

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISNN 2237-759X

## 'DREAMS SEEM KIND OF UTOPIC': VAGUE CATEGORY MARKERS IN A LEARNER CORPUS

### 'DREAMS SEEM KIND OF UTOPIC': MARCADORES DE LINGUAGEM VAGA EM UM CORPUS DE APRENDIZ

Mateus Emerson de Souza MIRANDA  
(Mary Immaculate College, University of Limerick)  
mateus.desouza@mic.ul.ie

**ABSTRACT:** The aim of this paper is to analyze Vague Category Markers (VCMs) in the oral discourse of Brazilian learners of English, by combining corpus pragmatics (CP) and contrastive interlanguage analysis (CIA) methods. The corpus chosen for this research is a Brazilian oral interlanguage sub-corpus from the Louvain International Database of Spoken English Interlanguage (LINDSEI-BR) (GILQUIN ET AL, 2010; MELLO ET AL, 2013). For comparison to native English speakers, the Santa Barbara Corpus of Spoken American English (SBCSAE) was used. The findings from the analyses provide insights into the use of VCMs by Brazilian learners of English by showing that many of them are still not used by learners. In addition, learners prefer certain markers over the ones preferred by native speakers. This information may be valuable to highlight markers that perhaps should be addressed in the language classroom or, at least, to point out some important aspects that English teachers should take into consideration.

**Keywords:** vague category markers; corpus pragmatics; learners.

**RESUMO:** O objetivo deste trabalho é analisar os Marcadores de Linguagem Vaga (MLVs) no discurso oral de aprendizes brasileiros de inglês, combinando métodos da pragmática de corpus (PC) e de análise contrastiva da interlíngua (ACI). O corpus escolhido para esta pesquisa é o sub-corpus oral brasileiro do Louvain International Database of Spoken English Interlanguage (LINDSEI-BR) (GILQUIN ET AL, 2010; MELLO ET AL, 2013). Para comparação, o Santa Barbara Corpus of Spoken American English (SBCSAE) foi usado. Os resultados das análises fornecem insights sobre o uso dos MLVs por aprendizes brasileiros de inglês mostrando que muitos deles ainda não são usados. Além disso, os alunos preferem certos marcadores àqueles preferidos por falantes nativos. Esta informação pode ser valiosa para destacar os marcadores que talvez devam ser abordados na sala de aula ou, pelo menos, apontar

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISNN 2237-759X

alguns aspectos importantes que os professores de inglês devem levar em consideração.

**Palavras-chave:** marcadores de linguagem vaga; pragmática de corpus; aprendizes.

## **0. Introduction**

### **1. Integrating Corpus Pragmatics and Contrastive Interlanguage Analysis**

Corpus Linguistics (CL) is an area of research that has been growing consistently throughout the last few decades. It can be applied in vastly different areas of inquiry as a methodological tool. Linguists argue that it can provide a better means for quantitative and qualitative investigations of linguistic features on data taken from written and spoken natural language. According to Morton, Walsh and O`Keeffe (2011), CL focuses on large-scale analyses, and it does not account for context. Therefore, it is insufficient to deal with spoken interaction features at the levels of the utterance or turn, but it can be integrated with other approaches to provide more up-close, descriptive studies.

This work aims to analyze the pragmatic features in the oral discourse of Brazilian learners of English, using a combination of CL and pragmatics, called corpus pragmatics (CP) and contrastive interlanguage analysis (CIA). Focus will be placed on the use of Vague Category Markers (VCMs) and, more specifically, in the analysis of *kind of*, as a pragmatic marker by Brazilian learners of English. According to Rühlemann and Aijmer (2015), corpus linguistics and pragmatics were considered exclusive due to their differences in methodology, with 'pragmatics relying on close horizontal reading and qualitative interpretation and corpus linguistics typically scanning texts vertically and processing data quantitatively' (p. 01). In addition, this combination is recent, and it was possible because corpus linguistics and pragmatists realized they could benefit by having both fields merged. Therefore, by combining both fields, there is a highly iterative form of pragmatics, traditionally more qualitative or horizontal, and the vertical methodology of CL, being more quantitative (RUHLEMANN; CLANCY, 2018).

According to Callies (2015: 39), 'CIA as introduced by Granger (1996) is probably the most widely used methodological approach in Learner Corpora Research' (GILQUIN, PAQUOT, 2008; DÍEZ-BEDMAR, CASAS PEDROSA, 2011; MENDIKOETXEA, LOZANO, 2018; DÍEZ-BEDMAR, PAPP, 2008; ALTENBERG, GRANGER, 2001). As claimed by

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

Granger (2009: 18), CIA 'involves quantitative and qualitative comparisons between native language and learner language (L1 vs. L2) and between different varieties of interlanguage (L2 vs. L2).' She also states that native speakers (NS) and non-native speakers (NNS) comparisons can highlight a range of features of non-nativeness in learner writing and speech, i.e., not only errors, but also instances of under- and overrepresentation of words, phrases, and structures (GRANGER, 2002). For this reason, the integration of CP and CIA can provide both quantitative and qualitative insights for linguistics research.

## 2. Vague Language

Vagueness is an important characteristic in informal conversation, and utterances appear less direct, authoritative, or assertive, thus indicating the share of knowledge between speakers and marking in-group membership (CARTER; MCCARTJHY, 2006). In the literature, vagueness, depending on its manifestation, is also defined as 'tags' (WARD; BIRNER, 1992), 'terminal tags' (DINES, 1980), 'general extenders' (OVERSTREET; YULE, 1997a, 1997b), 'extension particles' (DUBOIS, 1993), 'generalized list completers' (JEFFERSON, 1990), 'vague category identifiers' (CHANNEL, 1994), and vague category markers (O'KEEFFE, 2003; EVISON; MCCARTHY; O'KEEFFE, 2007; VAUGHAN; MCCARTHY; CLANCY, 2017). For this paper, VCMs are adhered to exclusively.

