

# Near-synonyms in Learner Essays: An Analysis Based on *Corpora*<sup>1</sup>

MATEUS EMERSON DE SOUZA MIRANDA

Bacharel em Letras/Inglês com ênfase em Estudos Linguísticos e graduando em Licenciatura/Inglês pela Universidade Federal de Minas Gerais (UFMG).  
e-mail: mateusesm@gmail.com

## 1. Introduction

Several studies in various fields of linguistics have offered a new perspective regarding language use (Biber et al., 1998). Some generalizations are based on empirical data, which consist of the patterns of differences and peculiarities of the use of a specific language for a specific group of people in different situations. Learner Corpora (Granger, 2002) are situated within the non-native varieties of English, which can be broken down into English as an Official Language (EOL), English as a Second Language (ESL) and English as a Foreign Language (EFL) (Figure 1). Additionally, Granger states that Learner Corpora cover the last two non-native varieties of English, since ESL refers to English acquired in an English-speaking environment (e.g. US), and EFL covers English learned in a classroom setting in a non-English-speaking country, such as Brazil, whose citizens speak Portuguese as their first language.

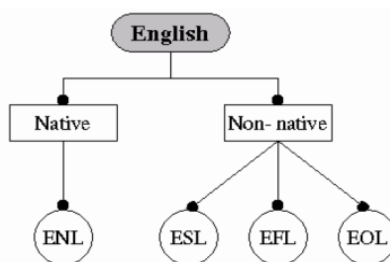


Figure 1. Varieties of English (Granger, 2002, p. 6).

---

<sup>1</sup> Este artigo é um recorte de minha monografia intitulada *Near-synonyms in learner essays: an analysis based on corpora*, sob orientação da Prof<sup>ª</sup>. Dr<sup>ª</sup>. Deise Prina Dutra, apresentada à Faculdade de Letras da UFMG, em 2015, como pré-requisito para obtenção do grau de Bacharel em Letras/Inglês com ênfase em Estudos Linguísticos. Na monografia, os pares *argue(d)* e *discuss(ed)* também foram analisados. O autor agradece as contribuições e esclarecimentos da Prof<sup>ª</sup>. Dr<sup>ª</sup>. Antoinette Renouf, da Birmingham City University (BCU) a esta pesquisa.

Learner corpora research as addressed earlier has some issues. Arts and Granger (2008), for instance, investigated the use of structures with connectors, adverbs, pronouns, and prepositions in non-native corpora. In Brazil, studies have focused on the occurrences of *for-clusters* on argumentative essays (Dutra and Silero, 2009), modality in writing (Tenuta, Oliveira and Orfanó, 2012), and quantifying expressions *a few* and *few* in a Brazilian learner corpus (Silero, 2014). For our study, Biber, Conrad and Reppen (1998) is central as they discuss how synonymous-appearing words are typically used in different ways, such as *to* and *too*. Biber et al. (1998) also present a study of the words *big*, *large* and *great* that thesauruses often consider as synonyms of *size*. In addition, Partington (1998) states that look-alike words are also a problem in translation studies because they are not always reliable translation equivalents.

The use of synonymous words by learners of English has not been widely researched. According to Biber et al. (1998), in the classroom, as in textbooks and lexicographic definitions, some words are characterized as synonymous or identical in meaning, and these may be similar only in some situational and contextual applications. The most commonly observed differences occur in patterns of use and frequency (Biber et al. 1998). Moreover, Moon (2010, p. 206) states that the concept *synonymous* implies interchangeable words in any context, though by contrast, there are very few 'perfect synonyms'; since, as proven empirically, the words which are usually considered synonymous are not to be applied in conveying the same meaning in all contexts (Moon, 2010). Corpora studies allow such claims to be tested, enabling contact with the real language.

Corpus Linguistics (CL) as a research methodology has traditionally offered a wide range of options in the area of Applied Linguistics (AL). However, this research will focus only on one area: the study of lexicon. More specifically, within the lexical aspects, a focus will be placed on near-synonyms misused by non-native English speakers by analyzing their semantic prosody and associated lexical repulsion. Hirst and Inkpen (2006), define near-synonyms as:

Words that are almost synonyms, but not quite. They are not considered fully intersubstitutable, but instead vary in their shades of denotation or connotation, or in the components of meaning they emphasize; they may also vary in grammatical or collocational constraints. (p.1)

It is understood that interlanguage refers to the state of a learner's language as it approximates to the target language (Selinker, 1972). Some inadequacies committed in the use of the word pairs: (a) *worry; concern* and (b) *worried; concerned* can be observed. The general objective of this work is to investigate and describe the use of the aforementioned words by native English speakers and Brazilian learners of the English language. This study takes into account the way these learners use these near-synonyms. More specifically, this study aims at examining the semantic prosody of near-synonym pairs misused by non-native English speakers: (a) *worry; concern* and (b) *worried; concerned*, as well as investigating if they are surrounded by a positive, negative, or neutral environment (Hunston, 2007: 249), and their lexical repulsion in a given context.

