and (2004) suggest in the case of Spanish students overuse of *can*), or to the way in which modality is and has been taught (primarily focusing on verbal modality).

Also, more essays –especially at A2 and C2- should be analyzed in order to confirm or discard some of the conclusions I have put forward in this study, since the lower number of A2 and C2 essays included in the WriCLE corpus might have led to establish merely apparently clear tendencies in the expression of modality. To this end, an appropriate statistical test should be applied in order to unveil the statistical significance of the results.

8. REFERENCES

- Aijmer, K. (1996) "I think – an English modal particle". In Swan, T. & Westvik, O.J. (eds) *Modality in Germanic Languages: historical and comparative perspectives.* Berlin: Walter de Gruyter.

- Aijmer, K. (2002). "Modality in advanced Swedish learners' written interlanguage". In Stephanie Petch-Tyson *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins Publishing.

- Bybee, Joan L. and Suzanne Fleischman (1995) *Modality in grammar and discourse*. Amsterdam: John Benjamins Publishing Co.

- Chafe, W. (1986) "Evidentiality in English conversation and Academic Writing". In Chafe, W. & Nichols, J. (eds) *Evidentiality: the linguistic coding of epistemology. Volume XX*. New Jersey: Ablex Publishing Co.

- Chocano G., R. Jiménez, C. Lozano, A. Mendikoetxea, S. Murcia, M. O'Donnell, P. Rollinson, I. Teomiro (2007). An exploration into word order in learner corpora: The WOSLAC Project. In: M. Davies, P. Rayson, S. Hunston, P. Danielsson (eds) *Proceedings of the Corpus Linguistics Conference 2007*, University of Birmingham.

- Coates, J. (1983) *The Semantics of Modal Auxiliaries*. London & Canberra: Croom Helm.

- Council of Europe (2001). *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

- Downing, A. & P. Locke (2002) *A University course in English grammar*. London: Routledge

- Eggins, S. (2004) *An introduction to systemic functional linguistics.* London & New York: Continuum International Publishing Group.

- Granger, S. (1993). "The International Corpus of Learner English", in J. Aarts, P. de Haan, and N. Oostdijk (eds.) *English language corpora: Design, analysis and exploitation*. Amsterdam: Rodopi.

- Granger, S. (1998) *Learner English on Computer*. London & New York: Longman

- Greenbaum, S. (1969) *Studies in English Adverbial Usage*. London: Longman.

- Greenbaum S., *et al.* (1985) *A Comprehensive Grammar of the English Language*. Essex: Longman.

- Jespersen, 1924 in Hoye, L. (1997) *Adverbs and Modality in English*. London & New York: Longman

- Halliday, MAK. (1985) *An introduction to Functional Grammar*. (1st ed.) Kent: Edward Arnold.

- Halliday, MAK. (1995) *An introduction to Functional Grammar*. (2nd ed.) Kent: Edward Arnold.

- Halliday, MAK and Christian Matthiessen (2004) *An introduction to Functional Grammar*. London: Hodder Arnold.

- Hinkel, E (1995) The use of modal verbs as a reflection of cultural values. *TESOL Quarterly*, 29(2), 325-343.

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

- Huddleston, R.D. (1976) Some theoretical issues in the description of the English verb. *Lingua* 40: 331-83

- Huddleston, R.D. and Pullum (2008) *A student's introduction to English grammar*. Cambridge: CUP

- Lightfoot (1979) *Principles of diachronic syntax*. Cambridge: CUP

- Lyons, J. (1977) *Semantics (I & II)*. Cambridge: CUP

- Martin, J. and Rose (2008) *Working with discourse: meaning beyond the clause*. London & New York: Continuum International Publishing Group

-Neff, J. *et al.,* (2003) "Evidentiality and the construction of writer stance in native and non-native texts". In Hladky, J. (ed.) *Language and Function: to the memory of Jan Firbas*. Studies in Functional and Structural Linguistics, v.49 (2003) Amsterdam: John Benjamins.

-Neff, Joanne, Francisco Ballesteros, Emma Dafouz, Mercedes Díez, Honesto Herrera, Francisco Martínez, Rosa Prieto, Juan Pedro Rica and Carmen Sancho (2003) "Contrasting learner corpora: the use of modal and reporting verbs in the expression of writer stance", in S. Granger and S. Petch-Tyson (eds.), *Extending the Scope of Corpus-based Research*, Amsterdam, Rodopi, 211-230.

