

**Universidade Federal de Minas Gerais
Faculdade de Letras**

RUBENS RODRIGUES VIEGAS JÚNIOR

**INVESTIGATING STATIVE VERBS USAGE BY BRAZILIAN LEARNERS
OF ENGLISH AS A SECOND LANGUAGE: A CORPUS BASED STUDY**

Belo Horizonte

2024

RUBENS RODRIGUES VIEGAS JÚNIOR

**INVESTIGATING STATIVE VERBS USAGE BY BRAZILIAN LEARNERS
OF ENGLISH AS A SECOND LANGUAGE: A CORPUS BASED STUDY**

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito para a obtenção do título de Mestre em Linguística Teórica e Descritiva.

Área de Concentração: Linguística Teórica e Descritiva

Linha de Pesquisa: Estudos Linguísticos Baseados em Corpora

Orientadora: Prof^a. Heliana Ribeiro de Mello

Belo Horizonte
Faculdade de Letras da UFMG

2024

V656i

Viegas Júnior, Rubens Rodrigues.

Investigating stative verbs usage by brazilian learners of english as a second language [manuscrito] : a corpus based study / Rubens Rodrigues Viegas Júnior. – 2024.

1 recurso online (121 f.: il., grafs., color., tab., p&b.) : pdf.

Orientadora: Heliana Ribeiro de Mello.

Área de concentração: Linguística Teórica e Descritiva.

Linha de pesquisa: Estudos Linguísticos Baseados em Corpora.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 115-118.

Apêndices: f. 119-121.

Exigências do sistema: Adobe Acrobat Reader.

1. Língua inglesa – Verbos – Teses. 2. Língua inglesa – Estudo e ensino – Falantes estrangeiros – Teses. 3. Linguística de corpus – Teses. I. Mello, Heliana. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 428.07



UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

INVESTIGATING STATIVE VERBS USAGE BY BRAZILIAN LEARNERS OF ENGLISH AS A SECOND LANGUAGE: A CORPUS BASED STUDY

RUBENS RODRIGUES VIEGAS JÚNIOR

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGÜÍSTICOS, como requisito para obtenção do grau de Mestre em ESTUDOS LINGÜÍSTICOS, área de concentração LINGÜÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudos Linguísticos Baseados em Corpora.

Aprovada em 05 de julho de 2024, pela banca constituída pelos membros:

Prof(a). Heliana Ribeiro de Mello - Orientadora

UFMG

Prof(a). Vitor Cordeiro Costa

UFMG

Prof(a). Bárbara Malveira Orfanó

UFMG

Belo Horizonte, 5 de julho de 2024.



Documento assinado eletronicamente por **Heliana Ribeiro de Mello, Professora do Magistério Superior**, em 08/07/2024, às 15:45, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Vitor Cordeiro Costa, Professor do Magistério Superior**, em 09/07/2024, às 15:14, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Barbara Malveira Orfano, Professora do Magistério Superior**, em 10/07/2024, às 11:39, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site

[https://sei.ufmg.br/sei/controlador_externo.php?](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0)

[acao=documento_conferir&id_orgao_acesso_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador 3271405 e o código CRC 70821CD3.

ACKNOWLEDGEMENTS

I want to express my gratitude and immense love to:

God. If You are not first, and all my victories are not to praise Your name, they are meaningless. I know Your ways are better than mine, and You have dreamed much more than a title with this achievement. Thank You for using this master's degree to connect me with many unknown aspects of my life. For this alone, I worship You, my Lord. There is no title, no treasure on Earth capable of buying what I have learned about my existence pursuing this master's degree.

My adviser, for not giving up on me when my internal world collided. I appreciate your support. It means a lot to me.

FAPEMIG, for the financial support given for one year.

My therapist, Renata Fátima. It would be so easy to write an entire thesis to thank you for everything you have done to help me along the way to finish the thesis presented here. How many times I cried, feeling unable to achieve this goal, and there you were, helping me to put the puzzle of my life together. You helped me to understand my behavior patterns and gave new meaning to what I have been through and am still going through in my entire life. Thank you very much.

All my friends, especially Carlos, Ceritelo, Izabelle, Jayne, Kayque, Keren, Marília, Mateus, Selma and Thaís. Your words lifted me and gave me the strength to carry on when the burden was too heavy. I know friends are gifts from God that He uses to bless one's life. I love each one of you.

My family.

My students.

AGRADECIMENTOS

Eu gostaria de expressar a minha gratidão e meu imenso amor a:

Deus. Se o Senhor não for o primeiro e todas as minhas vitórias não servirem para glorificar o Seu nome, elas são insignificantes. Eu sei que os Seus caminhos são melhores que os meus e que o Senhor sonhou muito mais do que um título para mim com esta vitória. Obrigado por usar o mestrado para me colocar em contato com tantos aspectos desconhecidos da minha vida. Isso já é motivo mais que suficiente para glorificá-Lo. Não há título, nem tesouro na Terra capazes de comprar o que eu aprendi sobre a minha própria existência durante o mestrado.

Minha orientadora, por não desistir de mim quando meu mundo interior colidia. Obrigado pelo seu suporte. Isso significa muito para mim.

FAPEMIG, pelo apoio financeiro durante um ano.

Minha terapeuta, Renata Fátima. Seria tão fácil escrever uma dissertação te agradecendo por tudo o que você fez para me ajudar ao longo do caminho para terminar a dissertação aqui apresentada. Quantas vezes eu chorei, me sentindo incapaz de conquistar esse objetivo, e você esteve lá, me ajudando a montar o quebra-cabeça da minha vida. Ajudando a entender meus padrões de comportamento, ajudando a ressignificar o que eu já enfrentei e o que eu tenho enfrentado em minha vida. Muito obrigado.

Todos os meus amigos, em especial Carlos, Ceritelo, Izabelle, Jayne, Kayque, Keren, Marília, Mateus, Selma e Thaís. As palavras de vocês me reergueram e me deram forças para continuar quando o fardo estava pesado demais. Eu sei que amigos são presentes de Deus que Ele usa para nos abençoar. Amo cada um de vocês.

Minha família.

Meus alunos.

“But as for you, be strong and do not give up, for your work will be rewarded.”

2 Chronicles 15:7

“So do not fear, for I am with you; do not be dismayed, for I am your God. I will strengthen you and help you; I will uphold you with my righteous right hand.”

Isaiah 41:10

ABSTRACT

Stative verbs are known for not accepting the progressive form in English (Falhasari et al., 2012). However, a change in the use of these verbs in the progressive form has occurred in the English language, and it is not restricted exclusively to non-standard varieties (Rautionaho, 2020; Martínez-Vázquez 2018). Aiming to investigate the use of stative verbs by Brazilian learners of English as a second language and compare it with the use by native American English speakers, transcriptions from two spoken corpora, the Louvain International Database of Spoken English Interlanguage-Brazil (Mello et al., 2013) and the Santa Barbara Corpus of Spoken American English (Du Bois et al., 2000-2005), were preprocessed using the Notepad++ software (Notepad++ Team, 2023). Subsequently, the verb "be" and other stative verbs belonging to the semantic categories of mental verbs (believe, know, prefer, think, understand), perception verbs (feel, hear, see, smell, taste), physical verbs (appear, associate, exist, look, seem), possession verbs (belong, have, hold, keep, own), and relationship verbs (dislike, hate, like, love, trust) were analyzed using LancsBox X software (Brezina; Platt, 2023). The results indicate that Brazilian learners and native American speakers share surprising similarities regarding the frequency and context of usage of the selected stative verbs. Little difference was found regarding the use of progressive forms and the semantic categories in which these verbs were grouped. The results here highlight the similarities in SV usage across different groups. This research contributes to learners' understanding of using stative verbs in English.