## 2. Literature Review

### 2.1. Interlanguage and Corpus Linguistics

Granger (2002) proposed that with the help of CL, studies can be carried on contrastive interlanguage analysis (CIA). According to Granger (2002), through the CIA, "Learners' and native speakers' data, or language produced by learners from different L1 backgrounds, can be compared." The first type of comparison is intended to illustrate not only errors, but also the level of under or overuse of a particular linguistic feature in the second language being learned. The second type of comparison aims to uncover L1 interference or transfer. Corpora data produced by learners from different backgrounds can also be compared between different corpora, with the goal of uncovering common features of the second language acquisition process by discarding specific L1 peculiarities (Granger, 2002, p. 3).

### 2.2 Semantic Prosody

The concept of Semantic Prosody (SP) was first introduced by Louw (1993) as the consistent aura of meaning with which a form is imbued by its collocates. Since this introduction, SP has been discussed, expanded upon, and questioned by many researchers; however, this subsection aims to present some definitions to clarify the analyses presented in the Results section.

According to Sinclair (1991), many uses of words and phrases show a tendency to occur in a certain semantic environment. For example, the verb *happen* is associated with unpleasant events, such as accidents. Sinclair (1996, p. 75), also defines semantic prosody as a functional choice which links meaning to purpose, and all subsequent choices within the lexical item relate back to the prosody. This definition points out three defining features of SP: functionality, linguistic choice, and communicative purpose.

Firstly, functionality occurs when a person chooses lexical items to make sentences, in addition to the lexical and grammatical rules which govern the grammaticality of the sentence. Other choices taken into consideration include the semantic preference and SP which are related to the functions. Secondly, linguistic choice happens when the combination of every collocation is not in the least arbitrary, but all words are in a mutually selective relation. Thirdly, communicative purpose is when the right SP is bound to express the attitudes of speakers/writers and their purpose with harmony and explicitness (Sinclair, 1996, p. 87).

Partington (1998) defines SP as an expansion in connotation that occurs beyond the isolated word. According to Partington, these semantic prosodies can be interpreted taking into account the impregnated connotation between the central word and its placement. The direction as well as prosody is seized from a lexical unit, which is wider than a word. Later on in his research, Partington (2004), classified SP into favorable, neutral, and unfavorable prosodies. A favorable affective meaning was labelled as positive, while an unfavorable affective meaning was judged as negative. When the context provided no evidence of SP, the instance was labeled as neutral.

Finally, Huston (2007) states that the semantic prosody can be apprehended through the interpretation of the text that co-occurs with the central word and, in context, this fact can assign a negative or positive quality to the lexical item example. Stubbs (1996) proposes that some words have a predominantly negative prosody, some of them have a positive prosody, and others are neutral. According to him, if the collocates that a node word attracts are mostly of strong negative semantic characteristics, the node word carries a strong negative prosody. If the collocates are mainly positive words, then the node word is imbued with a positive prosody. If both positive and negative collocates exist in the same context, the node word can be said to carry a neutral or mixed prosody.

### 2.3. Collocation and Lexical Repulsion

The concept of collocation is widely discussed in the field of CL. Sinclair (1991), who states the importance of analyzing words by how they are combined with one another in a span of words either on the left or on the right:

Collocation is the occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is a maximum of four words intervening. Collocations can be dramatic and interesting because of their unexpected nature, or they can be important in the lexical structure of the language because of their frequency. (...) Collocation, in its purest sense, as used in this book, recognizes only the lexical co-occurrence of words. (Sinclair, 1991, p. 170).

Biber et al. (1998) define a collocation as words that tend to co-occur more frequently with the lexical items in analysis. In their opinion, identifying the most frequently occurring words is an effective technique to begin the analysis of the construction of the meaning of a word. On the other hand, Renouf and Banerjee (2007, p. 415), present the idea of *repulsion*, which refers to pairs of certain words that do not occur together and are intuitively-observed in language use. *Repulsion* in these pairs of words not only occurs because they are semantically, grammatically, or morphologically incompatible, but also occurs where there seems to be no other plausible explanation other than a standard English convention (Renouf and Banerjee, 2007, p. 419).

## 3. Methodology

### 3.1. Near-Synonyms

First, to start our analysis, we used a tool within the *Corpus of Contemporary American English (COCA)* to check if the word pairs were considered synonymous to one another. This information was obtained through the regular expression [= \*] applied to each word: [=worry] and [=concern] and [=worried] and [=concerned]. COCA is a reference monitor corpus, or a compiled corpus whose contents are fixed for use as a reference, of English used in the United States, as well as the largest English corpus available. For this study, *worry/concern* were selected as verbs in the infinitive form, and *worried/concerned*

were selected as adjectives.