According to Carter and McCarthy (2006: 202), vague language 'involves the use of words and phrases such as thing, stuff, or so, or something, or anything, and so on, or whatever, kind of, sort of'. Taking the lexical item **thing** as an example, Fox (1998) states that it means **object**, as in the example 'What's that **thing** over there?' However, it tends to be used in a less concrete context, as in, 'We get blamed for all kinds of **things**'. He states that **thing** can be used any time one does not want to specify more precisely what is said, but learners rarely use it as frequently as native speakers do. Crystal and Davy (1975: 11) put forward more reasons for the use of vague language:

- a. Memory loss – the speaker forgets the correct word
- b. The language has no suitable exact word or the speaker does not know it
- c. The subject of the conversation is not such that it requires precision, and an approximation or characterization will do
- d. The choice of vague item is deliberate to maintain the atmosphere

According to O'Keeffe (2003: 06), VCMs are:

Recognizable chunks of language that function in an expedient way as linguistic triggers employed by speakers and decoded by participants who draw on their store of shared knowledge. It is argued that the meaning of vagueness categories is socially grounded and are co-constructed within a social group that has a shared socio-historic reality.

As Fernandez and Yuldashev (2011) note, most corpora studies on vague language are focused on English L1 contexts (O'KEEFFE, 2003; AIJMER, 2013; CHENG; O'KEEFFE, 2015; CLANCY, 2016; O'KEEFFE, CLANCY, ADOLPHS, 2011; VAUGHAN; CLANCY; MCCARTHY, 2017) and not many studies have focused on its use by English learners. For example, Cheng & Warren (2001) have shown, through a study of the use of vague language between native and non-native speakers of English in Hong Kong, that such conversational features impact successful intercultural communication. In the Shirato and Stapleton (2007) contrastive interlanguage study, they found that Japanese learners of English underuse vague language items and believe that the results brought attention to important pedagogical applications for learners to have effective interactions in English. Orfanó (2013) found, in a corpus collected in a classroom of Brazilian graduate learners of English as a second language, that vague markers are under-represented and varied in the number of occurrences and form when compared to native data. Aijmer (2004) has shown through a corpus of interviews that Swedish learners, despite having great command of English, use vague language to express uncertainty or hesitation, while NS tend to use them for face-saving or to express politeness. Also, they frequently use them as a strategy when having communication problems. Lin (2013) has analyzed conversation between a group of British and Taiwanese adolescents and found out that there are differences in the amount of their use within the Taiwanese group. For instance, the top vague expression in the British corpus has no occurrences among Taiwanese learners, calling attention to the importance of the inclusion of spoken features in teaching materials for the English as a Foreign Language (EFL) teacher.

This paper presents an integration of CP and CIA to identify and quantify the use of VCMs in the LINDSEI-BR. In order to compare the use by learners and native speakers of English, an investigation of how they are distributed quantitatively in LINDSEI-BR and in SBCSAE was conducted. Additional topics include if there is overuse or underuse of any of the forms investigated in both corpora and what characteristics of learners can be revealed from their usage of such items. More specifically, for this paper, closer attention is paid to the analysis of the most frequent

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISNN 2237-759X

item, **kind of**. The results of this research may contribute to the understanding of the Brazilian learners' usage of vague items. This information may be valuable to highlight what should be addressed in the language classroom or, at least, to point out some important aspects that English teachers should take into consideration. For example, Fox (1998: 29) states that 'most classroom language is planned (or, at least, semi-planned) and therefore is lacking in some of the common features of unplanned discourse, which include the vague language items such as 'something like that', 'something', 'things like that' and, of course, 'thing' itself'.

### 3. Methodology

#### 3.1 The corpora

The LINDSEI project (GILQUIN; DE COCK; GRANGER, 2010) is coordinated at the Centre for English Corpus Linguistics, at the Université Catholique de Louvain, and was initiated by Sylviane Granger. In Brazil, the compilation of LINDSEI-BR was coordinated by Professor Heliana Mello (MELLO ET AL, 2013) at the Federal University of Minas Gerais (UFMG), and was set at university level, from high intermediate to advanced proficiency students. The speech of each learner in LINDSEI-BR was recorded for about 20 minutes, allowing the participants to complete tasks and present a position on a subject. To ensure a broad range of discussion topics were available for transcription in the project, three tasks were assigned following the international project guidelines (GILQUIN; DE COCK; GRANGER, 2010: 8):

- 1) a topic chosen by the learner in narrative form:

**Topic 1:** An experience you've had which has taught you an important lesson. You should describe the experience and say what you have learnt from it.

**Topic 2:** A country you have visited which has impressed you. Describe your visit and say why you found the country particularly impressive.

**Topic 3:** A film/play you've seen which you thought was particularly good/bad. Describe the film/play and say why you thought it was good/bad.

- 2) discussion about various topics with the task interviewer:

Free discussion: Informal conversation about learners' lives such as likes and dislikes university life, their interests and hobbies etc.

- 3) a description of a scene in a picture.

Learner speech features that were noted during the orthographic transcription include: overlapping, backchanneling, contractions, and truncation, along with other examples.

The native corpus used for comparison in this study was the SBCSAE, which is available online for download (DU BOIS ET AL 2000–2005) and is part of the International Corpus of English (ICE). It has recordings of spoken American English spontaneous interaction across the United States. Although LINDSEI-BR has a controlled section, it also comprises a spontaneous conversation at the beginning between the interview and interviewee.

### 3.2 Procedure

For this paper, CL is considered as a methodological tool to investigate LINDSEI-BR. Thus, the goal of using CL is to automatically search linguistics features in the data. After choosing the corpora, a first quantitative analysis on the usage of VCMs was made in both corpora. Vaughan, McCarthy and Clancy (2017) drew a list of the common VCMs referred to in previous literature. The list is considered in this study as a starting point for the analysis. The VCMs identified by them are as follows:

1. *(and) things (like that)*
2. *sort of/kind of*
3. *(and) all (of) that*
4. *(and) (all) that/this sort/kind/type of thing*
5. *these/those/all sorts/kinds/types of thing(s)*
6. *(or) something (like that)*
7. *(or) anything (like that)*
8. *and stuff (like that)*
9. *and/or whatever*
10. *and that*
11. *and so on*
12. *and so forth*
13. *etc./etcetera*
14. *and whatnot*
15. *(and) this, that and the other*

The program AntConc (Anthony, 2016) was used to identify VCMs in the list above and take the occurrence of each in both corpora. Then, a word cluster list was generated in both corpora to find any other VCMs that were not on the list proposed by Vaughan, McCarthy and Clancy (2017) (e.g. two-word clusters, three-word clusters, four-word clusters). Cluster, n-grams, lexical bundles or chunks are frequently occurring words that constitute a pattern of use (GREAVES; WARREN, 2010). The cluster list choice over frequency lists is because frequency lists are simply

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISNN 2237-759X

the record of a text's lexical repetitions and do not reveal much (THORNBURY, 2010).