Keywords: stative verbs, progressive form, learners, corpus linguistics.

RESUMO

Verbos estativos são conhecidos por não aceitarem a forma progressiva no inglês (Falhasari et al., 2012) porém, uma mudança em relação ao uso desses verbos no progressivo vem acontecendo na língua inglesa, e ela não se restringe exclusivamente às variantes não padrão (Rautionaho, 2020; Martínez-Vázquez 2018) Com o objetivo de investigar o uso de verbos estativos por Brasileiros aprendizes de inglês como segunda língua, e comparar com o uso de falantes nativos do inglês americano, as transcrições de dois corpora orais, o Louvain International Database of Spoken English Interlanguage-Brazil (Mello et al., 2013), e o Santa Barbara Corpus of Spoken American English (Du Bois et al., 2000-2005), foram pré-processadas através do software Notepad++ (Notepad++ Team, 2023) e posteriormente o verbo “be” e outros verbos estativos pertencentes às categorias semânticas de verbos mentais (believe, know, prefer, think, understand), de percepção (feel, hear, see, smell, taste,), de sensações físicas (appear, associate, exist, look, seem), de posse (belong, have, hold, keep, own), e relacionamento (dislike, hate, like, love, trust) foram analisados com a ajuda do software LancsBox X (Brezina; Platt, 2023). Os resultados indicam que os aprendizes brasileiros e os falantes nativos americanos compartilham de semelhanças surpreendentes em relação a frequência e contexto de uso dos verbos estativos selecionados. Pouca diferença foi encontrada em relação ao uso das formas progressivas e as categorias semânticas nas quais esses verbos foram agrupados. A pesquisa aqui conduzida, contribui para a compreensão do uso dos verbos estativos no inglês por aprendizes.

Palavras-chave: verbos estativos, forma progressivo, aprendizes, linguística de corpus.

LIST OF IMAGES

Figure 3.1.1: BR001 Transcription screenshot	32
Figure 3.1.2: SBC001 Transcription screenshot	34
Figure 3.2.1: BR001 Preprocessed File	37
Figure 3.2.2: SBC001 Preprocessed File	40
Figure 3.3: LancsBox X main screen overview	41
Figure 3.3.1: LancsBox X Add Corpora screen	42
Figure 3.3.1.1: LancsBox X KWIC screen	43
Figure 3.3.2: LancsBox X GraphColl screen	44
Figure 3.3.3: LancsBox X Words Tool screen	45
Figure 3.4.4: LancsBox X Text Tool screen	46
Figure 3.3.6: Verb “be” VB query	49
Figure 3.3.6.1: Verb “be” VB query with semantic tag	49
Figure 4.3: USAS discourse domains	72

LIST OF GRAPHS

Graph 4.1: LINDSEI-BR SVs occurrences	52
Graph 4.1.2: LINDSEI-BR mental verb occurrences	53
Graph 4.1.3: LINDSEI-BR perception verb occurrences	55
Graph 4.1.4: LINDSEI-BR physical verb occurrences	56
Graph 4.1.5: LINDSEI-BR possession verb occurrences	58
Graph 4.1.6: LINDSEI-BR relationship verb occurrences	59
Graph 4.2 SBCSAE SVs occurrences	61
Graph 4.2.2: SBCSAE mental verb occurrences	63
Graph 4.2.3: SBCSAE perception verb occurrences	65
Graph 4.2.4: SBCSAE physical verb occurrences	67
Graph 4.2.5: SBCSAE possession verb occurrences	69
Graph 4.2.6: SBCSAE relationship verb occurrences	70
Graph 4.5: Comparison between LINDSEI-BR and SBCSAE SVs widespread usage	106

LIST OF TABLES

Table 3.2.1: Regex used in the preprocessing of LINDSEI-BR	35
Table 3.2.2: Regex used in the preprocessing of SBCSAE	38
Table 3.3.6: POS tags	47
Table 4.1.1: LINDSEI-BR verb “be” occurrences	53
Table 4.1.2: LINDSEI-BR mental verb tense occurrences	54
Table 4.1.3: LINDSEI-BR perception verb tense occurrences	55
Table 4.1.4: LINDSEI-BR physical verb tense occurrences	57
Table 4.1.5: LINDSEI-BR possession verb tense occurrences	58
Table 4.1.6: LINDSEI-BR relationship verb tense occurrences	60
Table 4.2.1: SBCSAE verb “be” tense occurrences	62
Table 4.2.2: SBCSAE mental verb tense occurrences	64
Table 4.2.3: SBCSAE perception verb tense occurrences	66
Table 4.2.4: SBCSAE physical verb tense occurrences	68
Table 4.2.5: SBCSAE possession verb tense occurrences	69
Table 4.2.6: SBCSAE relationship verb tense occurrences	71

LIST OF ABBREVIATIONS

CL – Corpus Linguistics

CQL – Corpus Querying Language

L2 – Second Language

LINDSEI-BR – Louvain International Database of Spoken English Interlanguage-Brazil

Regex – Regular Expressions

SBCSAE – Santa Barbara Corpus of Spoken American English

SLA – Second Language Acquisition

SVs – Stative Verbs

VB – Base Form

VBD – Past Form

VBG – Gerund or Present Participle Forms

VBN – Past Participle Form

VBP – Non 3rd Person Singular Present Forms

VBZ – 3rd Person Singular Present Forms

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	15
1.1 Objective	16
1.2 Outline of the chapters	16
CHAPTER 2: LITERATURE REVIEW	18
2.1 Corpus Linguistics	18
2.2.1 Written corpora	20
2.2.2 Spoken corpora	21
2.2.3 Learner corpora	24
2.2 Stative Verbs	25
2.2.1 An overview of stative verbs	25
2.2.2 Stative Verbs and Recent Studies.....	28
CHAPTER 3: METHODOLOGY	31
3.1 The corpora	31
3.1.1 LINDSEI-BR corpus	31
3.1.2 SBCSAE corpus	33
3.2 Notepad++	34
3.2.1 Preprocessing of LINDSEI-BE corpus	34
3.2.2 Preprocessing of the SBCSAE	37
3.3 LancsBox X	40
3.3.1 KWIC Tool (Key Word in Context)	42
3.3.2 GraphColl	43
3.3.3 Words Tool	44
3.3.4 Text Tool	45
3.3.5 Searching in LancsBox X	46
3.4 Querying LINDSEI-BR and SBCSAE	47
CHAPTER 4: RESULTS AND DISCUSSION	51
4.1 LINDSEI-BR SVs occurrences	51
4.1.1 LINDSEI-BR verb “be” occurrences	52
4.1.2 LINDSEI-BR mental verb occurrences	53
4.1.3 LINDSEI-BR perception verb occurrences	54
4.1.4 LINDSEI-BR physical verb occurrences	56
4.1.5 LINDSEI-BR possession verb occurrences	57