According to Biber et al. (1998), the research questions used in lexicographic studies to explore the meanings of words in corpus-based research can be utilized to show all the contexts in which lexical items take place. In order to investigate how and in what context the native speakers of English use certain collocates, an analysis was based on observation of the empirical data of the *Louvain Corpus of Native English Essays* (LOCNESS). LOCNESS (Granger, Sanders, and Connor, 2014) is a compilation of the writings of native English speakers and has 209,783 words. The participants were university undergraduates, and the participants' essays, which averaged the same length, were based on topics given by researchers.

To investigate how non-native speakers apply the pairs of near-synonyms, the 2009 version of the *Brazilian Subcorpus of the International Corpus of Learner English* (BRICLE), which contains 159,000 words, was used. These essays were compiled by different universities, such as the Pontifical University of São Paulo (PUC-SP), the State University of Rio de Janeiro (UERJ), and the Federal University of Minas Gerais (UFMG), among others. The essays were written by students in their own time (untimed) using language reference tools (dictionary, thesaurus, etc.), but consisted entirely of the students' own work. For example, students were not allowed to draw on other articles or books to write their essay, and they were not allowed to ask a native speaker of English for help. Additionally, they also had to fill out the information in the Learner Profile (LP; see APPENDIX A). Each essay had approximately 500 to 1,000 words and followed the compilation criteria of the *International Corpus of Learner English* (ICLE). ICLE is a corpus with 3.7 million words (Granger et al., 2009) and has essays produced by advanced learners of English from different countries like Finland, France, Germany, and Japan, among others. The essays follow 13 suggested titles (Granger et al., 2009):

1. Crime does not pay.
2. The prison system is outdated. No civilized society should punish its criminals; it should rehabilitate them.
3. Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value.
4. A man/woman's financial reward should be commensurate with their contribution to the society they live in.
5. The role of censorship in Western society.
6. Marx once said that religion was the opium of the masses. If he was alive at the end of the 20th century, he would replace religion with television.
7. All armies should consist entirely of professional soldiers: there is no value in a system of military service.
8. The Gulf War has shown us that it is still a great thing to fight for one's country.
9. Feminists have done more harm to the cause of women than good.
10. In his novel *Animal Farm*, George Orwell wrote, "All men are equal: but some are more equal than others." How true is this today?

11. In the words of the old song, "Money is the root of all evil."
12. In the 19th century, Victor Hugo said, "How sad it is to think that nature is calling out but humanity refuses to pay heed." Do you think it is still true nowadays?
13. Some people say that in our modern world, dominated by science, technology, and industrialization, there is no longer a place for dreaming and imagination. What is your opinion?

### 3.2. Semantic Prosody and Lexical Repulsion

Based on the concepts of semantic prosody described in the theoretical basis, a word can have, based on its surroundings, lexical items that express positive, neutral, or negative semantic prosody (Stubbs, 1996). Concordance lines of the near-synonyms were selected to provide a method of analyzing each specific word in LOCNESS and its environment (Anderson & Corbett, 2009). To set the semantic prosody of words in this study, we chose to select 10 words and context-related collocates to classify them into one of three options: positive connotation, neutral connotation, or negative connotation. The data analysis also included examining the lexical repulsion of the pairs (Renouf, 2007). These collocates were analyzed to determine if the semantic prosody could be related to any observed repulsion with each word pair.

To check the concordance lines in LOCNESS to define the semantic prosody and lexical repulsion, we used AntConc (Anthony, 2011), available free of charge online and necessary to read the data, because the essays were compiled in a text (.txt) format. Moreover, the writing sample in which the pairs appeared could be quickly identified and evaluated, if necessary, to obtain any other information in a general context.

## 4. Results and Analysis

### 4.1. Semantic Prosody in LOCNESS

#### Worry

Worry was found to have a negative semantic prosody in LOCNESS (Table 1). The majority of its collocates fell in the negative connotation category, with collocates such as *problem*, *risk*, *fired*, and *erase*.

Concordance lines for *worry* taken from LOCNESS:

1. "...today's storage techniques are so advanced, there is *nothing* to worry about."
2. "'Main *Problems*' are those which worry road and rail transport operators."
3. "[Lowering the *risk*] has, therefore, made the fight safer and reduced the need for worry."
4. "...if the producers just worry about their *money* and not build a relationship with their consumers, the consumers will find someone else to patronize."
5. "The police could worry more about *solving* more murder cases than a lot of these drug cases."

6. "You don't have to **worry** about being *fired*."
7. "Obviously, a united Europe is not going to *erase* centuries of culture to form one identifiable 'European culture'; but this might always be a **worry** for the British."
8. "When using abstinence, a person does not have to **worry** about *getting pregnant* because they will not be in the game."
9. "It seems to have two distinct *forms* of which one seems to **worry** the British public most; Political Union."