After collecting the quantitative data, AntConc was used again to analyze the concordance lines. These are 'a collection of the occurrences of a word-form, each in its own textual environment' (SINCLAIR, 1991: 32), since quantitative data alone cannot answer any questions that researchers bring to the study of discourse (THORNBURY, 2010).

#### **4. Analysis**

In this section, the results and analysis of the data from the qualitative analysis of the corpora data are presented. For this study, a subcorpus of LINDSEI-BR, with 65,323 words, and a subcorpus of the SBCSAE with 312,659 were used. With regard to the number of tokens, the corpus of NS is larger than NNS. Evison (2010) states that to compare the frequencies of occurrences in corpora of different sizes, the frequencies must be normalized to establish the right proportion between them. According to Biber, Conrad and Reppen (1998: 263) 'the raw frequency count should be divided by the number of words in the text, and then multiplied by whatever basis is chosen for norming'.

Since English learners are the focus of this research, data taken from LINDSEI- BR will be presented first and later compared to the native English corpus.

##### **4.1 Frequency of VCMs**

After retrieving the VCMs in both corpora, the results were quantified in Table 2. It is important to note that the VCM *and whatnot* was excluded from the list, since there were no occurrences of it in either corpus. Searching for possible other VCMs in the cluster lists, the occurrence of the following three other VCMs was found in both corpora and were added to the list: *and everything*; *(and) blah blah blah* and *or what* in the SBCSAE. Table 2 shows the VCMs that were investigated in LINDSEI-BR and SBCSAE. The chart displays their normalized frequency.

TABLE 2.

|    | <b>VCMs</b>  | <b>LINDSEI-BR</b> |    | <b>VCMs</b>  | <b>SBCSAE</b> |
|----|--|-------------------|----|--|---------------|
| 1  | <i>kind of</i>                                       | 211               | 1  | <i>kind of</i>                                       | 127           |
| 2  | <i>(and) (all) that/this sort/kind/type of thing</i> | 68                | 2  | <i>(or) something (like that)</i>                    | 117           |
| 3  | <i>(or) something (like that)</i>                    | 53                | 3  | <i>and stuff (like that)</i>                         | 70            |
| 4  | <i>and everything</i>                                | 39                | 4  | <i>sort of</i>                                       | 47            |
| 5  | <i>these/those/all sorts/kinds/types of X</i>        | 25                | 5  | <i>(and) all (of) that</i>                           | 39            |
| 6  | <i>(and) all (of) that</i>                           | 25                | 6  | <i>and/or whatever</i>                               | 30            |
| 7  | <i>and stuff (like that)</i>                         | 25                | 7  | <i>and everything</i>                                | 29            |
| 8  | <i>(and) things (like that)</i>                      | 18                | 8  | <i>(or) anything (like that)</i>                     | 23            |
| 9  | <i>and so on</i>                                     | 7                 | 9  | <i>these/those/all sorts/kinds/types of X</i>        | 20            |
| 10 | <i>and/or whatever</i>                               | 7                 | 10 | <i>(and) (all) that/this sort/kind/type of thing</i> | 17            |
| 11 | <i>etc./etcetera</i>                                 | 3                 | 11 | <i>and blah blah blah</i>                            | 16            |
| 12 | <i>and blah blah blah</i>                            | 3                 | 12 | <i>or what</i>                                       | 12            |
| 13 | <i>and so forth</i>                                  | 0                 | 13 | <i>etc./etcetera</i>                                 | 8             |
| 14 | <i>or what</i>                                       | 0                 | 14 | <i>(and) things (like that)</i>                      | 8             |
| 15 | <i>(or) anything (like that)</i>                     | 0                 | 15 | <i>and so on</i>                                     | 4             |
| 16 | <i>sort of</i>                                       | 0                 | 16 | <i>and so forth</i>                                  | 2             |
| 17 | <i>(and) this, that and the other</i>                | 0                 | 17 | <i>(and) this, that and the other</i>                | 1             |

Comparison of VCMs in LINDSEI-BR and the SBCSAE normalized per 250,000 words.

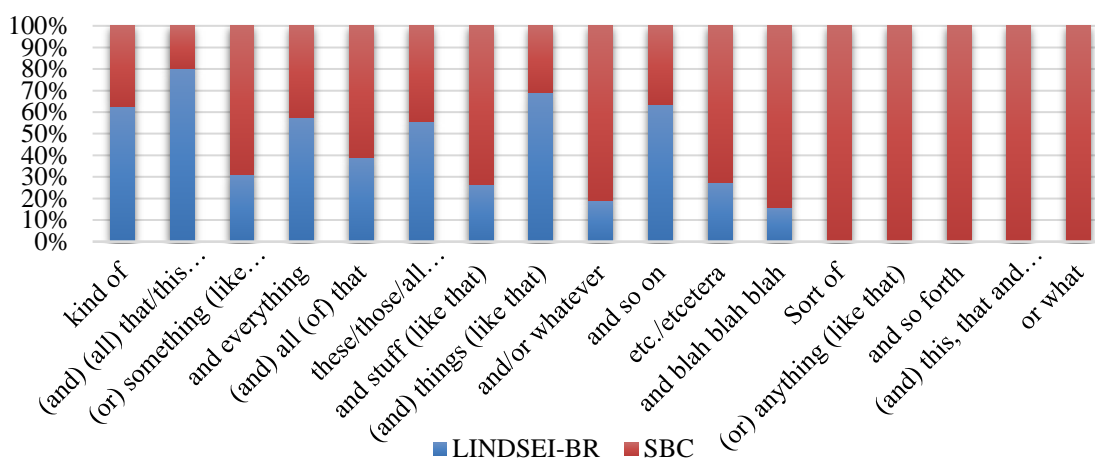
One major difference between LINDSEI-BR and SBCSAE is the frequency with which they are used. After totaling the number of occurrences in both corpora, there are 484 occurrences in the former and 570 occurrences in the latter. The SBCSAE has nearly 100 more uses than does LINDSEI-BR. Upon closer examination, multiple VCMs within the LINDSEI-BR corpus are not used at all, whereas other VCMs are only used sparingly. Moreover, it was also found that not all of them are used to mark groups or categories or function as triggers. The lack of usage in some of the VCMs in LINDSEI-BR, contrasted with higher frequencies in the SBCSAE, suggests that learners do not have the ability to use as wide



of a range of VCMs compared to the native speakers. This is illustrated more clearly in figure 1.