4.1.6 LINDSEI-BR relationship verb occurrences	59
4.2 SBCSAE SVs occurrences	60
4.2.1 SBCSAE verb “be” occurrences	61
4.2.2 SBCSAE mental verb occurrences	62
4.2.3 SBCSAE perception verb occurrences	64
4.2.4 SBCSAE physical verb occurrences	66
4.2.5 SBCSAE possession verb occurrences	68
4.2.6 SBCSAE relationship verb occurrences	70
4.3 Semantic Tags	71
4.3.1 Semantic Tags for Be in LINDSEI-BR	72
4.3.2 Semantic Tags for Mental Verbs in LINDSEI-BR	73
4.3.3 Semantic Tags for Perception Verbs in LINDSEI-BR	73
4.3.4 Semantic Tags for Physical Verbs in LINDSEI-BR	73
4.3.5 Semantic Tags for Possession Verbs in LINDSEI-BR	74
4.3.6 Semantic Tags for Relationship Verbs in LINDSEI-BR	74
4.3.7 Semantic Tags for “Be” in SBCSAE	75
4.3.8 Semantic Tags for Mental Verbs in SBCSAE	75
4.3.9 Semantic Tags for Perception Verbs in SBCSAE	76
4.3.10 Semantic Tags for Physical Verbs in SBCSAE	76
4.3.11 Semantic Tags for Possession Verbs in SBCSAE	77
4.3.12 Semantic Tags for Relationship Verbs in SBCSAE	78
4.4 Collocations	78
4.4.1 Be Collocates	79
4.4.2 Mental Verbs Collocates	82
4.4.3 Perception Verbs Collocates	90
4.4.4. Physical Verbs Collocates	97
4.4.5 Possession Verbs Collocates	99
4.4.6 Relationship Verbs Collocates	102
4.5 LINDSEI-BR and SBCSAE final remarks	106
CHAPTER 5: CONCLUSION	112
REFERENCES	115
APPENDIX A	119
APPENDIX B	120
APPENDIX C.....	121

CHAPTER 1

INTRODUCTION

In English, stative verbs (SVs) are well known for not accepting the progressive form (Falhasari et al., 2012), which is often said to be a characteristic of the so-called action or dynamic verbs. However, as stated by Hundt (2004), the use of the progressive in English has been growing in scope and frequency for at least two centuries. Freund (2016) also argues that an area that has continuously changed in current English is linked to the expansion of SVs usage in the progressive (e.g., **I'm enjoying** your new jacket).

Recent studies have shown that for learners with a higher level of proficiency, the use of SVs in the progressive form was more unacceptable when compared to native speakers (Falhasari et al., 2012); a continuous shift towards a more grammatical status of using the verb *love* in the progressive (Martínez-Vázquez, 2018); and that the use of SVs appeared in all tenses, except for the future progressive (Hanum, 2018). Bearing that in mind, this thesis aims to describe SVs usage among Brazilian learners of English as a Second Language and compare the results found in the former group to SVs usage among American English native speakers.

To carry out this study, two spoken corpora were selected: the Louvain International Database of Spoken English Interlanguage-Brazil (LINDSEI-BR) (Mello et al., 2013) and the Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois et al., 2000-2005). The methodology adopted here originates in Corpus Linguistics (CL), which has been proven helpful in empirical investigations as it allows the analysis of enormous amounts of data using computer software (McEnery and Hardie, 2012), which facilitates and saves time. In addition, corpus-based research has also contributed to the study of languages (Biber et al., 2021; Biber, 1991)

LINDSEI-BR and SBCSAE transcripts contain markers used to signalize speech features such as reactions and overlapping, and SBCSAE also contains timestamps in its transcripts. As such information is irrelevant to this study, they needed to be removed from the corpora. The Notepad++ software (Notepad++ Team, 2023) was used to achieve such a goal. Regular expressions (Regex) and the search for specific characters were applied in the software to make it possible to find and remove the necessary items. After that, the transcripts of both corpora were uploaded to LancsBox X software (Brezina; Platt, 2023) to be queried. LancsBox X software was used to tag

the corpora with Part of Speech and Semantic tags, and through Corpus Querying Language (CQL), the search for the frequency of occurrences, which verb forms and semantic contexts the verbs were used, and collocates was done.

The results found in this study can contribute to the existing literature on this topic and provide additional insights that might be useful for English language teaching.

1.1 Objective

The main objective of this thesis is to verify in a corpus how Brazilian learners of English use SVs. More specifically, the transcripts of semi-spontaneous spoken interactions of Brazilian learners of English will be examined through the analysis of selected SVs, and the results will be compared to occurrences of the same verbs in an American native spontaneous speech corpus. For that purpose, the following are the specific objectives of this research:

- Quantify the number of occurrences of the selected SVs.
- Identify the semantic category that is used the most.
- Identify the SVs that occur the most in the progressive form.
- Identify in which form the SVs were used the most.
- Identify the semantic contexts in which such verbs were used.
- Identify the collocates of the base form of these verbs.
- Compare the results found in the learner corpus to those found in the native speaker corpus.

1.2 Outline of the chapters

This thesis is structured into five chapters. As presented above, the introductory chapter offers an overview of the research topic, details about the selected corpora, and the software tools used for data processing and querying. Additionally, the objectives of this study are outlined in the latter part of the first chapter.

Chapter two provides a concise review of the literature supporting this study. It includes information on corpus linguistics as a methodology and details on the compilation of written, spoken, and learner corpora. The chapter also delves into previous discussions on SVs.

Chapter three elaborates on the methodology adopted in this thesis. It presents the two corpora used, introduces the software tools Notepad++ and LancsBox X, and explains the data processing and querying procedures.

Chapter four is dedicated to presenting and discussing the findings. It begins with analyzing the frequency of occurrences, followed by examining the semantic contexts in which the SVs are used. The chapter identifies the five strongest collocates of the base form of the verbs across both corpora, where applicable. It concludes by comparing these results to the literature reviewed in chapter two.

The final chapter, Chapter Five, summarizes the thesis, addressing key findings and discussing limitations encountered during the research. Following Chapter Five, the thesis includes the References and an Appendix.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we will first define CL as a field of study in linguistic research. Then, we will provide a brief overview of written, spoken, and learner corpora, including the process and guidelines followed by researchers during the compilation of these corpora. Special attention is paid to spoken corpora due to their relevance to the nature and purpose of this work. Finally, the differences between stative and action verbs are presented, and the semantic categories SVs can be grouped (mental, perception, physical, possession, and relationship). This chapter ends with information about some research findings about SVs in the context of CL and experiments.

2.1 Corpus Linguistics

According to Pearson (1998), linguists may assume different approaches when explaining what a corpus is, and some of the possible attempts are as follows:

- a) a group of texts chosen and arranged under clear linguistic standards to serve as a sample of the language (Sinclair, 1994).
- b) A subset of an electronic text library constructed following clear design standards for a particular objective (Atkins, Clear, and Ostler, 1992).
- c) A group of texts that can be used for linguistic analysis and are thought to reflect a particular language, dialect, or other subsets of languages (Francis, 1992).
- d) Several pieces of machine-readable text chosen to represent a language or variety (McEnery and Wilson, 1996).

