

Epistemic modality through the use of adverbs: a corpus-based study on learners' written discourse

Modalidade epistêmica por meio do uso de advérbios: um estudo baseado em *corpora* sobre o discurso escrito de aprendizes

Adriana Tenuta*
Ana Larissa A. M. Oliveira**
Bárbara Malveira Orfanó***

ABSTRACT: This paper discusses the grammatical category of modality and the variety of linguistic resources available for the expression of it, and presents a research that aimed at analyzing the expression of modality through the use of adverbs in academic writing. More specifically, the study presented investigated how Brazilian learners of English express modality through adverbs in their academic essays. Two corpora were used: a sub-corpus taken from the corpus of Brazilian Learners of English (CABrI) and another sub-corpus taken from the Louvain corpus of Native English Essays (LOCNESS). The prevalent adverbial items found in both corpora were identified and described, using corpus Linguistics tools. The analysis conducted revealed the rigidity of the expression of modality through adverbs in the learners' written discourse as opposed to a more varied way of this expression in the native speakers' data. This paper also discusses the way native speakers and learners differ in their written production and the possible pedagogical implications of these findings.

KEYWORDS: Corpora. Modality. Adverbs. Syntax. Learners' writing.

RESUMO: Este artigo discute a categoria gramatical da modalidade e a variedade de recursos linguísticos disponíveis para a expressão desta. O artigo apresenta uma pesquisa que teve como objetivo analisar a expressão da modalidade por meio do uso de advérbios na escrita de aprendizes. Mais especificamente, o estudo apresentado investigou como aprendizes brasileiros de Inglês expressam modalidade por meio de advérbios em redações produzidas em ambiente acadêmico. Foram utilizados dois *corpora*: um *sub-corpus*, retirado do *corpus* de Aprendizes Brasileiros de Inglês (Cabri) e outro *sub-corpus*, retirado do *corpus* Louvain de Redações em Inglês escritas por nativos (LOCNESS). Os itens adverbiais prevalentes em ambos os *corpora* foram identificados e descritos, usando ferramentas de Linguística de *corpus*. A análise realizada revelou a rigidez da expressão da modalidade por meio de advérbios na escrita dos aprendizes, bem como uma forma mais variada desta expressão nos dados de falantes nativos. Este artigo discute, ainda, a forma como os falantes nativos e os aprendizes diferem em sua produção escrita e as possíveis implicações pedagógicas desses resultados.

PALAVRAS-CHAVE: *Corpora*. Modalidade. Advérbios. Sintaxe. Escrita de aprendizes.

* Universidade Federal de Minas Gerais.

** Universidade Federal de Minas Gerais.

*** Universidade Federal de São João del Rei.

1. Introduction

This paper reports on a study that investigated, using corpora, how Brazilian learners of English express modality through adverbs in the production of essays. Corpus-based studies on learners' production of written discourse have caught the attention of many researchers from different domains. Despite the difficulties in compiling and analyzing students' production, recent findings have contributed to the understanding of these students' interlanguage by identifying linguistic features that are prevalent in their discourse (BERBER-SARDINHA; SHEPERD, 2008; DUTRA, 2009).

Following the Hallidayan model (HALLIDAY, 2004), modality conveys stance and attitude of the sender of a message. In this study, then, we use a learners' corpus, aiming at identifying how Brazilian learners of English express stance and attitude by employing modality elements containing an adverb in their academic writing. We shall compare their production to that of native speakers of English in the same setting, that is, in the academic writing scenario. By identifying the most used adverbs in the expression of modality in learners' essays, we may have a better account of these speakers' expressions of stance and attitude. For that purpose, two corpora were investigated: our reference corpus, CABrI (Corpus of Brazilian English Learners, in construction – BERBER-SARDINHA, 2001; DUTRA, 2009), and LOCNESS (Louvain Corpus of Native English Essays - GRANGER; DAGNEAUX; MEUNIER; PAQUOT, 2009). We believe that such an approach to the study of modality in English can contribute to the emerging area of corpora as well as to the study of syntax.

The structure of this paper is the following: introduction, literature review and theoretical framework, analysis, and conclusion.