When figure 1 was further analyzed, it was noted that the use of VCMs by non-native speakers of English was low overall. In fact, some VCMs, such as **sort of** and **and so forth**, among others, were not used at all in LINDSEI-BR. Therefore, the hypothesis is that the non-native speakers may not have been exposed to many VCMs when they learned English, and their vocabulary in this regard is subsequently lacking.

GRAPH 1.



Frequency of occurrence of VCMs in LINDSEI-BR and the SBCSAE.

In contrast to the NNS, the data from the SBCSAE show that VCMs are much more widely used. With the exception of **and whatnot**, all other VCMs identified by Vaughan, McCarthy and Clancy (2017) show up at least once. Additionally, with a few exceptions, NS demonstrate more total uses of VCMs for each type than do NNS.

Another major difference within the corpora can be seen between **kind of** and **sort of**. Both corpora demonstrate that **kind of** is used, though it is more frequent in LINDSEI-BR than it is in SBCSAE, with roughly 85 more occurrences. This is surprising because even by adding up the occurrences of **kind of** and **sort of** in the SBCSAE due to their synonymy (KIRK, 2015), NNS use it more. In contrast, **sort of** is not used at all in the subcorpus, whereas it is used 47 times in the SBCSAE. The difference between their use is significant because of how similar they are. The results could also be related to the proficiency level of LINDSEI-BR participants. De Cock (2000) research on the French component of LINDSEI states that there is an underuse of **sort of** in the corpus. De Cock (2000) defines the participants of LINDSEI-FR as advanced and finds

31 occurrences in the corpus of 90,000 words. Although LINDSEI-BR participants are from intermediate to advanced level of English, this information from the metadata is not precise because the level of proficiency was self-assessed by the students. Prodromou (2005) suggests that the more advanced the learner becomes, the better use of pragmatic markers such as **sort of** they will make.

To gain a reference point for the variance in use, **kind of** and **sort of** can be compared between two native speaker corpora to provide context into their use and the reason for no occurrence in LINDSEI-BR. The Corpus of Contemporary American English (COCA) and the British National Corpus (BNC), were analyzed for occurrences as seen in Table 3.

TABLE 3.

| Expressions | COCA    | BNC    |
|-------------|---------|--------|
| Kind of     | 208,078 | 15,814 |
| Sort of     | 91,333  | 22,559 |

**Kind of** and **sort of** in COCA and BNC.

Both markers were found in COCA and in the BNC, but the occurrences varied greatly. In COCA, **kind of** occurred significantly more often than in the BNC. In contrast, **sort of** occurred more frequently than **kind of** in the BNC. From this analysis, it was determined that American speakers of English prefers **kind of**, whereas British speakers of English prefers **sort of**.

According to Gilquin and Granger (2015), learners may have a preference to use certain markers over others. From the results in the chart, the Brazilian learners' preference is exclusively **kind of** which is used for different purposes in the subcorpus. In Brazil, English learners are greatly influenced by American culture. Many Brazilians listen to American music, read American novels, and watch American movies. English as a second language schools in Brazil generally prefer American English for instruction. On the other hand, British English is significantly less common; Brazilians tend not consume as much British music, literature, and film as they do the American varieties, and schools that teach British English also are not common in Brazil (FRIEDRICH, 2000). Therefore, learners are more strongly exposed to one type of English over the other and the discrepancy in their frequency of occurrence in the LINDSEI-BR corpus can be put into context. It appears that learners having been influenced by American English therefore have been exposed to more examples of **kind of** in literature, film, and music, therefore making them familiar with this marker and using it frequently. Their lack

of exposure to British English diminishes the chance they know the equivalent *sort of*; indeed, a lack of occurrences of this marker supports this hypothesis. Thus, the occurrence of **kind of** in the learner corpus is related to and influenced by the learners' exposure to American English.

For this paper, a focus on the analysis of **kind of** was given.

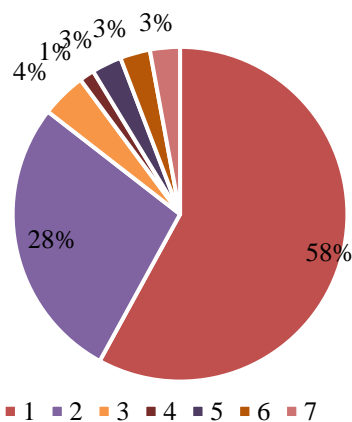
#### 4.2 The pragmatics of **kind of/sort of**

To support the quantitative analysis and refine the occurrences of **kind of** in LINDSEI-BR, seven functions of the pragmatic markers **kind of/sort of** described by Poos and Simpson (2002) were considered, as follows:

1. To express inexactitude
2. To soften the force of a stance and opinion
3. To mitigate criticism or request
4. To precede the use of sophisticated words
5. To establish and maintain rapport between interlocutors
6. To precede metaphors
7. To fill a pause and help the speaker keep the floor

When considering the functions described by Poos and Simpson (2002) in LINDSEI-BR, each concordance line was examined manually to classify the occurrences and it was found that learners tend to use of **kind of** to express inexactitude followed by softening the force of a stance and opinion as in figure 2:

FIGURE 2



Functions in LINDSEI-BR corpus.

In the examples analyzed from LINDSEI-BR below, overlapped speech is indicated with <overlap />.

### *Function 1: Expressing inexactitude*

Most of the uses of **kind of** in LINDSEI-BR are to be imprecise about something, or it is used when learners do not know how to explain something or they do not remember the exact word at the moment of speaking, as shown in extracts (1) and (2).