Although authors may not wholly agree with the definitions mentioned above, they all agree that a corpus is not made of randomly selected data. However, how is such data selected? Sinclair (1991, p.13) states that “certainly the bulk of any lay discussion about corpora concerns the criteria for text¹ selection,” and according to him, the first thing to be considered in the process is the aim of the corpus being

¹ McEnery and Hardie (2012, p.2), define *text* as a file of machine-readable data. According to them, these files do not need to be textual, and there are certainly examples nowadays of files of video data being used as corpus texts.

compiled². This is because Linguistics can be branched into many areas (Lindquist and Levin, 2018).

For example, there are studies in sociolinguistics (focused on the relationship between society and language), psycholinguistics (focused on the relation between mind and language), neurolinguistics (focused on the relationship between neurological processes in the brain and language), and so on, thus, defining the aim of a study will guide the compiler through the other stages in the compiling process.

McEnery and Hardie (2012) state that CL can be considered different from any other topic studied in linguistics because it does not focus on any specific aspect of a language. For them, CL can be considered a valuable methodology in qualitative and quantitative studies. Thus, this approach can be used in any area of Linguistics because it offers many procedures for studying languages.

For Sinclair (1991), the design of the corpus would be the second thing to be considered. A corpus can be made of written texts, spoken transcriptions, or both, although he also states that most corpora “keep well away from problems of spoken language” (Sinclair, 1991, p. 15). This may be related to the fact that compiling a spoken corpus is much more laborious, time-consuming, and expensive (McEnery and Hardie, 2012; Mello, 2014).

When discussing the design of a corpus, Biber (1993, p. 243) states that:

“Some of the first considerations in constructing a corpus concern the overall design: for example, the kinds of texts included, the number of texts, the selection of particular texts, and the length of text samples. Each of these involves a sampling decision either conscious or not.”

These things are relevant in CL because a corpus should offer balance and representativeness. Without the latter, assumptions about a language as a whole cannot be made, making corpus-based grammars and dictionaries, for instance, impossible (Biber, 1993).

When considering the size of a corpus, Davies (2015) states it can vary from a small size (1-5-million-word approximately) to the web being used as a corpus (containing billions of words). The author exemplifies the varied sizes with the following corpora:

² The process of building a corpus is called compilation.

1. Small 1–5-million-word, first-generation corpora like the **Brown Corpus** (and others in the Brown “family,” such as the LOB, Frown, and FLOB).
2. Moderately sized, second-generation, genre-balanced corpora, such as the 100-million-word **British National Corpus**.
3. Larger, more up-to-date (but still genre-balanced) corpora, such as the 450-million-word **Corpus of Contemporary American English** (COCA).
4. Large text archives, such as Lexis-Nexis.
5. Extensive text archives, such as Google Books.
6. The Web as a corpus, seen here through the lens of Google-based searches.

In CL, computers are used to “read, search and save the time of human analysts when manipulating the data” (McEnery and Hardie 2012, p. 2), and as a corpus may contain millions of words, “it is certainly tough to search such a large corpus by hand in a way which guarantees no error” (McEnery and Hardie 2012, p. 2). However, Biber (1993) states that a corpus is not a selection of random files to be used in an analysis. When using CL to answer a research question, “the corpus data must be well matched to that research question” (McEnery and Hardie 2012, p. 2).

Diverse types of corpora are available for research, such as written language corpora, spoken language corpora, historical corpora, first language learning corpora, second and foreign language learning corpora, parsed corpora, and multimodal corpora. The purpose of this work is not to explore and elaborate on every type of corpora; thus, only a glimpse of written, spoken, and learner corpora is given in the following pages.

2.1.1 Written corpora

According to the number of corpora available for research, there is no question that corpora of written texts top the list. This is because compiling such corpora is more accessible, cheaper, and quicker than other corpora (McEnery and Hardie, 2012; McEnery and Brookes, 2022). However, we are not to say that building a written corpus does not involve challenges and difficulties.

One of the first things to consider when compiling a corpus is its purpose. Knowing the answer to this question allows the compiler to make smarter choices and achieve better results in the designing process (McEnery and Hardie, 2012; McEnery and Brookes, 2022).

Available written corpora can be split into two main categories:

1. *specialized corpora*, designed to represent a specific genre or variety of language and
2. *general corpora*, which are much larger since they are thought to represent a language and its use on a broader scale.

Therefore, an existing general corpus will serve well if the researcher aims to study a particular linguistic feature on a broad scale (McEnery and Hardie, 2012; McEnery and Brookes, 2022). However, building their corpus may be wiser if the researcher intends to study a particular topic within a particular period (McEnery and Hardie, 2012; McEnery and Brookes, 2022).

As mentioned before, when compiling a corpus, no matter its type, one should achieve representativeness (balance and size). However, another thing that must be considered is authenticity. McEnery and Brookes (2022) define authenticity as the natural quality of language. They also state that “capturing and including authentic language in a corpus tends to pose much less of an obstacle for the collection of written texts than it does for spoken texts” (McEnery and Brookes 2022, p. 36). As mentioned by the authors, some problems concerning authenticity in written corpora are the nonstandard spelling of words and the volume of online language produced by so-called social media “bots”³.

2.1.2 Spoken corpora

Despite their importance and contributions to CL studies, Lindquist and Levin (2018) state that spoken language corpora are insufficiently represented in CL. As mentioned before, this is because compiling a spoken corpus costs more money and is technically more challenging, thus contributing to the lack of abundance of such corpora compared to written corpora.

Spoken corpora can be split into two different categories, according to Mello (2014): 1) non-spontaneous spoken corpora that encompass "planned speech, read written text, scripted speech, and even task-based prepared speech" (Mello, 2014, p.

³ Bots can automatically generate large volumes of social media content very quickly. Such content can also be difficult for the untrained eye to distinguish from content that is produced by humans. (McEnery and Brookes, 2022, p. 36)

29), and 2) spontaneous spoken corpora that include "non-planned speech that is performed at the same time that it is structured" (Mello, 2014, p. 29).

It is worth mentioning that before computers started to be used to compile and analyze corpora, the so-called early corpora, those dating from the 1980s, were primarily spoken-based (Staples, 2015). However, those corpora were often small and were used to study phonetic features, leading to significant criticism from researchers who would claim such corpora, because of the size, did not represent speech and could not be used in quantitative analyses (Staples, 2015).

Adolphs and Knight (2014) state that, besides all the challenges and difficulties in compiling spoken corpora, because of the contributions they have made, we can see a strong interest in the development of spoken corpora. Moreover, Mello (2014) argues that in the past 20 years or so, the compilation of spontaneous speech corpora has grown significantly due to technological advances.