2. Theoretical framework

For a better understanding of how modality is conceptualized, we have structured the theoretical framework as follows: (1) outlining the main characteristics of modality, (2) describing the realization of modality in English and (3) discussing the interface between corpora and grammar in academic writing by reviewing previous research in the area.

2.1 The Expression of modal values in English

Mood and other modality resources are means for the expression of the speaker's attitude or commitment regarding the content of a proposition (PALMER, 1974). According to Palmer, mood is realized by verbal morphology, whereas modality is a feature related to a variety of linguistic phenomena, as described by Downing and Locke (2006), among which modal verbs play a central role. Modality is to be understood as a grammatical category that covers notions such as possibility, probability, necessity, volition, obligation and permission.

Modality, therefore, can be connected to basic logical meanings, categorized under a few types: (a) epistemic, (b) deontic and (c) dynamic (DOWNING; LOCKE, 2006), of which the first two (epistemic and deontic) are the central ones.

Epistemic modality is the expression of the various degrees of certainty/uncertainty about facts, events, situations, and, thus, it is related to limitations on the speaker's knowledge about these same facts, events, situations. Consequently, epistemic modality refers to meanings related to inference, prediction, expectation, and probability (BIBER: 1999; DOWNING; LOCKE, 2006). Epistemic modality expressed through different means is illustrated below:

1. *It might rain tomorrow.* (modality realized by a modal verb)
I expect that he be happy. (modality realized by *expect* + an embedded clause in the subjunctive)
It's very unlikely that they will accept our offer. (modality realized by an adverb)

Deontic modality, on the other hand, refers to meanings such as permission and obligation of various kinds, ranging from very strong to a milder obligation. Thus, deontic modality, differently from epistemic, is, associated with authority and judgment, rather than with knowledge or prediction. For this reason, deontic modality comprises language resources used to influence people to do (or not to do) things, whereas epistemic modality is used to express what speakers think is likely to happen.

In spite of the fact that epistemic and deontic meanings are different, the same modal verbs can be used in the expression of one or the other, depending on the context given.

2. *It must have been him.* (epistemic)
You must leave now. (deontic)

Additionally, on many occasions, it is necessary to consider the context of use for the precise interpretation of modality meanings. This is the case of the example below, in which *must* can express either the epistemic meaning of prediction or possibility (contextualized as: *I assume you are patient, given certain evidences*) or the deontic meaning of obligation or necessity (contextualized as: *there is a need for you to be very patient, according to my understanding of the situation*).

3. *You must be very patient.*

Although modality is centrally related to epistemic or deontic meanings, as we have stated, there are also other kinds of meanings associated with modality, all of them, however, play a more peripheral role in syntax and are grouped under the label *dynamic*.

These dynamic meanings are described as ability and courage (DOWNING; LOCKE, 2006) and ability, volition and courage (HUDDLESTON; PULLUM, 2005). They are often expressed by modal verbs, like *can* and *will*, and by semi-modals, like *dare*.

Some examples of dynamic modality are displayed below:

4. *I can speak Spanish.* (ability)
I daren' t say this. (courage)

In certain cases, we can interpret the same occurrence as dynamic and as epistemic, since both types of meanings can be identified in the occurrence:

5. *You can 't be right.* (probability and/or ability)
She can play the piano. (possibility and/or ability)
I can speak four languages. (possibility and/or ability)

Modality also conveys meanings related to the concept of remoteness, illustrated in the examples below by Huddleston and Pullum (2005).

6. *If she liked the place, she would have stayed.* (remote)

There are authors that group modality meanings differently. Biber (1999), in this corpus-based reference grammar, identifies three categories of modal verbs: (a) permission/possibility/ability - *can, could, may, might*; (b) obligation/necessity - *must, should, (had) better, have (got) to, need to, ought to, be supposed to*; and (c) volition/prediction - *will*,

would, shall, be going to. This categorization does not correspond exactly to the distinction deontic/epistemic adopted in this work.