**Table 1.** Collocates used to determine the semantic prosody of *worry* in LOCNESS

Collocate	Positive Connotation	Neutral Connotation	Negative Connotation
Nothing		X	
Problem			X
Risk			X
Money			X
Solve	X		
Fired			X
Erase			X
Pregnancy			X
Form		X	

### Concern

The lexical items surrounding *concern* in LOCNESS (see table 2) indicate that a neutral semantic prosody is associated with the usage of the word, with collocates such as *effect*, *aspect*, *publicity*, and *water* used in assigning the neutral prosody.

Concordance lines for *concern* taken from LOCNESS:

1. "...leaders of the Methodist church have expressed great **concern** at the *effect* on low income families who spend more than they can afford on tickets."
2. "The *condition* of roads is often a local issue rather than of national **concern**."
3. "So, the harmful *aspects* of genetic manipulation are very few in number, and most people need not **concern** themselves with worrying about it."
4. "Beef has been given bad *publicity* by the press, and it is becoming a major **concern** with the public."
5. "The *gap* between primary and secondary education became a major **concern** for the later governments."
6. "Pattullo merely addresses the very real **concern** of sexual *feelings* that may arise if homosexuals and heterosexuals share the same barracks."
7. "After authorities investigated the scene of the crime, they drove to the home of O.J. Simpson, supposedly out of **concern** for his *safety*."

8. "So if the *cost* of punishment is at all a **concern** for you, then it is obvious that capital punishment is the wrong direction to go."
9. "Also, like the text says, companies need to have a greater **concern** for *water* and why it should not be polluted."

**Table 2.** Collocates used to determine the semantic prosody of *concern* in LOCNESS

Collocate	Positive Connotation	Neutral Connotation	Negative Connotation
Effect		X	
Condition		X	
Aspect		X	
Publicity		X	
Gap			X
Feelings	X		
Safety		X	
Cost		X	
Water		X	

### The adjective "Worried"

In contrast to its infinitive form, in LOCNESS, *worried* (Table 3) was found to have a neutral or negative semantic prosody; however, comparing it to *concerned* (Table 4), its semantic prosody is more negative, which is consistent with the infinitive form. Collocates include: *lack*, *distance*, *role*, and *lose*.

Concordance lines for *worried* taken from LOCNESS:

1. "Charities were **worried** that they would *lose* out."
2. "Clubs and charities who sold scratch cards as a source of income were **worried** because they thought *nobody* would buy their cards."
3. "Many people are **worried** about the *lack* of democratic control over the Community decision making process and the voting systems used to enact legislation."
4. "Instead of the law enforcement worrying about the dangerous criminals such as murderers and rapists they are **worried** about controlling the *use* of drugs."
5. "[A]nalysts were **worried** that more people would start to *gamble*."
6. "While people are **worried** about whether or not *testing* is ethical they should think about their morals."
7. "Dora also is **worried** by the *distance* separating the Organisation and the people it is fighting for."



8. "So, why are they more interested in themselves and not **worried** about the *fans*?"
9. "This example showed how a company was willing to give in at first and build a relationship before they **worried** about *profit*."
10. "Perhaps because they are **worried** and uncertain as to their *role* and Britain's in this new state, they prefer to believe that it is simply not possible or even probable."

**Table 3.** Collocates used to determine the semantic prosody of *worried* in LOCNESS

Collocate	Positive Connotation	Neutral Connotation	Negative Connotation
Lose			X
Nobody			X
Lack			X
Use		X	
Gamble			X
Testing		X	
Distance		X	
Fan	X		
Profit	X		
Role		X	

### **Concerned**

Correlating with the infinitive form, the semantic prosody of the adjective *concerned*, based on the lexical items *idea*, *proposal*, *future*, and *nature*, was neutral (Table 4).

Concordance lines for *concerned* taken from LOCNESS:

1. "He felt that man was too **concerned** with *ideas*, and not with the precious side of human life, and the need to preserve it."
2. "He is basically a capitalist of nature, selfish, individualistic and very **self-concerned**."
3. "Being **concerned** with the *exploits* of terrorists it fails to deal with social and political conditions in any detail."
4. "Although life is futile, people are still free to choose although the *masses*, as far as Camus is **concerned**, are unaware of this freedom."
5. "The main *proposal* for reform, put forward by Jean Zay & later the Languin-Wallon plan which formed the basis for the eventual reforms, **concerned** what was called 'orientation'."
6. "As far as the immediate *future* is **concerned** a certain loss of sovereignty is the

- price we have to pay for material progress.”
7. “The final argument against condom distribution in schools is **concerned** with the *dependability* of condoms.
  8. “Why should we be **concerned** with the *life* of a violent criminal?”
  9. “The average criminal can work one hour a day and make \$300.00 without having to be **concerned** with *deductions* for medical insurance and taxes.”
  10. “People have also become **concerned** with *nature*.”