-Neff, J., Ballesteros, F., Dafouz, E., Martínez, F. and J. P. Rica (2004): "Formulating writer stance: a contrastive study of EFL learner corpora", in U. Connor and T. A. Upton (eds.), *Applied Corpus Linguistics. A Multidimensional Perspective*, Amsterdam and New York, Rodopi, 73-89.

- Nuyts, J (2001) *Epistemic modality, Language and Conceptualization*. Amsterdam: John Benjamins Publishing.

- O'Donnell, M. 2008. "Demonstration of the UAM CorpusTool for text and image annotation". *Proceedings of the ACL-08:HLT Demo Session (Companion Volume)*, Columbus, Ohio, June 2008. Association for Computational Linguistics. pages 13–16.

- Palmer, F. (1979) *Modality and the English modals*. London and New York: Longman

- Palmer, F. (1986) *Mood and Modality* (1st ed.). Cambridge: CUP

- Palmer, F. (2001) *Mood and Modality* (2nd ed.). Cambridge: CUP

- Palmer, F.(2003) “Modality in English: Theoretical, descriptive and typological issues”. In Facchinetti, Krug and Palmer (eds) *Modality in Contemporary English*. NY-Berlin: Mouton de Gruyter.

- Perkins (1983) *Modal expressions in English*. In Nuyts, J. (2001) *Epistemic modality, Language and Conceptualization*. Amsterdam: John Benjamins Publishing

- Plank (1984). *The modals story retold*. In Nuyts, J (2001) *Epistemic modality, Language and Conceptualization*. Amsterdam: John Benjamins Publishing

- Rollinson, P. & Mendikoetxea, A. (2010). Learner corpora and second language acquisition: Introducing WriCLE In: J. L. Bueno Alonso, D. Gonzáliz Álvarez, U. Kirsten Torrado, A. E. Martínez Insua, J. Pérez-Guerra, E. Rama Martínez & R. Rodríguez Vázquez (eds.) *Analizar datos>Describir variación/Analysing data>Describing variation*. Vigo: Universidade de Vigo (Servizo de Publicacións), pp. 1-12.

- Selinker, L. IRAL - International Review of Applied Linguistics in Language Teaching. Volume 10, Issue 1-4, Pages 209–232,

-Soufleros, E. (2010) “Patterns of modality in argumentative essays by L1 Greek and L1 Spanish advanced EFL learners.” Home page. Web. 18 July, 2011.

- Talmy and Sweetser in Nuyts, J. (2001) *Epistemic modality, Language and Conceptualization*. Amsterdam: John Benjamins Publishing.

- Thompson, P. (2000) “Modal Verbs in Academic Writing”. In Kettemann, B. and G. MARKO (Eds.) *Language and Computers, Teaching and Learning by Doing Corpus Analysis*. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz, 19-24 July, 2000, pp. 305-325(21). Amsterdam: Rodopi.

- UCLES (2001). *Quick Placement Test (Paper and pencil version)*. Oxford: Oxford University Press.

Appendix 1:

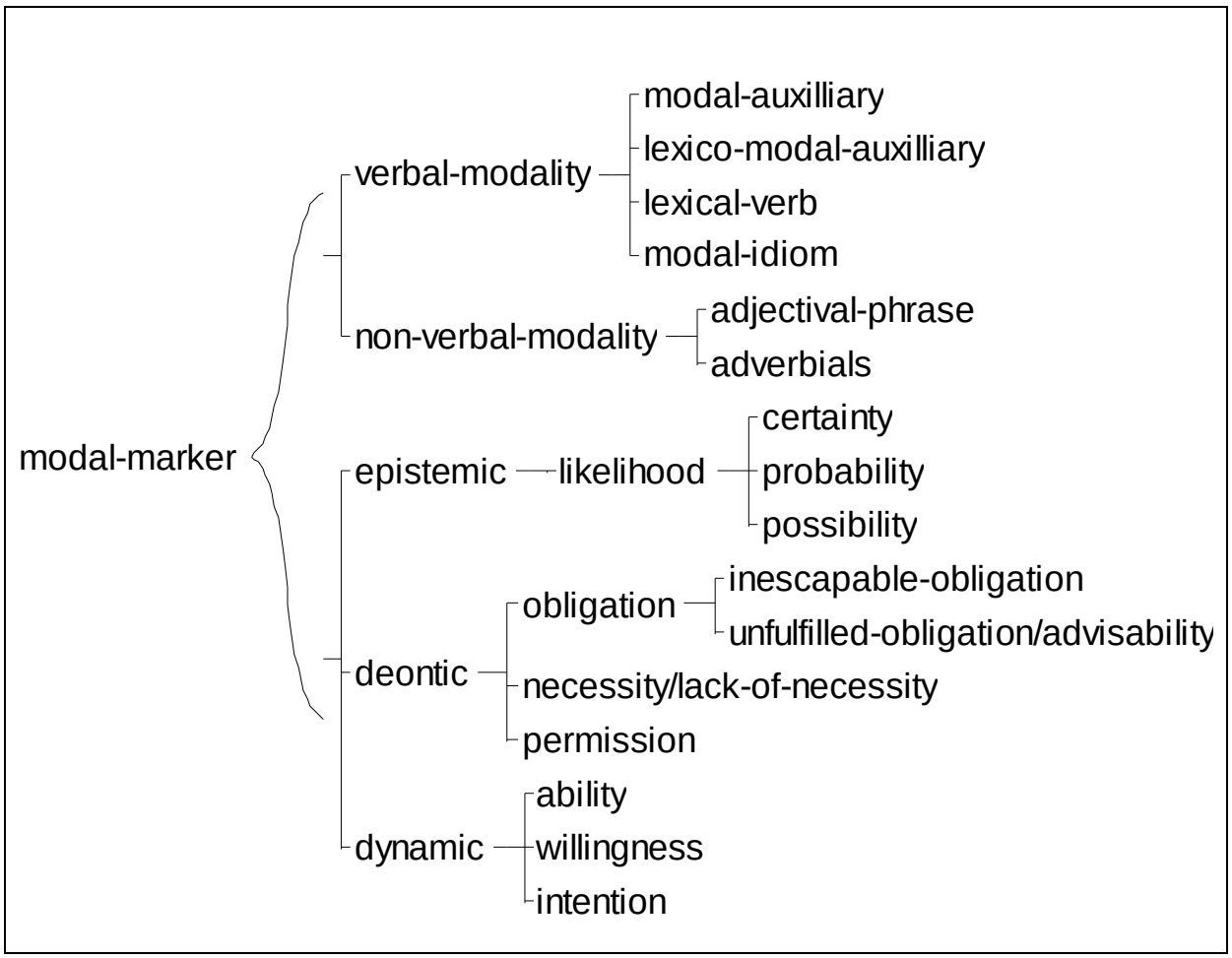
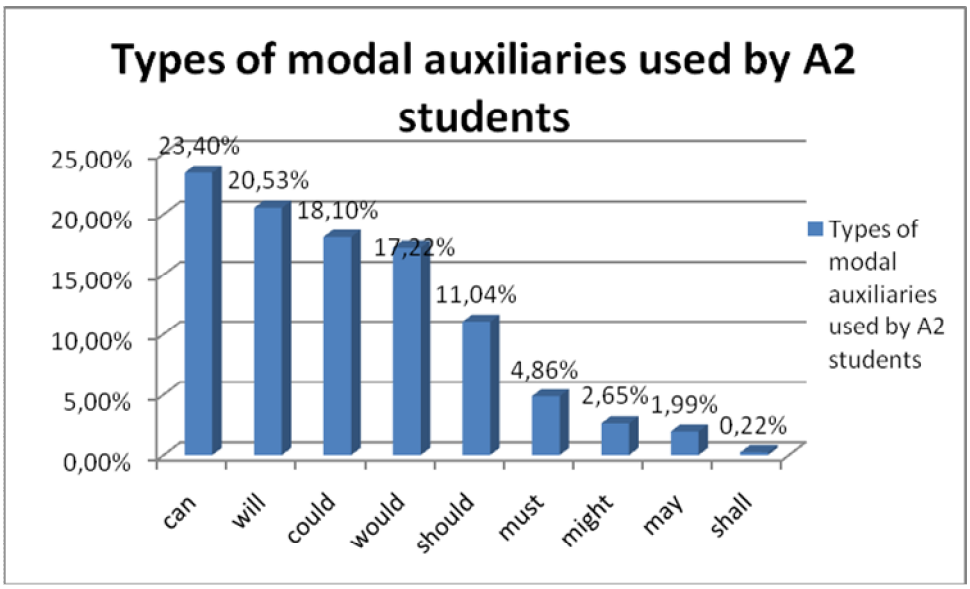
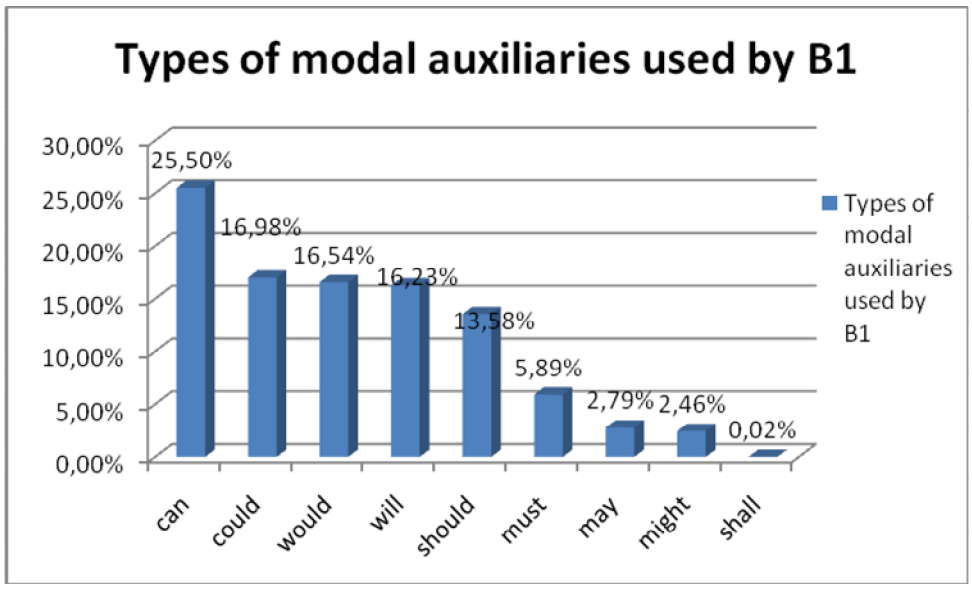