From the perspective of Biber (1999), modals are divided into three groups, namely, modals, marginal auxiliary verbs and semi-modals. The first group encompasses *can, could, may, might, shall, should, will, would* and *must*. These modals have a number of specific features, such as (a) being invariant forms, (b) preceding the subject in yes-no questions and (c) being followed by a verb in the bare infinitive. Marginal auxiliary verbs correspond to *need (to), ought to, dare (to)* and *used to*. According to Biber (1999), these marginal auxiliary verbs are rare and occur almost only in British English. Fixed idiomatic phrases, such as (*had*) *better, have to, (have) got to, be supposed to* and *be going to*, are called semi-modals by Biber (1999). Semi-modals differ from central modals because they can be marked for both tense and person. Besides, they can also occur as non-finite forms.

Taking a multi-dimensional feature of modality into account, Carter and McCarthy (2006) state that the best candidates for modality meanings is the closed class of modal verbs, but the 'list' contains others, which are very high in frequency and carry related meanings. These include lexical modals, such as the verbs *look, seem* and *sound*; the adjectives *possible* and *certain*, and the adverbs *maybe, probably, definitely, apparently, and possibly*. The authors also suggest that the domain of modality needs to be expanded beyond the closed class of modal verbs, which is not a new idea, and they provide compelling evidence of the ubiquity of modality items in everyday spoken and written discourse.

The varied expression of modality is presented below, based on Downing and Locke (2006), considering two basic situations: modality expressed in the verb group and expressed elsewhere in the clause

When expressed in the VG, modality can be realized by:

- The modal verbs *may, might, should, must, can, would, will, ought to, shall, could, need*; the semi-modals (modals in certain uses): *need, dare, wish*
- The lexical auxiliaries (chain-like structures with primary verbs *be* and *have*): *be able to, be apt to, be due to, be going to, be liable to, be likely to, be certain to, be sure to, be to, be unlikely to, be supposed to, have to, have got to, had better, would rather, would sooner*.
- The phased structures composed of: *need, want, regret, try, manage, hesitate, happen, chance, tend, seem, appear, pretend* (in any tense) + a V in *-ing* or

infinitive; subjunctive forms in embedded clauses, introduced by verbs such as: *expect, suppose, recommend, require, request, suspect, intend, think, guess, assume.*

- The lexical verbs such as *allow, beg, command, forbid, guarantee, guess, promise, suggest, warn..*
- The imperative forms.
- The past tense to indicate remoteness from reality, as in *I thought I'd go along with you, if you don't mind*; and conditional structures, as in *If you went, I would go too.*

Modality expressed elsewhere in the clause, may be found in adverbs and sentence modifiers: *maybe, supposedly, perhaps, possibly*; predicate adjectives: *possible, impossible, likely, conceivable, doubtful, certain, sure, positive* and nouns such as *possibility, probability, chance, likelihood.*

In contrast with modal verbs, adverbs, which are the focus of this study, are numerous in the area of certainty expression. According to Chafe (1986), all adverbs can be considered *evidential*¹ and can be classified along dimensions such as reliability and degree of expectation. *Certainly*, for example, indicates that the speaker expresses his or her assessment of the proposition. *Obviously* and *clearly* could be classified as markers of induction and of courses expresses that something is in line with expectations.

Modality can also be expressed in different points of the clause, concomitantly. Downing and Locke (2006) refer to this realization as modal harmony. According to them, modal harmony can be illustrated by the following example:

7. *I doubt she could possibly have said that.*

2.2 Corpora and grammar

The study of grammar is relatively recent in Corpus Linguistics, since the lexicon used to be the unit of investigation by excellence in early corpora studies. However, advances in

¹ In a broad definition of evidentiality, epistemic modality is a subcategory of evidentiality, which is neither marked for the mode of knowing nor for the source of knowledge and therefore is distinct from evidentiality in a narrow sense. In English, all evidential adverbs are modal.

automatic tagging and parsing, as well as the appearance of reasonably sized corpora containing detailed grammatical annotation have progressively enabled corpus linguists to shift their attention towards genuinely grammatical issues.

Fulfilling the main objective of this paper, we shall discuss relevant corpus based research in academic writing.

2.3 Empiric research on academic writing and modality from a Corpus Linguistics perspective

In the past few years, research on academic discourse has flourished promoting an interesting debate on the language of higher education. Many studies have concentrated on different aspects of how both native speakers and non-native speakers organize their essays.