**Table 4.** Collocates used to determine the semantic prosody of *concerned* in LOCNESS

Collocate	Positive Connotation	Neutral Connotation	Negative Connotation
Idea		X	
Self		X	
Exploit			X
Masses		X	
Proposal		X	
Future		X	
Dependability		X	
Life	X		
Deduction			X
Nature		X	

#### 4.2. Lexical Repulsion with Word Pairs

The data analysis also included examining the lexical repulsion, association, or neutrality of collocates as they occurred with the word pairs. The collocates described in the semantic prosody were analyzed to determine if the semantic prosody could be related to any observed repulsion with each word pair.

**Table 5.** Collocates attracted by *worry/worried* and repelled by *concern/concerned*

Collocate	Prosody
Risk	Negative
Problem	Negative
Money	Neutral to negative

Role	Neutral
Erase	Negative

**Table 6.** Collocates attracted by *concern/concerned* and repelled by *worry/worried*

Collocate	Prosody
Life	Positive
Effect	Neutral
Proposal	Neutral
Nature	Neutral
Cost	Neutral

The collocates attracted by *worry/worried* (Table 5) are all neutral to negative. The most commonly used were *risk*, *problem*, and *money*, which in this case the semantic prosody is neutral to negative, as it is a matter of personal finance. All of these collocates have in common a negative aspect about a potential situation in the future, which is not observed in *concern/concerned*. Therefore, *worry/worried* tend to attract negative collocates while *concern/concerned* repel them.

As observed in the concordance lines, the semantic prosody of the most common collocates (Table 6) such as *life*, *effect*, and *proposal*, observed with *concern/concerned* is predominantly neutral. This word pair tends to attract neutral collocates due to an overreaching concern of a particular issue, which is best used in a neutral connotation. This neutral method of conveying concern is not observed in *worry/worried*; therefore, *concern/concerned* attract neutral collocates, but on the other hand, *worry/worried* repel them.

#### 4.3. Data from Br-ICLE

To start the analysis of frequency of errors, we present the occurrences of the words investigated in Br-ICLE in infinitive nodes (Table 7) and in adjective nodes (Table 8).

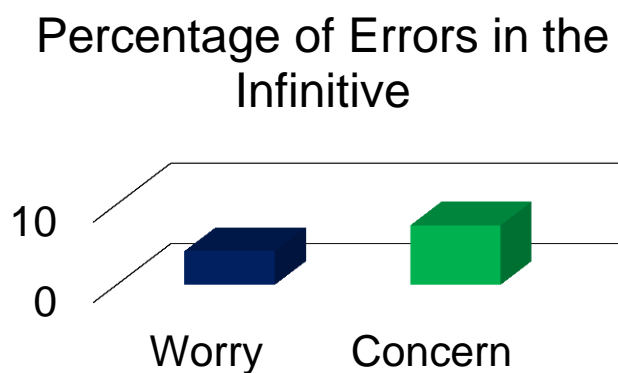
**Table 7.** Frequency of infinitive nodes found in Br-ICLE corpus

Node	Frequency
Worry	24
Concern	27

**Table 8.** Frequency of adjective nodes found in Br-ICLE corpus.

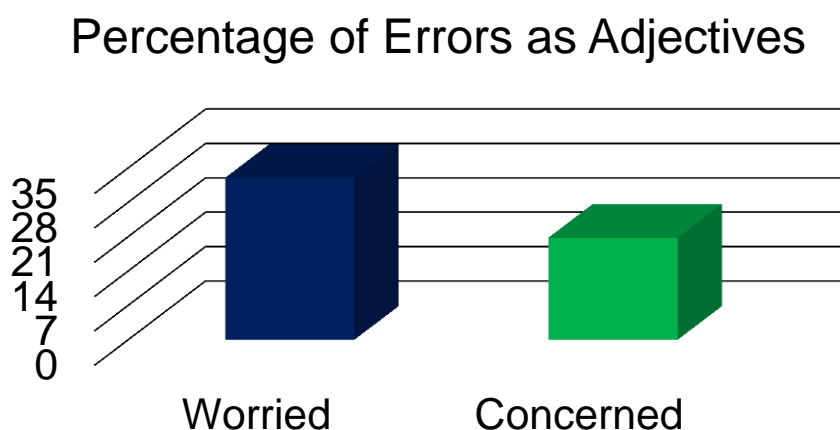
Node	Frequency
Worried	30
Concerned	48

Taking the semantic prosody of the pairs into consideration, we now analyze the concordance lines in Br-ICLE (See APPENDIX B) to see the use the near-synonyms by learners. Figure 2 shows *worry* was misused by the students in 1 occurrence (4,2%), while *concern* was used incorrectly in 2 occurrences (7,4%).