Concerning the compilation of spoken corpora, Sinclair (2005) has offered some guidelines that are still useful to date in the compilation of both written and spoken corpora. According to Sinclair (2005), corpus contents should be chosen based on their communicative function in the community of origin, regardless of the language they encompass. Furthermore, builders of corpora should endeavor to ensure that their corpus is as reflective as possible of the language it represents. Additionally, only components of corpora explicitly designed for independent contrast should be compared. Moreover, criteria for structuring a corpus should be few, distinctly separate, and collectively efficient in outlining a representative language or variety. In addition, information beyond the alphanumeric string of words and punctuation in a text should be stored separately and integrated when necessary for applications. Whenever feasible, language samples for a corpus should consist of complete documents or transcriptions of entire speech events or closely approach this objective, resulting in substantial variations in sample size. Also, a corpus's complete design and composition should be thoroughly documented, including information on content and the rationale behind decisions made. Importantly, corpus builders should prioritize notions of representativeness and balance, even though these goals are not precisely defined or entirely achievable; they should guide the corpus design and component selection. Moreover, subject matter control in a corpus should be based on external, not internal, criteria. Finally, a corpus should seek component homogeneity while ensuring sufficient coverage.

When discussing the architecture of a spontaneous speech corpus, Mello (2014) states that it must contain formal and informal language registers. Moreover, it is necessary to have a well-prepared team to work on recording the audio, transcribing, aligning it, and segmenting it depending on the study type to be carried out. Mello (2014, p. 33) also argues that “to achieve the largest possible diaphasic variation, it is more relevant to record a larger number of spoken texts (especially informal ones, which are the most relevant) than to have longer and fewer texts”.

According to Mello (2014), speech can be of two different natures - informal and formal - the former is the most important one because that is the moment speakers are freely using language without paying attention to or controlling it. Mello (2014) also discusses how the interactions can be grouped (public, private, family), pointing out that this task should not be done intuitively. As stated by her, speech corpus should contain monologue texts (the development depends primarily on one person), dialogue texts (two people interacting), and conversation texts (three or more people interacting among them), and the compiler must pay close attention to achieve balance in the distribution among these kinds of texts to achieve speech variation.

Another essential aspect to observe when compiling a spoken corpus is the availability of metadata (information about the speaker, usually an acronym, such as age, profession, gender, number of words in the text, etc.). Mello (2014, p. 48) states that “Metadata should be recorded as soon as possible after recording to guarantee the preservation of fundamental information that otherwise might be easily lost or forgotten.”

The quality of the equipment to be used in the recording process is also of immense importance, as stated by Mello, and it must include the following:

“a portable digital recorder with large memory storage and long duration batteries; wireless lapel microphone system with receiver and transmitter in the least invasive format possible; mixer system that allows for a large number of microphone inputs (for several lapel microphones to be used by interactants in conversations). It is useful to have an omnidirectional microphone with a tripod and radio transmission system. An omnidirectional microphone can be used to record conversations in stationary situations and very little gain environments, such as outdoors, for example. These are few situations; However, omnidirectional microphones are advantageous because they are even less invasive than lapel ones.” (MELLO, 2014, p. 49)

It is essential to pay attention to such details because the quality of the equipment adopted and the acoustic quality of the audio recorded are linked, and audio with a substandard quality cannot be used in phonetic studies, for example (Mello, 2014). Concerning the microphones to be adopted, Mello states they must be as less invasive as possible so that the participants can forget them. Thus, the first minutes of the recording serve as a “warm-up,” as they lead the participants to forget they are being recorded.

As previously mentioned, spoken corpora can be segmented, which helps study tonal units. Mello (2014) highlights the importance of having a well-trained team to work on the corpus segmentation and statistically validate the results found. We will not discuss this topic in this thesis because it is not of paramount importance to the analysis proposed here. Another feature to be considered is the annotation of the corpus. For this task, free software is available to assist the compiler.

To conclude this overview of spoken corpora, the last feature to be discussed is the transcription of a spoken corpus. It can be phonetic, prosodic, or orthographic according to the needs of its compiler (Mello 2014). As for the segmentation, the transcription of a spoken corpus also requires a well-trained team to work on this task to avoid mistakes (Mello 2014). Mello (2014, p. 58) points out that the CHAT⁴ architecture is a “well-established example of transcription parameters to be followed”. She also states that “corpus transcription is a chore that, necessarily, needs to be well planned, and one the reasons is that “transcription criteria define, to a large extent, the kinds of research that will be allowed by the corpus data” (Mello 2014, 58).

2.1.3 Learner corpora

According to Granger et al. (2015), studies related to second language acquisition (SLA) are not new in linguistics; however, with the rise of CL, these studies have seen a significant advance in their methodology. Before CL, the data used in such studies were restricted to a small number of students or just one individual. Therefore, a lack of representativeness and naturalness (due to highly controlled tasks) was a reality (Granger et al., 2015).

⁴ see MacWhinney, 2000 for more information.

Granger et al. (2015) argue that, like any corpus, the learner corpus is defined as a collection of machine-readable authentic texts (including transcripts of spoken data) sampled to be representative of a particular language or language variety. What distinguishes the learner corpus is its representation of language as generated by foreign or second language (L2) learners (Granger et al., 2015).

When discussing the typology of learner corpora, Granger et al. (2002) state that learner corpora are typically monolingual, although a limited number of learner translation corpora have been assembled. Moreover, existing learner corpora include samples of non-specialist language. Additionally, contemporary learner corpora often have a synchronic nature, meaning they depict learner use at a specific moment in time. The existence of longitudinal corpora, which track the progression of learner use over an extended period, is quite rare. Granger et al. (2002) also argue that this scarcity is attributed to the substantial challenges in compiling such corpora, as they necessitate tracking a learner population over months or, preferably, years. However, researchers interested in understanding the development of learners' proficiency commonly gather "quasi-longitudinal" data, where information is collected from a homogeneous group of learners at various proficiency levels (Granger et al., 2002).

Learner corpora have immensely contributed to developing material used in language teaching (Granger et al., 2002; Granger et al., 2015). According to McEneary and Xiao (2011), it can help in the following areas: 1) Syllabus Design and Materials Development. 2) Language Testing, and 3) Teacher Development.

2.2 Stative Verbs

The first part of this chapter describes how SVs behave in the English language, focusing on their semantics. The second part presents information derived from the results of recent studies about the use of SVs in the progressive form.

2.2.1 An Overview of Stative Verbs

According to Rothmayr (2009, p. 3), "Vendler 1957 was the first to group verbs along the lines of event structure". In his paper on Aspect, dated 1957, Vendler (1957, p. 143) argued that "the use of a verb may also suggest the particular way in which the verb presupposes and involves the notion of time". However, the possible distinctions

made among verbs that suggest processes, states, dispositions, occurrences, and so on could not be explained in terms of time alone; factors such as “the presence or absence of an object, conditions, intended states of affairs, also enter the picture” (Vandler. 1957, p. 143). Bearing that in mind, he grouped verbs into four distinct types. Firstly, state verbs, which are atelic, and the situations they describe are durative and do not involve change. Secondly, activity verbs, like state verbs, are atelic, and the situations they describe are durative but involve change. Thirdly, accomplishment verbs, that are telic, and the situations they describe are durative and involve change. Finally, achievement verbs, being telic, and the situations they describe are non-durative but involve change.

The first distinction discussed by Vandler (1957) involves the difference between “verbs that possess the continuous tenses from verbs that do not” (p. 144). To exemplify, he explains that the question “What are you doing?” can be answered with something like “I am running,” but not with “I am knowing”. On the other hand, it is possible to ask something like “Do you know...?” and have as an answer “Yes, I do.”, but something like “Do you run?” “Yes, I do” may feel awkward. That is because running and other verbs, such as writing, describe a process going on in time. Thus, they “consist of successive phases following one another in time” (Vandler, 1957, p. 144) and are compatible with the progressive tense.