Following a topic-oriented approach to academic writing, Hinkel (1995) showed how topics could influence the use and distribution of certain modal expressions in learners' production. In essays written in the discipline of Education, Chinese, Japanese and Vietnamese learners overused *must* and *should*. Native speakers, on the other hand, did not employ *must* when talking about political and educational issues. Hoyo (1997) conducted an experiment with Spanish learners of English and native speakers of English on how modal values are expressed in English. The author designed a series of tests in which the participants were required to fill in gaps in a text with modals and/or adverbs. He observed that learners had trouble combining modal and adverbs to express attitude. Learners did not perceive the combinatory potential of modal expressions, which in turn hindered their performance in academic writing.

Biber (1999, 2001, and 2004) has contributed immensely to our understanding of academic discourse. Following a frequency-driven approach, Biber and his colleagues identified the most frequent bundles in academic discourse mainly describing their grammatical characteristics.

Hyland (2008) explores the structure and function of four word-bundles in a corpus of academic discourse. The data for his study consists of three corpora (research articles, doctoral thesis and master's dissertations) comprising 3.5 million words. The author suggests that the presence of certain bundles, for example, *as a result of* can help identify different text genres. The results indicated that students draw on different resources to develop their arguments. There were fewer lexical bundles in doctoral thesis, while in master's

dissertations, a greater number of bundles were found. This might be an indication that less proficient students rely more on formulaic expressions due to restrict vocabulary.

In another study, focusing on the field of second language academic writing, Hyland and Milton (2010) show how L2 writers differ significantly from native speakers in that the former group relies on a more limited range of items, offering stronger commitments, and exhibiting greater problems in conveying a precise degree of certainty. The authors also found that, in comparison to the L1 writers, the L2 writers rely on a more limited range of grammatical resources, including particular modal verbs and the expression *I think*. The authors posit that this distinction may be a consequence of non-native speakers' limited language repertoire, which does not enable them to adjust different levels of stance when building up their writing.

In a similar direction and adopting a frequency-driven approach, Chen and Baker (2010) identify the most frequent lexical bundles in three corpora: a) a sub-corpus from FLOB (academic prose section), b) BAWE-CH (Chinese students of English), and c) BAWE- EN (English students). Their comparative study showed that there are differences and similarities between native speaker and learner academic writing. The use of lexical bundles in non-native and native student essays, for example, is very similar: from a structural point of view, they both have more verb phrase based bundles and discourse organizers than native expert writing whereas native professional writers exhibit a wider range of noun phrase based bundles and referential markers.

Following a pragmatic-functional approach, Simpson-Vlach and Ellis (2010) also looked at the most common lexical bundles in academic discourse, focusing on both oral and written corpora. They used the Michigan Corpus of Academic Spoken English (MICASE) and the oral academic part of the British National Corpus (BNC) also including in their research the Hyland Corpus (2004) and the written BNC files of various academic subjects. They extracted three and four word *n-grams* and had ESP instructors judge if the lexical bundles were chunks, if they had a function or if they were expressions that were worth teaching. As a result, they proposed the Academic Formulas List (AFL) with 435 lexical bundles distributed in 18 subcategories. However, research on the expression of modality in academic writing is scarce: the next paragraphs will outline previous studies in the area that prove to be significant to our analysis.

3. Data and Methodology

As we have already mentioned, this study uses two corpora: a sub-corpus taken from the corpus of Brazilian Learners of English (CABrI) and another sub-corpus taken from the Louvain Corpus of Native English Essays (LOCNESS). CABrI is composed of academic essays written by advanced undergraduate students (B1 to C1 according to the Common European Framework) from the Language Course at Universidade Federal de Minas Gerais. Students from the Liberal Arts course are asked to write argumentative essays ranging from 300 to 500 hundred words. Students have to choose from 13 different titles such as *Crime does not pay* and *Feminists have done more harm to the cause of women than good* to write their essays. Learners' writings are converted to text files and stored, so corpus methodological tools can be employed for analysis. The texts chosen to compose the sub-corpus belong to the American argumentative section. In total, the LOCNESS sub-corpus used in this study contains 60,241 words. Now, CABrI contains around 36,187 words. The Louvain Corpus of Native English Essays (LOCNESS) presents essays written by American and British speakers, ranging from academic to literary texts.