**Figure 2.** Percentage of errors in Br-ICLE corpus for each word in the infinitive form.

When analyzing the adjective synonym-pairs, we found 10 occurrences (33%) of *worried* used incorrectly, and 10 occurrences (21%) of *concerned* were not applied correctly (Figure 3).



**Figure 3.** Percentage of errors in Br-ICLE corpus for each word as adjectives.

#### 4.6. Discussion - Comparison of Use between LOCNESS and Br-ICLE

The Br-ICLE corpus was compared to the native English speaker corpus LOCNESS to determine semantic prosody patterns. Our research highlights the ability of native English speakers over Brazilian learners to distinguish positive, neutral, and negative semantic prosody as follows.

When *worry* was searched in the concordance lines, a negative semantic prosody was found; “problem”, “risk”, “erase”, and “fired” were associated with the negative prosody. The uses of *concern* showed a neutral semantic prosody, with “aspect”, “condition”, “effect”, and “water” used. By analyzing the concordance lines, we see underlying negative and neutral semantic prosodies for *worry* and *concern*, respectively. The Brazilian learners were more aware of the semantic prosody of these pairs; in fact, there were only three total misuses of the two combined.

Surprisingly, when these misuses were discovered, the sentence structure around these collocates was incorrect. We found that the implied neutrality or negativity of the following sentences was not evident to the learners, resulting in the misuse of the words.

1. [This] involves sensibility, technical knowledge, concern about human problems, expectations about life and construction of a better world.
2. Even though, some restaurant owners are not satisfied with this prohibition because they are concern with the probability of losing smokers customers.
3. [The government employees] do not worry about the system.

In the first sentence, *concern* is used to talk about human problems; this contrasts with what we observed in LOCNESS, as “problem” was a negative connotation with *worry*. Correcting this portion of the sentence would read: “...beginning to *worry* about human problems...” because from this sentence, it is implied a negative problem has not begun to be addressed yet. The second sentence describes a business and its relationship with losing customers who smoke. Losing business is a negative consequence of a new prohibition, so *concern* is not the appropriate choice in this case. The corrected portion of this sentence would read: “...because they *worry* about the likelihood...” Finally, in the last sentence, it is not stated if there is a problem with the system, so as written, it shows a neutral connotation; therefore, *worry* is inappropriate for this sentence. The sentence structure should be corrected to say: “[The government employees] do not show *concern* for the system.”

Br-ICLE indicated the learners did not understand when to use the words *worried* and *concerned* based on their prosody. *Worried* had a neutral to negative semantic prosody, with “lose”, “lack”, “use”, and “role” used in its vicinity. *Concerned* was predominantly neutral in its semantic prosody; “idea”, “self”, “future”, and “nature” were used. The Brazilian learners struggled with the use of these two words in Br-ICLE; there were ten mistakes we found per collocate in the corpus. The learners often chose *worried* where there was a neutral prosody, and they frequently chose *concerned* when there was a negative prosody; this is in contrast to their choices using the infinitive forms of the near-synonyms, as the learners seemed to be more aware of prosody in that case. In conclusion, the adjectives did not match the infinitive form in the Brazilian learner corpus; the learners did not understand the semantic prosody of the words when they used the in-

correct near-synonyms in the sentences.

The predominant mistakes could be seen when a neutral word was used and *concerned* was more appropriate, and when a negative word was used, *worried* was more appropriate. I have chosen two sentences from both *worried* and *concerned* to illustrate this point.

1. Marx was worried, at his time, about religion and experts are worried, nowadays, about television.
2. Nowadays we are much more worried about economy than ecology.
3. [F]amilies are falling apart because parents are more concerned about how to make more money than how to instruct their children; workaholic parents are careless about family.
4. The [routines] have made us usually concerned only about things we need to do.

In the first set of sentences, the underlined *worried* will be examined. If written correctly, the experts would be worried about the “influence” of television, not about television itself. “Influence” has a neutral connotation, so *concerned* would be the appropriate choice. The second sentence implies there is something wrong with the economy, so people are *worried* about it; however, there is no connotation in the sentence to suggest anything is negative or positive, only a neutral connotation. Therefore, *concerned* is a more appropriate choice to use.

The second set of sentences will focus on the underlined *concerned*. Sentence 3 is overwhelmingly negative in connotation as the focus of the family is on money rather than the health of the family; however, the parents’ *worry* is with their wealth rather than the implied “well-being” of their family, so *worried* should be used in the place of *concerned*. Sentence 4 appears correct as written, but in fact, the “routines” and “things” referred to in this sentence are actually issues that need to be resolved; this implies a negative connotation to the sentence that would not be obvious to a Brazilian learner of English. Therefore, the proper choice for this sentence should be *worried*.