Table 11 The full network of modality used in this study

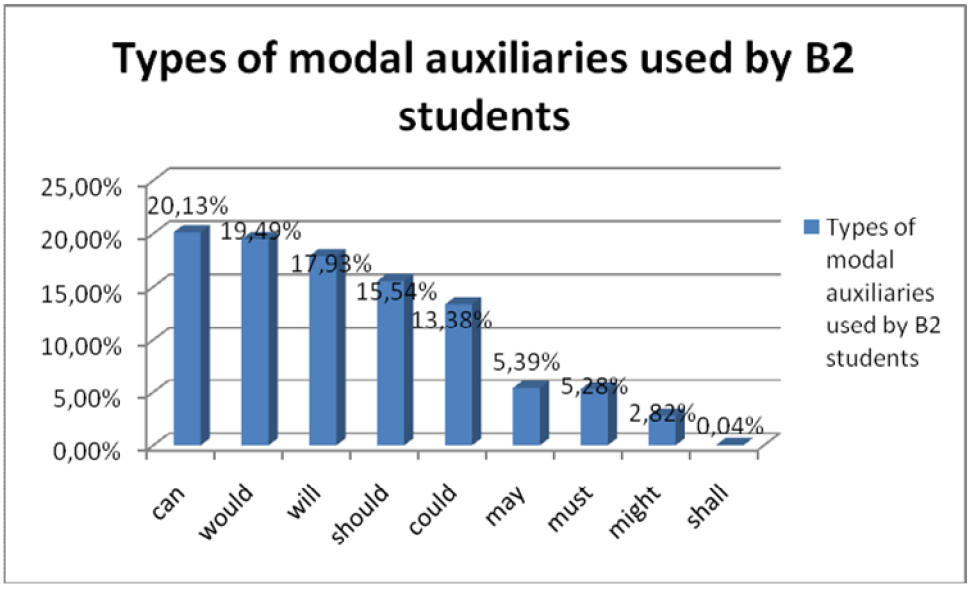
Appendix 2:



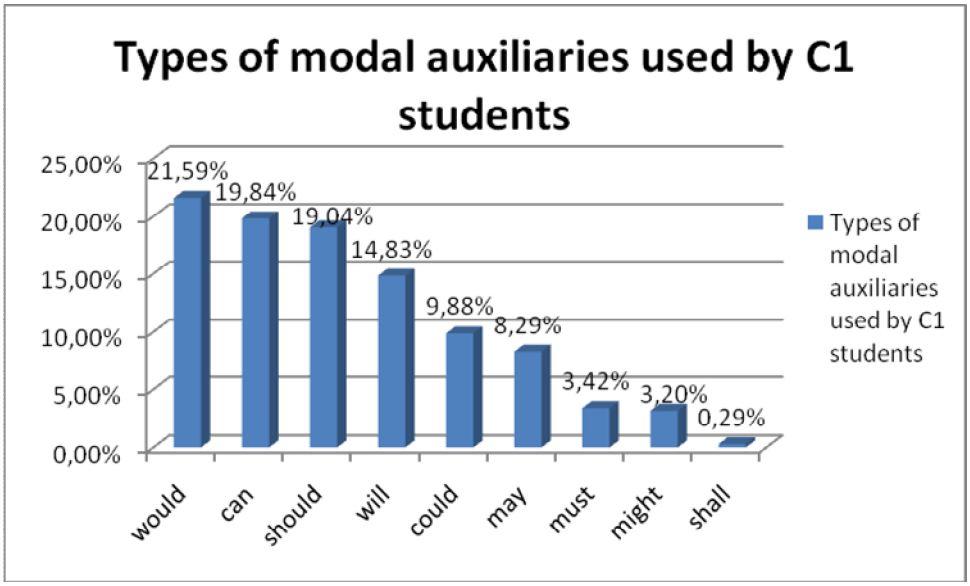
Graph 58: Elements realizing the modal auxiliary verb category in A2 essays



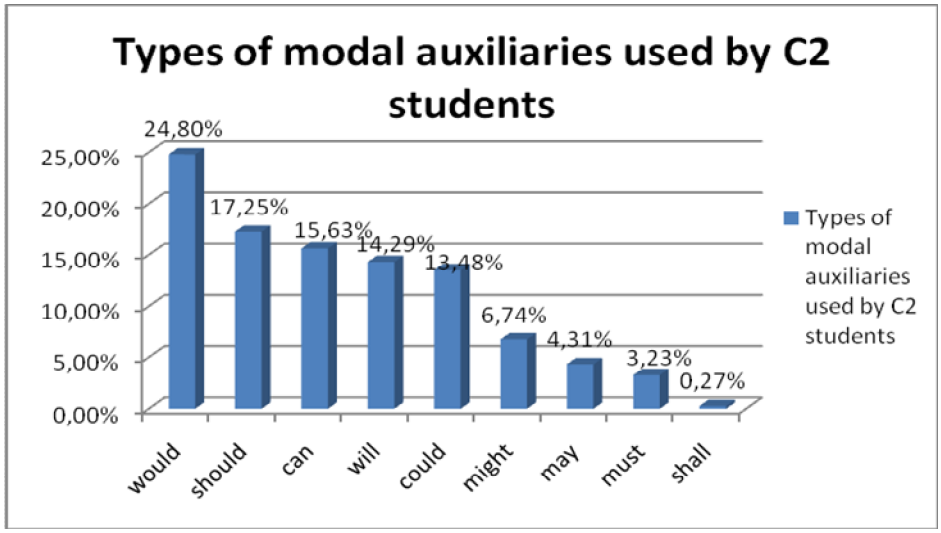
Graph 59: Elements realizing the modal auxiliary verb category in B1 essays



Graph 60: Elements realizing the modal auxiliary verb category in B2 essays



Graph 61: Elements realizing the modal auxiliary verb category in C1 essays



Graph 62: Elements realizing the modal auxiliary verb category in C2 essays