(1) (LINDSEI-BR)

[talking about a book]

<B> I like most the way<?> that she describe things and other .characters . she: is a young .girl she's about twenty-five or something like that . she lives . I don't know if is in= in Paris but I know it's in France .<overlap /> and: it's in Paris <laughs> good to know .and: (eh) .. she's really= really . (eh) (eh) . likes to describe things to describe . her father's . (er) .. his= job and the: the way he do things . and her mother . (eh) way of always getting nervous and things like that . it's . (eh) **kind of** a comedy a= a romantic comedy but . the the comedy part is not . as .we Brazilians used<?> to see . so that was . I found it more . interesting in the movie <laughs></B>

<A><overlap /> it's in Paris . Montmartre ... [?pause?\_01:23-02:01] what about her love </A>

(2) (LINDSEI-BR)

[talking about plans for the future]

<B> training yes probably next year I'll start teaching English and . I'm not sure if I really want to do this . but my experience <XX> training is okay . I'm changing my mind not change my mind but change my feeling about what is to work in a public school . in fact I like I'll be here for the next three or four years . I will try to do a **kind of** <foreign> concurso </foreign> you know and if I . get a good position or something stable I think I'll stay here .if not I will move from Brazil for sure </B>

<A> (mhm) </A>

According to Lakoff (1973), hedges are words that fuzziness are implicitly involved. In (1) the learner is imprecise when describing the book, and in (2) **kind of** precedes the word **concurso** in Portuguese which means that the participant intend to take a civil service examination, but forgets how to say the word in English and instead uses Portuguese. In addition, in (2) it is vague because the learner may want to refer to having similar jobs to a civil servant. In the examples above, they can also signalize a difficulty of the learners to articulate the intended meaning, but at the same time, their use can facilitate the flow of the learners' discourse (PRODRUMOU, 2005).

*Function 2: Softening the force of a stance or opinion*

According to Poos and Simpson (2002), softening an opinion is closely related to the function mentioned above. In example (3), the learner is describing a moment of being involved with drugs and when describing the parents' reaction to it **kind of** is used to soften the father's reaction who dealt better with the situation compared to the mother, but who still suffered.

(3) (LINDSEI-BR)

[talking about an experience in life]

<A> and you see them now </A>

<B> (eh) I see some . nowadays this is really different .because we don't have any= anything to talk about .. and then I realized that I was really blind about everything . I was really .. <XX> but nowadays I prefer to stay . (eh) in the boat I am <laughs><X><overlap /> (mm) I Icrea= (mm) with= I've created a **kind** . nowadays I've solved it .. but I think there still a gap between my . me and my mother especially . cos we suffered a lot . (eh) my daddy **kind of** . (erm) . dealt better with it . started being my friend instead of being my daddy and h= he helped me a lot . with this behavior . but my mom . (mm) . yeah I created a big ha= a big gap . between us . but we . nowadays it's better sometimes . he's afraid . of me getting . in this stuff again .. (er) .. </B>

<A><overlap /> what about your parents position now .. do you have a good relationship . with them ... sure .. relax<XX></A>

Also used as a hedge to mitigate, the learner in (4) uses **kind of** to soften his opinion that he is advanced in the course because he did not exactly follow the course curriculum and took classes from more advanced terms. To downtone the force of his opinion once the interviewer can disagree with him, he prefers to mitigate his assumption.

(4) (LINDSEI-BR)

[talking about the course at the university]

<B> yeah that might be possible (eh) this is my third term here so <laughs> actually the third <overlap /> this is the third only the third it's like (erm) <overlap /> I did I did some tests then I I managed to get some some subjects from seventh term then then sixth term fifth term so that's why I'm **kind of** advanced in it in the course but . (eh) actually if I I had to say (eh) which term am I in <overlap /> so I would say oh the third <laughs> yeah yeah in theory the third but (erm) I studied some things in advance so that's why I'm <laughs> doing phonology right now <overlap /> yeah so that's why . . so yeah maybe yeah it can be linguistics I'm I'm interested in that yeah . good</B>

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

<A><overlap /> yes it's your third </A>

### *Function 3: Mitigating criticism or requests*

In example (5) when talking about the future, the learner states he is not willing to teach after finishing the course at the university, especially teaching teenagers, and has an opinion that since he works with teenagers in a rich neighborhood, he wants to convey his idea by criticizing they are spoiled because they are from the rich neighborhood. According to Poos and Simpson (2002), the use of **kind of** here would be to soften 'a bald assertion and mitigating a remark that could potentially threaten the interlocutor's positive or negative face' (p. 16).

(5) (LINDSEI-BR)

[talking about future plans]

<A>and . which are your .. plans after graduation </A>

<B><overlap /> (mm) </B>

<A><overlap /><XX> teaching </A>

<B><overlap /> actually I intend no </B>

<A> no </A>

<B> the problem is that I teach to teenagers and teenagers suck they don't wanna be there especially because it's **kind of** a rich neighborhood, so they're spoiled they're hard to deal with so my my plan my </B>

### *Functions 4 and 5: Preceding the use of sophisticated vocabulary or jargon words/ to establish and maintain rapport between interlocutors*

Some occurrences of **kind of** were found in LINDSEI-BR that preceded the use of sophisticated words, like in (6) **kind of utopic** to position himself in relation to the interlocutor (POOS; SIMPSON, 2002). Additionally, to maintain rapport with the interviewer, **kind of** is also used in the same extract to keep the conversation friendly.

(6) (LINDSEI-BR)

[talking about future plans]

<B> [...] yeah (eh) . my future is like I created my future in the present I'm not that **kind of dreamful** . I was dreaming I don't like to <overlap /> yeah but I had ... has plans yeah I said dream yeah .. I said dreams yeah I . yeah<laughs> it's true . it's true ... but I I I yeah . I said the wrong word . plans are better yeah plans are better . dreams . dreams seem **kind of utopic** like a <laughs> like romantic stuff you know ..