## 5. Conclusion

Two sets of word pairs were researched in native and non-native English speaking corpora, and the questions initially posed were answered. The words in each pair were found to appear in specific semantic prosody environments. *Worry/worried* were found in predominantly negative connotations, while *concern/concerned* were found in predominantly neutral connotations. Each word pair repelled the other word pair’s most frequent collocates on the basis of semantic prosody. In the infinitive form, the semantic prosody of the surrounding environment was found to have a tremendous impact on the choice of word in the pair by both native and non-native English speakers, but as adjectives, the non-native English speakers struggled with using the correct collocate. Using these comparative studies, teachers can prepare activities for students in order to make them understand better the use of near-synonyms by introducing their students to semantic prosody.

## 6. References

- Aarts, Jan., & Granger, Sylviane. Tag sequences in learner corpora: a key to interlanguage grammar and discourse, in: Granger, S. (ed.) *Learner English on Computer*. London, New York: Addison Wesley Longman, 1998, p. 132-142, 2008.
- Anthony, Laurence. Software AntConc free of charge. The software is available to download at < [http://www.antlab.sci.waseda.ac.jp/antconc\\_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html)>.
- Anderson, Wendy & Corbett, John. *Exploring English with online corpora: an introduction*. London: Palgrave Macmillan, 2009.
- Biber, Douglas; Conrad, Susan & Reppen, Randi. *Corpus Linguistics: Investigating Language Structure and Use*. New York, NY: Cambridge University Press, 1998.
- COCA – Corpus of Contemporary American English. Retrieved March 28, 2015, from [corpus.byu.edu/coca](http://corpus.byu.edu/coca).
- Dutra, Deise & Silero, Rejane. *O uso de for: uma análise de itens linguísticos em corpus de aprendizes brasileiros*, in: Shepherd, T. M. G.; Sardinha, T. B. & Pinto, M. V., (ed.), *Caminhos da Linguística de Corpus*, (pp. 325-342). São Paulo: Mercado de Letras, 2009.
- Granger, Sylviane. A bird's-eye view of learner corpus research, in: Granger, S., Hung, J., Petch-Tyson, S. (ed.), *Computer learner corpora, second language acquisition and foreign language teaching*, (pp. 3-33). Amsterdam, The Netherlands: John Benjamins, 2002.
- Granger, Sylviane; Dagneaux, Estelle; Meunier, Fanny & Paquot, Magali (ed.). *International Corpus of Learner English* (Vol. 1) Louvain-la-Neuve: Presses Universitaires de Louvain, 2009.
- Hunston, Susan. Semantic Prosody Revisited. *International Journal of Corpus Linguistics* (vol. 12), 249-268, 2007.
- Hirst, Graeme & Inkpen, Diana. Building and using a lexical knowledge base of near synonym differences. *Computational Linguistics*, 32, 223-262, 2006.
- Louw, Bill. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In: Baker, M., Francis, G. & Tognini-Bonelli, E. (ed.), *Text and technology: in honour of John Sinclair* (pp. 157-176). Amsterdam: John Benjamins, 1993.
- Moon, Rosamund. What can a corpus tell us about lexis?, in McCarthy, M., & O'Keeffe, A (ed.), *The Routledge Handbook of Corpus Linguistics* (pp. 197-211). Londres: Routledge Taylor & Francis Group, 2010.
- Partington, Alan. *Patterns and meanings: using corpora for English language research and teaching*. Philadelphia, PA: John Benjamins, 1998.

Partington, Alan. Utterly content in each other's company: semantic prosody and semantic preference. *International Journal of Corpus Linguistics* (vol. 9). 131-156, 2004.

Renouf, Antoinette & Benerjee, Jayeeta. Lexical repulsion between sense-related pairs. *International Journal of Corpus Linguistics*, 12 (3), 415-444, 2007.

Selinker, Larry. Interlanguage. *International Review of Applied Linguistics*, 10, 209-231, 1972.

Silero, Rejane. *Os quantificadores a few e few: questões de interlíngua e prosódia semântica em corpus de aprendizes*. Dissertação (Mestrado). Universidade Federal de Minas Gerais. Programa de Pós-Graduação em Estudos Linguísticos, UFMG, Minas Gerais, 2014.

Sinclair, John. *Corpus, concordance, collocation*. Oxford: Oxford University Press, 1991.

\_\_\_\_\_. The Search for Units of Meaning. *TEXTUS*, v. 9, n. 1, 75-106, 1996.

Stubbs, Michael. *Text and corpus analysis: computer-assisted studies of language and culture*. Oxford: Blackwell, 1996.