For this analysis, first, word lists (frequency lists) were generated and adverbs with the potential to function as modal items were isolated. This procedure, according to O'Keeffe, McCarthy and Carter (2007), proves to be essential in identifying the core vocabulary of English, and it is considered a good resource for pedagogical purposes, which is one of the aims of this study. Comparing frequency lists is, then, an appropriate starting point; however, relying only on frequency lists would not be sufficient. For that reason, in order to get a better notion of the pragmatic function of the adverbs involved in the construction of modality in the essays under investigation, the next step was to compile lexical bundle lists containing an adverb with a modal function. After identifying the most common items in this list, concordance lines were analyzed so that the bundles containing modality adverb could be observed in their particular contexts through manual search of the data.

4. Analysis

We started analyzing the most frequent adverbs found in the learner corpus in order to verify their function in the data. These results are presented in the table below.

Table 1- Most frequent adverbs found in CABrI

Item	Raw Freq.	Freq. per million words
<i>probably</i>	13	472
<i>certainly</i>	9	326
<i>maybe</i>	9	326
<i>likely</i>	9	326
<i>simply</i>	7	254
<i>unfortunately</i>	5	181
<i>actually</i>	5	181
Total	44	1,594

As the main aim of this paper is to compare learners' data with native speakers' production, we shall analyze LOCNESS for corresponding results. In table 2, then, we present the most frequent adverbs found in LOCNESS.

Table 2- Most frequent adverbs found in LOCNESS

Items	Raw Freq.	Freq. per million words
<i>likely</i>	27	448
<i>certainly</i>	19	315
<i>probably</i>	17	282
<i>perhaps</i>	16	265
<i>surely</i>	13	215
<i>possibly</i>	12	199
<i>maybe</i>	11	182
<i>unlikely</i>	7	116
Total	122	1,952

Confronting tables 1 and 2, we observe that native speakers seem to use more adverbs than non-native speakers do. In addition, when analyzing the frequency of adverbs in each group, we can observe that the native speakers' use of adverbs is more evenly distributed than that of learners. This first finding is in line with Holmes (1998), is that it provides evidence of a more balanced use of linguistic resources by native speakers to express modality. Throughout this paper, we will make the case that learners tend to use a more fixed set of expressions to convey modality. In table 3, we contrast the adverbs found in both corpora, aiming to determine differences and/or similarities in them.

Table 3- Distribution of adverbs in the two corpora (raw results)

Items	LOCNESS	CABrI
<i>Likely</i>	27	9
<i>Certainly</i>	19	9
<i>Probably</i>	17	13
<i>Perhaps</i>	16	4
<i>Surely</i>	13	1
<i>Possibly</i>	12	0
<i>Maybe</i>	11	9
<i>Unlikely</i>	7	0

Complementing the quantitative analysis, we also submitted the data to statistic tests, believing that they would strengthen the argument that there is a striking difference between native speaker and learner corpora in the use of adverbs expressing modality. This difference is outlined in table 4.

Table 4 - Expected contingency table between both corpora

Adverbs	NS	NNS
<i>Likely</i>	26.3	9.70
<i>Certainly</i>	20.5	7.54
<i>Probably</i>	21.9	8.08
<i>Perhaps</i>	14.6	5.39
<i>Surely</i>	10.2	3.77
<i>Possibly</i>	8.77	3.23
<i>Maybe</i>	14.6	5.39
<i>Unlikely</i>	5.11	1.89

Chi-square = 18.1
degrees of freedom = 7
probability = 0.011

Based on the results of the statistical test, we can affirm that the difference in the use of adverbs in the corpora analyzed is significant. The fact that the *p value* is zero reinforces the claim that there is a striking difference between the corpora analyzed, as the results of the *chi square* (*chi square* = 18.1) have proved. From this preliminary analysis, one can speculate that the use of adverbs to express modality is underrepresented in learners' academic essays, as stated in a research carried out by Tenuta, Oliveira, Orfanó (in press), which has shown that Brazilian learners rely on a rigid set of verbs to express modal values. However, the analysis proposed here intends to go beyond stating statistical differences. In fact, it aims at understanding the linguistic features that make up for these differences and their implications

for the learners' written discourse production. In order to do so, at this point of the analysis, we shall concentrate on the most common bundles containing an adverb in CABrI and LOCNESS. The following table shows this distribution in both corpora.