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

we<overlap /> yeah yeah . but . I don't wanna get married now <laughs> it's just a plan . I wanna solve my career first [...] **</B>**

### *Function 6: To precede metaphors*

Similar to how **kind of** preceded complex vocabulary in the example above, it is also used to precede metaphors as in (7). According to Poos and Simpson (2002), 'these hedges also seem to function as metapragmatic markers — drawing the listeners' attention to the non-literal terminology and attesting to the speakers' self-consciousness about using an overt metaphor' (p. 17).

(7) LINDSEI-BR

[talking about a movie]

**<A>**how is how is the end **</A>**

**<B>**when . in the end . they stay together <laughs> but . it's not that obvious because= she: . she **kind of** . afraid of love . and .. and a certain<?> way . and there is a moment that they: they make a parallel between them . and .the fact that she she grew up . taking home (eh) classes because her mother think that she has a heart problem [...] **</B>**

### *Function 7: Filling a pause and helping the speaker to keep the floor*

In this function, the speaker is unsure of how to connect one part of the conversation to the next part. Therefore, the parts before and after **kind of** may not naturally connect, meaning it is used in this case to hedge. It can also be used to pause while the speaker comes up with the correct word or phrase to continue the conversation. Example (9) shows that the speaker uses **kind of** to pause before coming up with "suitable environment" to continue the conversation. The other speaker does not interject during this pause.

(8) LINDSEI-BR

[talking about future plans]

**<A>**and do you think that now: you can teach . adults ... <overlap /> (mhm) **</A>**

**<B>**yes I hope so . I started with kids because I thought I would feel comfortable <overlap /> and learn how to be a teacher you know because it would be a more: **kind of** . suitable environment for me ..so: I= I di= I wanted . when I come back f= if I pass please <laughs> (eh) if I come back from the exchange program I I think I'm gonna try <name of the first school> again **</B>**

As a comparison, the SBCSAE was analyzed to see if uses of **kind of** mirrored or differed from the uses in LINDSEI-BR. Some of the uses were identical between the two corpora. First, one of the same uses of **kind of** was in front of adjectives to soften opinions, though this happened more frequently than in LINDSEI-BR. It was also noted that complex vocabulary (4) was introduced more frequently than in the learner corpus (**nebulous, anthropomorphic, bogus, crusade** as examples).

Although there were these similarities, the SBCSAE showed a few variations in use that did not appear in LINDSEI-BR. The SBCSAE featured uses of **kind of** to precede expletives (**fucking, pissed**) to mitigate them (3). No expletives were found in the learner corpus. According to Thomas (1983), taboo language is extremely informal and thus difficult for learners to manage in their speech without risking failing socially and pragmatically. **Kind of** occurred more in the LINDSEI-BR and **kind of like** occurred only in the SBCSAE but not in the learner corpus. **Kind of a** is used in the SBCSAE to exemplify something, whereas **kind of like** is used in the native corpus as a hedge for (2). Finally, **this/that kind of thing** was observed. This VCM is used to share knowledge, specifically when one person thinks the other person will know or understand what he or she is talking about.

According to Poos and Simpson (2002), each of these categories of meaning is much fuzzier than has been so far described. Most instances of **kind of** and **sort of** in any given text are difficult to place in any one of these categories. Many are entwined with connotations of politeness, accommodation, vagueness, and understatement, among other possibilities of interpretation (POOS; SIMPSON: 18).

Through learner corpora analyses, studies have been done on real language in use that contribute to the understanding of learners' competence. When considering the use of **kind of** in LINDSEI-BR and its multifunctionality listed by Poos and Simpson (2002) in Section 3.2, the analysis has shown that learners may use the marker to convey a range of functions, but it is mainly used to express inexactitude followed by softening the force of a stance of opinion which represents an overuse of these functions. On the other hand, some functions appear exclusively in the native corpus. This might be because students were not introduced to the other functions since pragmatic markers are considered an under taught aspect in L2 classrooms (GRANGER; TYSON, 1996) and therefore Brazilian learners are not aware of their full functions, since they prefer to use the simpler ones.



## 5. Final remarks

Through CP, the usage of vague markers on the continuum learners – native experts was conducted and explored both quantitatively and qualitatively. As with many other learner corpora studies, the study also relies on the contrastive interlanguage analysis method, through which it was possible to highlight features that are distinctive for the interlanguage of learners (GAETANELLE; GRANGER, 2015). The combination of corpus pragmatics and contrastive interlanguage analysis provides a better means to quantitative and qualitative investigations on data taken from the spoken corpora.

The findings from this analysis provide insights into the use of vague language by Brazilian learners of English by showing that markers such as **sort of** and **and so forth** are still not used by learners, evidenced by the hypothesis that they may not have been exposed to them. In addition, learners prefer certain markers over the ones preferred by native speakers, which is the case of **kind of, and everything, and things like that**, among others. Within a range of functions of the pragmatic markers such as **kind of**, learners tend to use them to express inexactitude followed by their use to soften the force of a stance and opinion. The lack of use regarding other pragmatic functions may be that its use may increase according to the learner's proficiency level (HELLERMAN; VERGUN, 2007). It is also important to highlight that different markers and functions may be found in corpora with different designs such as those in Orfanó's (2013).

In this study, the main goal of using CP was to establish distinctive characteristics in terms of vague language and having information from the native speaker corpus as a reference of comparison for the analysis. Results from this research may inform EFL materials and raise learners' pragmatics awareness in the L2 classroom. According to Orfanó (2013), 'it would be useful to include more authentic spoken language data in the classroom, with activities based on corpora in which vague items could be explored' (p. 376). Despite the many benefits to studying learners' interlanguage, many available corpora are not calibrated to the Common European Framework (CEFR; Council of Europe, 2001; Companion Volume, 2018) which is the case of LINDSEI-BR. This makes it difficult to establish learners' competence and identify functions at each CEFR level. An important limitation of this study is that LINDSEI-BR had to be analyzed as a whole, which shows the importance of compiling calibrated corpora for studies on learner language. In this way, the use of language features across the levels can be tracked.

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

## References

AIJMER, K. The meaning and functions of the Swedish discourse marker *alltsa*-Evidence from translation corpora. *Catalan journal of linguistics*, Vol. 6, p. 31-59, 2007.

AIJMER, K.; RUHLEMANN, C. (Eds). *Corpus pragmatics: A handbook*. Cambridge: Cambridge University Press, 2015.