Tenuta, Adriana; Oliveira, Ana & Orfanó, Bárbara. How Brazilian learners express modality in their writing: a corpus-based study on lexical bundles. *Revista Intercâmbio*, (pp. 1-15). São Paulo: LAEL/PUCSP, v. 25, 2012.

**Artigo recebido em 01/08/2016 ; aprovado para publicação em 25/10/2016**

**RESUMO:** A escolha da palavra errada pode transmitir conotações indesejadas, implicações e/ou posturas inadequadas, e a escolha entre quase-sinônimos, palavras que compartilham o mesmo significado central, porém que diferem em suas nuances, só pode ser uma alternativa se tivermos conhecimento sobre suas diferenças (Hirst and Inkpen, 2006). Partindo desse pressuposto, neste trabalho, os dados de dois corpora (um de falantes nativos de inglês, o *Louvain Corpus of Native English Essays (LOCNESS)* e um de aprendizes de inglês, o *Brazilian Subcorpus of the International Corpus of Learner English (Br-ICLE)*) foram examinados a fim de que fossem selecionados dois conjuntos de pares considerados quase-sinônimos: (a) *worry; concern* and (b) *worried; concerned*. A principal descoberta da pesquisa é de que uma conotação positiva, neutra ou negativa tem um impacto significativo sobre a forma como as palavras, em cada par, são usadas no corpus de nativos. No entanto, no corpus de aprendizes, a falta de compreensão da prosódia semântica dos pares de palavras investigados pode ter afetado a maneira como não nativos usam tais palavras.

**PALAVRAS-CHAVE:** quase-sinônimos, aprendizes, linguística de corpus.

**ABSTRACT:** Choosing the wrong word can convey unwanted connotations, implications, or attitudes; and the choice between near-synonyms, words that share the same core meaning but differ in their nuances, can be made only if the knowledge about their dif-



ferences is available (Hirst and Inkpen, 2006). Based on this assumption, the data of two corpora (one of native English speakers, the *Louvain Corpus of Native English Essays (LOCNESS)* and one from learners of English, the *Brazilian Subcorpus of the International Corpus of Learner English (Br-ICLE)*) were examined to select two sets of *near-synonym* pairs: (a) *worry; concern* and (b) *worried; concerned*. The main research finding is that a positive, neutral, or negative connotation has a significant impact on the way the words in each word pair are used in the native English corpus. However, in the non-native corpus, the lack of understanding of semantic prosody nuances for the investigated word pairs may affect how non-native speakers use such words.

**KEY-WORDS:** near-synonyms, learners, corpus linguistics.

APPENDIX A – Br-ICLE’s Learner Profile

LEARNER PROFILE

Text code : (do not fill in)

Essay :

Title :

Approximate length required :      -500 words      +500 words

Conditions :                              timed                              untimed

Examination :                              yes                              no

Reference tools :                              yes                              no

What reference tools ?

Bilingual dictionary :

English monolingual dictionary :

Grammar :

Other(s) :

Surname :

First names :

Age :                              Male

Female

Nationality :

Native language :

Father’s mother tongue :

Mother’s mother tongue :

Language(s) spoken at home : (if more than one, please give the average % use of each)

Education :

Primary school - medium of instruction :

Secondary school - medium of instruction :

Current studies :

Current year of study :

Institution :

Medium of instruction :

English only

Other language(s) (specify)

Both

Years of English at school :

Years of English at university :

Stay in an English-speaking country :

Where ?

When ?

How long ?

Other foreign languages in decreasing order of proficiency :

I hereby give permission for my essay to be used for research purposes.

Date : .....

Signature : ...

## APPENDIX B – Concordance lines from Br-ICLE

<b>Word</b>	<b>Original Text from Br-ICLE</b>	<b>Edited Text</b>
Worry	They do not worry about	They do not show concern with
Concern	because they are concern with	because they worry about
Concern	concern about human problems	beginning to worry about
Worried	experts are worried	experts are concerned
Worried	people are worried about money	people are concerned about money
Worried	people are worried about material	people are concerned with material
Worried	people are really worried	people are really concerned
Worried	We need more people worried	We need more people concerned
Worried	They are not worried with	They are not concerned about
Worried	he was worried about nature	he was concerned about nature
Worried	Mr. Bush is not worried	Mr. Bush is not concerned
Worried	more worried about another	more concerned about another
Worried	are even more worried	are even more concerned
Concerned	in which a man concerned	in which a man worried
Concerned	supposed to be concerned	supposed to be worried
Concerned	parents are more concerned	parents are more worried
Concerned	people might not be concerned	people might not be worried
Concerned	made us usually concerned only	made us usually worried only
Concerned	I am also concerned by	I am also worried by
Concerned	have to be concerned about safety	have to be worried about safety
Concerned	universities should be concerned	universities should be worried about
Concerned	more concerned about quality of life	more worried about quality of life
Concerned	concerned with their own salaries	worried about their own salaries