However, as stated by Vandler (1957), depending on the context of verbs compatible with the progressive tense are used, there may be a change in their meaning or, let us say, in the idea the speaker wants to convey. If we say, “Someone is running or pushing a cart” and compare it with “Someone is running a mile or drawing a circle,” we can assume that running or pushing a cart has no terminal point. It may take some time, but there is no set amount of time that it must take.

Vandler (1957) continues to explain that, on the other hand, running a mile or drawing a circle, are actions that need to be completed to say someone ran a mile or drew a drawing. They may also go on for a time, but it takes a particular time to draw a circle or run a marathon. Therefore, running a marathon and drawing a circle must be finished, but talking about finishing “running a cart” or “pushing a cart” makes no sense. Vendler (1957, p.146) then calls “the first type, that of “running,” “pushing a cart,” activity terms, and the second type, that of “running a mile” and “drawing a circle,” accomplishment terms”. Now, let us turn to the verbs that lack the continuous aspect. Vendler (1957) states that verbs like “knowing” and “recognizing,” for example, do not

indicate processes going on in time, even though they might be based on a topic for a certain period and either be true or false. Moreover, some of the verbs in this group “can be predicated only for single moments, while others can be predicated for shorter or longer periods of time” (Vandler, 1957, p. 146).

Consider the following situations used by Vendler (1957) to compare the verbs reach and win to the verbs know and believe. According to him, someone reaches the top or wins a race at a definite moment. Therefore, these situations will not be factual until the top is reached and the race is won. On the other hand, we love or recognize something for a short or a long time, i.e., there is no definite time. Vendler (1957) calls verbs like “reach” and “win” achievement verbs, whereas verbs like “love” and “recognize” are called stative verbs. Now that some of the characteristics of all the groups proposed by Vendler (1957) have been presented let us focus on the group most relevant to this thesis, the stative one.

According to Vendler (1957), verbs such as “to know,” “to think,” and “to understand” fall within the group of verbs that have conceptual divergences of their own. For example, the verb “to know” can be used in two basic senses: the first one to refer to a process, as in “He is thinking about Jones,” and the second one to refer to a state, as in “He thinks that John is a rascal,” thus, “the first sentence can be used to describe what a person is doing, whereas the second cannot” (Vendler 1957, p. 152).

Saeed (2009), when discussing the aspects of meaning at the level of the sentence, also argues that stative verbs classify situations into two diverse types: static and dynamic. In static situations, the speaker wants to convey steady information. There is no focus on the moment it began or ended, nor on the internal phases or changes. Saeed (2009) also points out that stative verbs do not accept the progressive form in English, additionally, he states that verbs can have more than one meaning, and some meanings can be more stative than others, which allows the use of the progressive form, for example, the verb have.

- I have a car. (state)
- *I'm having a car.
- I am having second thoughts about this. (dynamic)

Grammar books, such as Biber et al (2021) usually mention SVs as non-compatible with the progressive form, as mentioned before, and SVs often refer to:

- thoughts and opinions: *agree, believe, doubt, guess, imagine, know, mean, recognize, remember, suspect, think, understand*
- feelings and emotions: *dislike, hate, like, love, prefer, want, wish*
- senses and perceptions: *appear, be, feel, hear, look, see, seem, smell, taste*
- possession and measurement: *belong, have, measure, own, possess, weigh.*

As stated by Biber et al. (2021), dynamic verbs are described as freely occurring in the progressive aspect, on the other hand, verbs with stative senses are described as not occurring in the progressive. “However, it turns out that both dynamic and SVs are included among the most common verbs in the progressive – and that both dynamic and SVs are included among the verbs that rarely take the progressive” (Biber et al., 2021, p. 471) Sawn (2016, p. 42), also argue that “occasionally 'non-progressive' verbs are used in progressive forms to emphasize the idea of change or development.

Examples:

- 1) *These days, more and more people **prefer** / **are preferring** to retire early.*
- 2) *The water **tastes** / **is tasting** better today.*
- 3) *As I get older, I **remember** / I **'m remembering** less and less.*
- 4) *I'm **liking** it here increasingly as time goes by.*

2.2.2 Stative Verbs and Recent Studies

Rautionaho (2020) conducted a study aimed at investigating the stative progressive usage in world Englishes (outer and inner circle varieties). According to Rautionaho (2020), many authors claim that the use of stative progressive is associated with outer circle varieties of English, and this is known as a phenomenon that differentiates such varieties from the inner circle ones. In her study, Rautionaho (2020) compared three outer-circle English varieties (Indian English, Hong Cong English, and Nigerian English) to three inner-circle varieties (British English, Irish English, and New Zealand English), and tested if such declaration about the outer-circle varieties was indeed valid.

Rautionaho (2020) claims that previous studies showed that there is a consensus on the following: SVs are “more frequent in spoken language, in the present tense, and without the occurrence of modal auxiliaries” (Rautionaho, 2020, p. 8).

Rautionaho (2020) also points out that considering the semantic categories⁵ SVs are grouped (relational, cognitive, affective, perception, stance), “relational and cognitive verbs are usually the least compatible with the progressive” (Rautionaho, 2020, p. 6), “affective verbs, do occur with the progressive when they express temporariness or highlight an emotion” (Rautionaho, 2020, p. 6), and the “two remaining categories are considered ‘less’ stative, in the sense that some perception verbs and most stance verbs may be used almost interchangeably in the progressive and the non-progressive” (Rautionaho, 2020, p. 6).

Based on such information and using data that comes from the International Corpus of English (ICE), Rautionaho (2020) results showed that all the varieties investigated had occurrences of stative progressive forms, however, Indian English was the one with the most occurrences. According to Rautionaho (2020), this is not enough to establish that outer circle varieties present a wider use of the progressive. Moreover, this affirmation must be clearly defined, as the other two outer circle varieties did not present the same results in Indian English.

Another study aiming at understanding more about the use of SVs in the English language was conducted by Martínez-Vázquez (2018). Using three different corpora, the Corpus of Contemporary American English (COCA), the Corpus of Global Web-Based English (GloWbE), and the Corpus of Historical American English (COHA), Martínez-Vázquez (2018) investigated the usage of the verb “love” in inner and outer varieties of English as well. According to her, “to fully understand these emerging uses of the so-called progressive states, we must consider the syntactic elements of the construction and the meaning and pragmatics of the so-called stative verb” (Martínez-Vázquez, 2018, p. 1), so focusing on that, the results found by Martínez-Vázquez (2018) showed that the use of the verb “love” in the progressive form has grown in scope and frequency, especially in the last decade. Aiming at comparing the results, Martínez-Vázquez (2018) also analyzed the verb “like”. She states that her results showed that “like, has a much lower incidence and less variety of uses in the progressive than love” (Martínez-Vázquez, 2018, p. 22).

⁵ It seems there is no consensus about the names of the semantic categories among authors who have worked with SVs. The categories names presented by Rautionaho (2020) slightly differ from the ones adopted in this thesis. Here, as presented in Chapter 4, the categories adopted are mental, perception, physical, possession, and relationship.