ALTENBERG, B.; & GRANGER, S. The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2). <http://dx.doi.org/10.1093/applin/22.2.173>, 2001, p. 173-195.

ANTHONY, L. AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University, 2016.

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus linguistics*. Cambridge: Cambridge University Press, 1998.

CALLIES, M. Learner corpus methodology. In: S. Granger, G. Gilquin.; F. Meunier. (Eds). *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139649414.003, 2015, p. 35-56.

CARTER, R.; MCCARTHY, M. *Cambridge grammar of English*. Cambridge: Cambridge University Press, 2006.

CHANNELL, J. *Vague language*. Oxford: Oxford University Press, 1994.

CHENG, W.; WARREN, M. The use of vague language in intercultural conversations in Hong Kong. *English World-Wide*, 22, 2001, p. 81-104.

CLANCY, B.; O'KEEFFE, A. Pragmatics. In: D. Biber and R. Reppen. (Eds). *The Cambridge Handbook on Corpus Linguistics*. Cambridge: Cambridge University Press, 2015, p. 235-251.

CLANCY, B. *Investigating Intimate Discourse: Exploring the Spoken interaction of Families, Couples and Friends*. London: Routledge, 2016.

COUNCIL OF EUROPE. *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: CUP, 2001.

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

COUNCIL OF EUROPE. The European Framework of Language Learning, Teaching and Assessment Companion Volume with New Descriptors, 2018.

CRYSTAL, D.; DAVY, D. *Advanced Conversational English*. London: Longman, 1975.

DE COCK, S. Repetitive phrasal chunkiness and advanced EFL speech and writing. In: C. Mair and M. Hundt. (Eds). *Corpus linguistics and linguistic theory*. Amsterdam: Rodopi, 2000, p. 51-68.

DE COCK, S. Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literature (BELL)*, new series 2, 2004, p. 225-246.

DÍEZ-BEDMAR, M. B.; CASAS PEDROSA, A. V. The use of prepositions by Spanish learners of English at University level: A longitudinal analysis. In: N. Kübler. (Ed). *Corpora, Language Teaching, and Resources: From theory to Practice*. Bern: Peter Lang, 2011, p. 199-220.

DÍEZ-BEDMAR, M.B.; PAPP, S. The use of the article system by Chinese and Spanish learners. In: G. Gilquin, Sz. Papp; M. B. Díez-Bedmar. (Eds). *Linking up contrastive and learner corpus research*. Amsterdam and New York: Rodopi, 2008, p. 147-175.

DINES, E. Variation in discourse – and stuff like that. *Language in society*, 1, 1980, p. 13-31.

DU BOIS, S. Extension particles, etc. *Language variation and change*, 4, 1993, p. 179-203.

DU BOIS, J. W.; CHAFE, W. L.; MEYER, C.; THOMPSON, S. A.; ENGLEBRETSON, R.; MARTEY, N. Santa Barbara Corpus of Spoken American English, Parts 1-4. Philadelphia: Linguistic Data Consortium, 2000-2005.

EVISON, J., MCCARTHY, M. J.; O'KEEFFE, A. Looking out for love and all the rest of it: vague category markers as shared social space. In: Cutting, J. (Ed). *Vague Language Explored*. Basingstoke: Palgrave, 2007, p. 138-157.

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

EVISON, J. What are the basics of analysing a corpus? IN: O'Keeffe, A.; McCarthy, M. *Routledge Handbook of Corpus Linguistics*. Oxford: Routledge, 2010, p. 122-135.

FERNÁNDEZ, J.; YULDASHEV, A. Variation in the use of general extenders and stuff in instant messaging interactions. *Journal of Pragmatics*, 43, 2610-2626. 10.1016/j.pragma.2011.03.012, 2011.

FRIEDRICH, P. English in Brazil: Functions and attitudes. *World Englishes*, 19, 2, 2000, p. 215-223.

FOX, G. Using corpus data in the classroom. In: Brian Tomlinson. (Ed). *Materials development in language teaching*. Cambridge, 1998, p. 25-43.

GILQUIN, G.; GRANGER, S. Learner language. In: D. Biber; R. Reppen. (Eds). *Cambridge Handbook of Corpus Linguistics*. Cambridge: Cambridge University Press, 2015, p. 418-436.

GILQUIN, G.; PAQUOT, M. Too chatty: Learner academic writing and register variation. *English Text Construction* 1(1), 2008, p. 41-61.

GILQUIN, G., DE COCK, S.; GRANGER, S. *The Louvain International Database of Spoken English Interlanguage*. Handbook and CD-ROM. Louvain-la-Neuve: Presses universitaires de Louvain, 2010.

GRANGER, S. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In: Aijmer K.; Altenberg B.; Johansson M. (Eds). *Languages in Contrast. Text-based cross-linguistic studies. Lund Studies in English*. Lund: Lund University Press, 1996, p. 37-51.

GRANGER, S. The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In: A. Karin. (Ed). *Corpora and Language Teaching*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/scl.33.04gra>, 2009, p. 13-32.

GRANGER, S. A bird's-eye view of learner corpus research. In: Granger, S.; Hung, J.; Petch-Tyson, S. (Eds). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/llt.6>, 2002, p. 3-36.

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

GRANGER, S.; TYSON, S. Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 1996, p. 17-27.

GREAVES, C.; WARREN, M. What can a corpus tell us about multi-word units? In: A. O'Keeffe; McCarthy, M. (Eds). *Routledge handbook of corpus linguistics* [electronic resource]. London: Routledge, 2010, p. 212-226.

HANNAY, M.; MARTÍNEZ-CARO, E. Thematic choice in the written English of advanced Spanish and Dutch learners. In: G. Gilquin, S. Papp; B. Diez. (Eds). *Linking up Contrastive and Learner Corpus Research*. Language and Computer Series. Amsterdam: Rodopi, 2008, p. 227-253.

HELLERMANN, J.; VERGUN, A. Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics*, 39, 2007, p. 157-179.

HONG, H.; CAO, F. Interactional meta discourse in young EFL learner writing: A corpus-based study. *International Journal of Corpus Linguistics*, 19(2), 2014, p. 201-224.