Table 5- Distribution of bundles containing an adverb in the data

Bundles	CABrI- raw freq	Freq. per m words	LOCNESS-raw frequency	Freq. per m words
<i>likely to</i>	9	33	16	265
<i>Almost certainly</i>	0	0	4	66
<i>certainly not</i>	0	0	4	66
<i>will probably</i>	9	32	0	0
<i>would probably</i>	0	0	6	99

Submitting the data to a *chi-square* test, we found that the difference between the corpora is significant, as illustrated below. Observing the *p value* from the results from table 6, one can verify that when looking, in particular, at bundles containing an adverb, the difference between the two corpora is significant. The items were analyzed and compared with results from LOCNESS as follows:

Table 6 - Expected contingency table

Bundles	CABrI	LOCNESS
<i>Likely to</i>	9.38	15.6
<i>almost certainly</i>	1.50	2.50
<i>certainly not</i>	1.50	2.50
<i>will probably</i>	3.38	5.62
<i>Would probably</i>	2.25	3.75

Chi-square = 23.4
degrees of freedom = 4
probability = 0.000

After identifying the most common bundles in both corpora, we focused on the analysis of each bundle independently. In CABrI, there are only two bundles being used, whereas in LOCNESS, we found four bundles. This fact reinforces the claim that native speakers express modality not only by using different adverbs, but also by combining them in different bundles. The only two bundles used by learners in the corpus analyzed were *likely to* and *will probably*.

4.1 *Likely to*

There are 33 occurrences of the bundle in CABrI and 265 in LOCNESS. This difference was expected since, as previously mentioned, there is lower frequency of adverbs expressing modality in the learners' productions. We considered important to speculate on this situation focusing on the learners' written discourse, bearing in mind that frequency differences between two datasets can indicate either overuse or underuse of linguistic features, which poses interesting pedagogical issues involving the teaching and learning of English.

Extract 1 - Example of *likely to* from LOCNESS²

Now that rail privatisation has gone ahead, many people are likely to lose faith in trains, due to the perceived inefficiency of the operators (for example the timetable book full of errors or the recent survey in Which? magazine about overcharging). Fares are likely to increase, and many rural lines that used to be subsidised by the government face closure.

Fig. 1 shows concordance lines for the bundle *likely to* in CABrI. In all examples, it is possible to see that the epistemic use is prevalent, which might be due to the text genre in focus. In texts of this genre, the writer, very frequently, has to commit him/herself, in different degrees, to the certainty of occurrence of a fact.

N	
1	ished severely and consequently they are more likely to commit crimes again.
2	country side towns around Brazil we are very likely to find a huge number of
3	by creating artificial dreams, graduates are likely to be shocked or unsure in
4	This kind of proficiency is more likely to be developed if one
5	This kind of proficiency is more likely to be developed if one
6	arget our limited resources for programs most likely to reduce recidivism and
7	Not likely to happen.
8	y and dreams be profitable, but they are also likely to be crucial ways to make
9	ot prepare graduate students to what they are likely to face in real life.
10	nt agencies show that such tragedies are more likely to occur to young adults

Fig. 1- Concordance lines for the bundle *likely to* in CABrI.

The function of this bundle is similar in both corpora; however, the frequency is significantly higher in the native speaker corpus. *Likely to* is the bundle containing an adverb preferred by native speakers to express possibility/probability, while English learners expressed possibility/probability by using the bundle *will probably*.

² All extracts from the learner corpus have been preserved as written by students. Hence, corrections by any type are not included in this study.

The low frequency of modal adverbs in the learner's academic writing was previously acknowledged by Tenuta, Orfanó, Oliveira (in press). The authors showed that learners seemed to rely on modal verbs to express epistemic modality. However, in the corpus investigated, learners used a very narrow range of modal verbs with epistemic meaning, for example, mainly *can* and *will*. In this regard, the findings revealed that the distribution of modals in the native corpus was more varied, since native speakers employed, for example, *should*, *could* and *would*.