Falhasari et al. (2012) did not conduct a study based on corpus linguistics, however, the results found by the authors can also add to the results of the studies previously mentioned. Falhasari et al. (2012) tested the judgment of 70 learners of English as a foreign language on the use of the following SVs: *like*, *coast*, *realize*, *know*, *believe*, *understand*, *hate*, and *doubt*. In the test, these verbs were used in the progressive form in sentences and the participants had to determine whether the sentences were grammatically correct or not. If the decision was not correct, participants should then fix the mistake in a way that was grammatically accepted.

The results obtained by Falhasari et al. (2012) showed that advanced learners considered the use of the progressive verbs previously mentioned unacceptable compared to the native participants. The native speakers considered the usage of the progressive form acceptable with most of the verbs. Therefore, it is possible to conclude that the learners follow the rules presented in grammar books that state these verbs cannot be used in the progressive.

We can finish this chapter assuming that SVs in the progressive form will be found in the corpora used in this thesis. However, to what extent and how spread in the semantic categories (mental, perception, physical, possession, and relationship) are they?

CHAPTER 3 METHODOLOGY

This chapter provides details about the two corpora used in this thesis and the software adopted to query both corpora. We start by explaining how the two corpora were compiled and their availability. Then, we explain how the texts of the two corpora were preprocessed to be used in this study presenting the software adopted for this stage. The last part of this chapter deals with the querying process of both corpora using the software LancsBox X.

3.1 The corpora

This section presents the two corpora selected for this study. Both corpora were chosen because they were compiled under excellent parameters and guidelines aiming at achieving balance and representativeness, which is of paramount importance for studies carried out using this methodology.

3.1.1 LINDSEI-BR corpus

LINDSEI-BR⁶ is a Brazilian sub-corpus of COBAI - *Corpus Oral Brasileiro de Aprendizizes de Inglês*, which is a branch of a larger project named LINDSEI⁷ (*Louvain International and Database of Spoken English Interlanguage*; Gilquin et al, 2010).

The LINDSEI project was launched in 1995, five years after the release of the International Corpus of Learner English (ICLE)⁸, and the goal was to offer a spoken counterpart to ICLE. According to Gilquin et al. (2010), the LINDSEI corpus brings together oral data produced by advanced learners of English from different mother tongues. According to the LINDSEI website, in the partners section, to date, nineteen components have been completed (Basque, Brazilian Portuguese, Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Greek, Italian, Japanese, Lithuanian, Norwegian, Polish, Spanish, Swedish, Taiwanese, Turkish), and five are under progress (Arabic (Lebanon), Arabic (Saudi Arabia), Croatian, Estonian, Iranian).

⁶ Available at: http://c-oral-brasil.org/cobai_lindsei_br.html

⁷ Available at: <https://uclouvain.be/en/research-institutes/ilc/cecl/lindsei.html>

⁸ Available at: <https://uclouvain.be/en/research-institutes/ilc/cecl/icle.html>

3.1.2 SBCSAE corpus

Published by the Linguistic Consortium (LDC)⁹, SBCSAE is also available on the TalkBank and Department of Linguistics of USCB¹⁰ websites. In addition, SBCSAE is part of the International Corpus of English (ICE) and is the main source of data for the spontaneously spoken portions of the American component of ICE.

According to Du Bois et al. (2000-2005), SBCSAE was compiled to represent a wide variety of people from different ages, genders, financial, geographical, and professional backgrounds in the United States of America. This corpus contains four parts and was compiled by researchers from the Linguistics Department of the University of California, Santa Barbara, under the guidance of John W. Du Bois.

Du Bois et al. (2000-2005), also state that the SBCSAE contains approximately 249 thousand words, split into four parts that include transcriptions, audio, and timestamps which correlate transcription and audio at the level of individual intonation units, as can be seen in the figure below. Most interactions are face-to-face conversations, however, other ways of interactions are available, for instance, card games, town hall meetings, telephone conversations, classroom lectures, food preparation, on-the-job talk, sermons, story-telling, tour-guide spiels, and more.

⁹ Linguistic Data Consortium: <https://www ldc. upenn. edu/>.

¹⁰ The Santa Barbara Corpus: <http://www. linguistics. ucsb. edu/research/santa-barbara-corpus>.

Figure 3.1.2: SBC001 Transcription screenshot

```

SBC001 - Bloco de Notas
Arquivo  Editar  Formatar  Exibir  Ajuda
0.00 9.21      LENORE:      ... So you don't need to go ... borrow equipment from anybody,
9.21 9.52      to --
9.52 14.10     ... to do the feet?
14.10 15.78     ... [Do the hooves]?
15.01 16.78     LYNNE:      [(H)=] <YWN Well,
16.78 18.32     we're gonna have to find somewhere,
18.33 18.85     to get,
18.85 20.69     (Hx) ... something (Hx) YWN>.
20.69 21.19     DORIS:      .. So,
21.19 21.74     [~Mae-] --
21.26 22.24     LYNNE:      [I'm gonna] (Hx) --
22.24 23.23     DORIS:      [2~Mae ~Lynne XX2]
22.28 24.25     LYNNE:      [2(H) We're not2] gonna do the feet today,
24.25 25.07     I'm gonna wait till like,
25.07 26.14     early in the morning=,
26.14 26.62     .. to do those,
26.62 26.87     cause y- --
26.87 28.37     I mean you get s= o ti=red.
28.37 30.10     (H) ... n- you just,
30.10 32.36     ... it takes % --
32.36 32.59     well,
32.59 33.90     it takes me longer than most people,
33.90 34.22     cause you know,
34.22 35.23     I'm not as stro=ng and,
35.23 36.88     (H) ... and I'm not as good,
36.88 39.30     as like somebody that would do it .. all the ti=me.
39.30 39.80     .. You know.
39.80 40.27     .. I mean,
40.27 41.39     .. (H) % <X Oh X> --
41.39 42.84     I mean I trim horses=,
42.84 43.66     and stuff like that,
43.66 44.20     but I mean,
44.20 44.92     I'm not like,
44.92 47.60     (H) .. <@ I'm no=t uh= @>,
Ln 1, Col 1      100%      Unix (LF)      1 nova notificação

```

Source: Designed by the author, 2024.

3.2 Notepad++

As mentioned in section 3.1, and as shown in Figures 3.1.1 and 3.1.2, LINDSEI-BR and SBSCAE transcriptions contain markers for many speech features. As this study focuses on analyzing SVs purely in a linguistic context, such markers are irrelevant to this study as they do not deal with the investigation of prosody. Thus, it was necessary to preprocess the corpora to erase nonrelevant characters, and the tool adopted was Notepad++¹¹. This software can be downloaded at no cost from its website, and it can be used as a replacement for Windows Notepad, as well as a source code editor.

3.2.1 Preprocessing of LINDSEI-BR

After loading all fifty TXT files of LINDSEI-BR to Notepad++, through the command *Ctrl + h*, regular expressions (Regex) were used to find what needed to be erased from the files. It is necessary to mention that, when using Regex, the box for

¹¹ Notepad++: <https://notepad-plus-plus.org/>

such a query needs to be selected, otherwise it will not find what is being searched for. For all the queries, the box to differentiate lower case letters from upper case letters was always selected. This was necessary because some of the markers used in the transcription of the texts included upper cases. The following table shows what Regex and sequence they were applied in the preprocessing of LINDSEI-BR.