HUNSTON, S. *Corpora in applied linguistics*. Cambridge: Cambridge University Press, 2002.

JEFFERSON, G. List construction as a task and resource. In: G. Psathas. (Ed). *Interaction competence*. Lanham, MD: University Press of America, 1990, p. 63-92.

KIRK, J. 'Kind of and sort of: Pragmatic Discourse Markers in the SPICE-Ireland Corpus'. In: Amador-Moreno, C.; McCafferty, K.; Vaughan, E. *Pragmatic Markers in Irish English*. Amsterdam & Philadelphia: John Benjamins, 2015, p. 89-113.

LAKOFF, R. The logic of politeness Or, minding your p's and q's. In: C. Corum, T. Cedric Smith-Stark, & A. Weiser. (Eds). *Papers from the 9th Regional Meeting of the Chicago Linguistic Society*. Chicago Linguistic Society, 1973, p. 292-305.

LEECH, G. Frequency, corpora and language learning. In: MEUNIER, F.; DE COCK, S.; GILQUIN, G.; PAQUOT, M. (Ed). *A Taste for Corpora. In honour of Sylviane Granger*. Amsterdam/Philadelphia: Benjamins, 2011, p. 7-32.

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

LIN, Y. Vague Language and Interpersonal Communication: An Analysis of Adolescent Intercultural Conversation. *International Journal of Society, Culture & Language*, 1(2), 2013, p. 69-81.

MENDIKOETXEA, A.; LOZANO, C. From corpora to experiments: methodological triangulation in the study of word order at the interfaces in adult late bilinguals (L2 learners). *Journal of Psycholinguistic Research*, 47(4), 2018, p. 871-898.

MELLO, H.; ÁVILA, L.; NEDER NETO, T.; ORFANÓ, B. LINDSEI-BR: an oral English interlanguage corpus. In: VII GSCP International Conference: Speech and Corpora, Belo Horizonte. *Proceedings of the VII GSCP International Conference: Speech and Corpora*. Florença, Itália: Firenze University Press, 2013, p. 85-86.

MORTON, T.; WALSH, S.; O'KEEFFE, A. Analyzing university spoken interaction: a corpus linguistics/conversation analysis approach. *International Journal of Corpus Linguistics*, 16(3), 2011, p. 236-345.

O'KEEFFE, A. 'Like the wise virgins and all that jazz' – using a corpus to examine vague language and shared knowledge". *Language and Computers*, 52(1), 2003, p. 1-20.

O'KEEFFE, A.; ADOLFS, S. Response tokens in British and Irish discourse. In: Schneider, K.; Barron, A. (Eds). *Variational Pragmatics*. Amsterdam: John Benjamins, 2008, p. 69-98.

O'KEEFFE, A.; MCCARTHY, M. J. (Eds). *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 2010.

O'KEEFFE, A.; CLANCY, B.; ADOLPHS, S. *Introducing pragmatics in use*. London, Routledge, 2011.

O'KEEFFE, A., CHENG, W. Vague Language. In K. Aijmer and C. Ruhlemann. (Eds). *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press, pp.686-869, 2015.

ORFANÓ, B. Analysing the use of vague language in spoken interlanguage: A corpus-based study of a group of Brazilian university students learning English as a second language. In: GRANGER, S; GILQUIN, G; MEUNIER, F. *Twenty Years of Learner Corpus Research*:

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

*Looking back, Moving ahead*. Corpora and Language in Use – Proceedings 1. Louvain-la-Neuve: Presses universitaires de Louvain, 2013, p. 367-376.

OVERSTREET, M.; YULE, G. On being explicit and stuff in contemporary American English. *Journal of English Linguistics*, 25(3), 1997a, p. 250-258

\_\_\_\_\_. Locally contingent categorization in discourse. *Discourse Processes*, 23, 1997, p. 83-97.

POOS, D.; SIMPSON, R.C. Cross-disciplinary comparisons of hedging: Some findings from the Michigan Corpus of Academic Spoken English. In: REPPEN, R.; FITZMAURICE, S. M.; BIBER, D. (Eds). *Using Corpora to Explore Language Variation*. Amsterdam: Benjamins, 2002, p. 3-23.

PRODROMOU, L. "You see, it's sort of tricky for the L2-user": *The puzzle of idiomaticity in English as a lingua franca*. Thesis. University of Nottingham, 2005.

RUHLEMANN, C.; CLANCY, B. 'Corpus linguistics and pragmatics'. In: Norrick, N.; Ilie, C. (Eds). *Pragmatics and its Interfaces*. Amsterdam: John Benjamins, p. 241-266, 2018.

SHIRATO, J.; STAPLETON, P. Comparing English vocabulary in a spoken learner corpus with a native speaker corpus: Pedagogical implications arising from an empirical study in Japan. *Language Teaching Research*, 11, 2007, p. 393-412.

SINCLAIR, J. *Corpus, concordance, collocation*. Oxford: Oxford University Press, 1991.

THOMAS, J. Cross-Cultural Pragmatic Failure. *Applied Linguistics* 4(2): 91-112. Winter, S. and Gärdenfors, P. 1995. Linguistic Modality as expressions of Social Power. *Nordic Journal of Linguistics*, 18, 1983, p. 137-166.

THORNBURY, S. What can a corpus tell us about discourse? In: O'Keeffe, A; McCarthy, M. (Eds). *The Routledge Handbook of Corpus Linguistics*, 2010.

MIRANDA, Mateus Emerson de Souza. 'Dreams seem kind of utopic': vague category markers in a learner corpus. *Revista Intercâmbio*, v.XLIV: 84-107, 2020. São Paulo: LAEL/PUCSP. ISSN 2237-759X

VAUGHAN, E.; MCCARTHY, M. J.; CLANCY, B. Vague category markers as turn-final items in Irish English. *World Englishes*, 36 (2), 2017, p. 208-223.

VYATKINA, N.; CUNNINGHAM, D. J. Learner corpora and pragmatics. In: S. Granger, G. Gilquin,; F. Meunier. (Eds). *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press, 2015, p. 281-305.

WARD, G.; B. BIRNER. The semantics and pragmatics of "and everything". *Journal of pragmatics*, 19, 1992, p. 205-214.