4.2 Will probably

This bundle follows the pattern *will probably* + verb.³ We tend to conclude that the bundle *will probably* might be more easily accessed by learners, becoming active in discourse through less mental effort (CHAFE: 1994), mainly because *will probably* is lexically and structurally closer to the Portuguese language. In this study, learners used *will probably* three times more than native speakers did.

Extract 2- Example from CABrI

There will probably be many reasons for dreaming and three possible - and believable – ones could be its profitability (for the enterprises which provide entertainment, for example), its help in making us stand and try to change our stressful reality, and the health benefits it provides us. Oddly enough, imagining can make a big profit from generating – and selling – brilliant ideas.

N	Concordance
1	is. All you have to do is read and read. In fact, you will probably get all
2	e she does not recognize. As for Dee's sister, she will probably feel
3	he references (authors, books, concepts, etc) that will probably help
4	e way she could free herself from her family. They will probably go along
5	no chance to see in loco how things really work, it will probably be very
6	e inhabit a modern and industrialised world. There will probably be many
7	ity degrees, such as philosophy and anthropology, will probably have
8	ity degrees, such as philosophy and anthropology, will probably have
9	er mother, Dee is not the same anymore and they will probably be apart

Fig. 2- Concordance lines for the bundle *will probably* in CABrI

In the next sections, we analyzed the bundles found in LOCNESS.

³ We also observed that there was a high frequency in the use of *will probably* + a linking verb, since a third of the occurrences with the bundle *will probably* followed this pattern. This finding will be further investigated.

4.3 Would probably

The use of the bundle *would probably* in LOCNESS might reflect a feature more closely related to formal register, considered adequate to the essay genre. We might speculate that learners do not use it because they might not be very proficient in the use of linguistic features that would enable them to express a degree of both certainty and formality in written discourse, as seen in the concordance lines below. We could argue that the bundle has a twofold discursive function, which was employed by native speakers and ignored by learners in the corpora analyzed.

Extract 3- Example of *would probably* from LOCNESS

If boxing was made illegal, then the sport would still take place, but it would not be under the control of doctors or referees and there would be no proper association to represent rules and regulations. It would probably become even more dangerous. Of course there are advantages banning the sport but they are very small compared to the disadvantages.

1	By banning fox hunting the law would probably drive the spor
2	ulations showed that charities would probably lose out, coll
3	primary school and her mother would probably die while the
4	It would probably become even mo
5	veral thousands of pounds and would probably have to sell u
6	were the highest bidder as they would probably hide it as it

Fig. 3- Concordance lines for the bundle *would probably* in LOCNESS.

5. Conclusion

In this study, we analyzed epistemic modality, which was a major feature in academic essays of both native speakers and learners of English. The choice for epistemic modality is justified by to the fact that this text genre favors this kind of modal expression, used to mitigate the level of commitment of the speaker to the certainty of the facts being stated.

Our findings indicate that the expression of epistemic modality in the native speakers' written production of academic essays, referring to, specifically, the use of adverbs, seems to be more varied than in the learner's production. This can be evidenced, for example, by the non-use, from the part of learners, of expressions such as like *likely to*, which proved to be widely used by native speakers. On the same token, the data analyzed show that the learners' preferred expression, including adverb to express epistemic modality was *will probably*. These findings strengthen the argument that teachers and material developers should introduce students to a more varied set of vocabulary as well as structural choices, so that

their writing skills can be improved towards the production of texts which are better elaborated and more responsive to the grammatical and semantic, but also to the pragmatic demands of genre.

Another important remark about the learners' written production concerns the fact that, in this study, learners seemed to ignore the form *would probably*, which is highly used by native speakers as a way to express remoteness and little commitment to the certainty of facts. This finding also supports the claim that learners tend to rely on a very rigid set of structures to realize modality. It also poses the issue of the role of language instruction and material design to raise students' awareness of how they can structure their discourse in order to comply successfully with the requirements of academic writing.

We would also like to comment on the importance of analyzing empirical language data for a broader understanding of how native speakers and learners can differ in their production and, if that is the case, decide on the best teaching strategies to help learners write more fluently and effectively.

Also, from a pedagogical perspective, research studies like the one conducted by Holmes (1998) have shown that, in general, the use of modality in English has been underrepresented in teaching materials, since this grammatical category is often dealt with through the restricted use of modal verbs. The present study argues along the same lines, as it also shows that students seem to rely on modal verbs instead of using a variety of modal forms, which is present in the native speakers' production.

Finally, we consider it is important that instructors provide students with opportunities to engage in reflection on their own stance, using strategies while writing, as it has also been postulated by Silver (2013), from a critical perspective.

References

BERBER SARDINHA, A.P. O *corpus* de aprendizes Br-ICLE. **Intercâmbio**, v.10, 2001, p. 227-239.

BERBER SARDINHA, A. P. **Linguística de Corpus**. Barueri: Manole, 004.

BERBER SARDINHA, A. P.; SHEPERD, T. An online system for error identification in Brazilian learner English. **Anais do 8th Teaching and Language Corpora Conference**. Lisboa: Associação de Estudos e de Investigação Científica do ISLA-Lisboa, 2008, p. 257-262.

BIBER, D. et al. **Longman Grammar of Spoken and Written English**, 1999.

BIBER, D. Stance in spoken and written university registers. **Journal of English for Academic Purposes**, 5, 2006, 97–116. **crossref** <http://dx.doi.org/10.1016/j.jeap.2006.05.001>

BIBER, D.; CONRAD, S.; LEECH, G. **Longman Student Grammar of Spoken and Written English**. Essex: Longman, 2002.

CHAFE, W. The nature of consciousness. In: **Discourse, consciousness and time**. Chicago & London: The University of Chicago Press, 1994, pp. 26-40.

DOWNING, A.; LOCKE, P. **English Grammar: a University Course**. New York: Routledge, 2006, second edition

DUTRA, D. P. Conscientização linguística através de corpora online. **Caderno de resumos do 17 INPLA**. São Paulo: PUC-SP, 2009, p. 62.

DUTRA S.; DAGNEAUX, E.; MEUNIER, F.; PAQUOT, M. **International corpus of Learner English-Version2**. Louvain-La-Neuve: UCL Presses Universitaires de Louvain, 2009.

HALLIDAY, M. A. K.; HASAN, R. **Cohesion in English**. London: Longman, 1976.

HALLIDAY, M. A. K. **An Introduction To Functional Grammar**. London: Hodder Arnold Publication, Revised Edition, 2004.

HOLMES, R. Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. In **English for Specific Purposes**, 2007, 16: 321-3. **crossref** [http://dx.doi.org/10.1016/S0889-4906\(96\)00038-5](http://dx.doi.org/10.1016/S0889-4906(96)00038-5)

HOYE, L. **Adverbs and Modality in English**. London & New York: Longman, 1997.

HYLAND, K.; MILTON, J. Qualifications and certainty in L1 and L2 students' writing. **Journal of Second Language Writing**, 6(2), 1997, pp. 183-205. **crossref** [http://dx.doi.org/10.1016/S1060-3743\(97\)90033-3](http://dx.doi.org/10.1016/S1060-3743(97)90033-3)

O'KEEFFE, A.; MCCARTHY, M.; CARTER, R. **From corpus to classroom: Language use and language teaching**. Randi Reppen, Northern Arizona University, 2007. **crossref** <http://dx.doi.org/10.1017/CBO9780511497650>

PALMER, F. R. **The English verb**. London: Longman, 1974.

SILVER, M. The stance of stance: a critical look at ways stance is expressed and modeled in academic discourse. **Journal of English for Academic Purposes**, 2003, vol. 2, issue 4: 359-374. **crossref** [http://dx.doi.org/10.1016/S1475-1585\(03\)00051-1](http://dx.doi.org/10.1016/S1475-1585(03)00051-1)

TENUTA, A. M. **Tempo, modo e aspecto verbal na estruturação do discurso narrativo**. Dissertação (Mestrado em Linguística). Belo Horizonte: UFMG, 1992.

TENUTA, A. M. A., OLIVEIRA, A. L. A. M., ORFANÒ B. M.
How Brazilian learners express modality in their writing: a corpus-based study on lexical bundles. São Paulo: **Revista Intercâmbio**, no prelo.

Artigo recebido em: 07.10.2014

Artigo aprovado em: 11.12.2014

Letras & Letras