Table 3.2.1: Regex used in the preprocessing of LINDSEI-BR

Regex	Used for	Replaced by
<code><(?!V?[AaBb]\s*+)[^>]+></code>	erasing anything in between < >, except <A>, , , .	nothing.
<code>\[.*?\]</code>	erasing anything in between [].	nothing.
<code>\(.*?\)</code>	erasing anything in between ().	nothing.
<code>\s+</code>	Erasing double spaces and empty lines.	one space in between the characters.

Source: Elaborated by the author, 2024.

The first Regex on the table above was used to erase markers related to the text's metadata and reactions, such as laughing and overlapping. The excerpt below from BR001 shows some of these markers in bold.

<h nt="BR" nr="BR001">

<S>

<A> hi **<first name of interviewee>** . I like to interview you informally on things of interest in your life for about fifteen minutes. (erm) .. **<click>** you should be **<X>** to be able to talk for three or five minutes about . one of the three topics. which one that you: ... **<overlap />** which .. topic number two **<overlap />** okay ****

The markers <A>, , , , indicate when a turn in the conversation begins and ends, with A belonging to the interviewer, and B to the interviewee. These mark-ups were not excluded to help visualize which speaker was responsible for the utterance.

The second Regex erased information related to the speaker's pauses, as can be seen in the excerpt from BR002 that follows.

<A> <overlap /> never heard<?> ... **[?pause?_04:05-04:20]** so . do you think that: . your: graduation . will help you: to: . improve: your English your skills . and: you want to be: . a professor . a teacher . English teacher

Some other reactions manifested by the speakers related to taking time or expressing interest and understanding were marked as well, however, such reactions were marked in between parenthesis, thus, the third Regex was used to remove them, and they appear in bold in the excerpt from BR003.

 ok ... ok . **(erm)** ... there is one girl who is being painted . by an artist . and: he= he pictures her . **(em)** I= exactly the way she is . and then when she sees the picture she criticises it she says I'm not like this . and: this is wrong .. and then he repaints her really beautiful . her hair is beautiful **(eh)** her lips her nose everything so perfect and then she's show= she shows the picture to her friends saying see that is me look how beautiful I am <laughs>

After erasing all the markers mentioned above, the last Regex on Table 3.2.1 was used to fix all the double spaces and remove empty lines. The result of a preprocessed transcription is seen below.

Table 3.2.2: Regex used in the preprocessing of SBCSAE

Regex	Used for	Replaced by
<code>\d+\.\d+</code>	erasing the numbers in the time stamps and the “.” in between them.	nothing.
<code>\d+</code>	erasing all the left numbers	nothing.
<code>\(. *?\)</code>	erasing anything in between ().	nothing.
<code>\s+</code>	Erasing double spaces and empty lines.	one space in between the characters.

Source: Elaborated by the author, 2024.

The numbers in the time stamps in the transcripts contain a “.” between them; however, more numbers are used in the transcripts in SBCSAE for different purposes. Thus, it was necessary to use different Regex to erase all the numbers from the files. The first Regex erased the numbers belonging to the timestamps, as shown in the excerpt below from SBC001.

0.00 9.21 LENORE: ... So you don't need to go ... borrow equipment
from anybody,

9.21 9.52 to --

9.52 14.10 ... to do the feet?

The other numbers were erased using the second Regex in Table 3.2.2. These numbers were used in the transcript and have no connection to time. They can be seen in the following excerpt from SBC002.

15.01 16.43 JAMIE: ... Tap?
 16.43 16.98 [X] [2X2] --
 16.50 17.00 HAROLD: [What] [2was the2],
 16.60 17.00 MILES: [2<X They had X>2] --

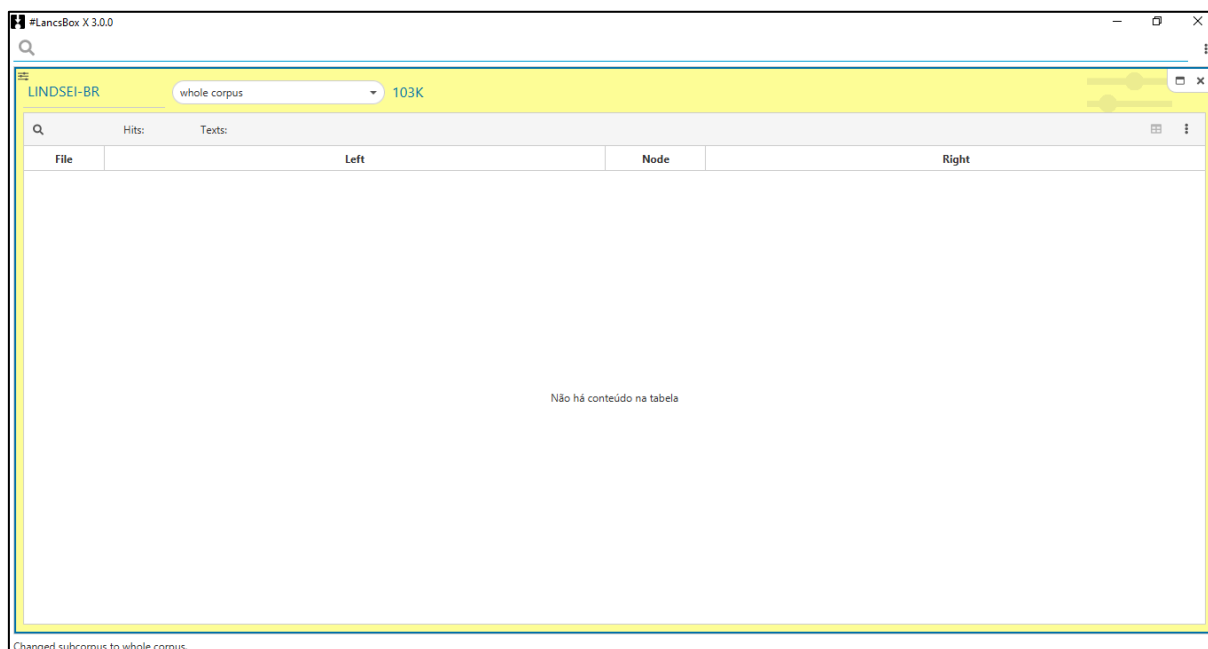
Some SBCSAE information was transcribed between parentheses, primarily to indicate noise. Such information is irrelevant to this study as it does not interfere with the usage of SVs. For this reason, the third Regex in Table 3.2.2 was used to erase markers and words indicating noise. The excerpt from SBC003 shows some of the items erased from the files.

429.35 429.85 MARILYN: [(THROAT)]
 429.35 429.95 PETE: [Or what].
 429.95 430.40 MARILYN: Yeah.
 430.40 431.10 PETE: .. Unh[hunh].
 430.65 431.68 MARILYN: [(H) And some] guy who- --
 431.68 433.53 .. who knows a good thing when he's on to it,
 433.53 434.83 .. (H) calls them up.

After cleaning the corpus from the markers mentioned above, we were left with some characters and words that were one by one removed using *Ctrl + h*. The characters and words are the following ones: [,], @, <, >, ~, \$, %, &, *, --, ?, !, ., .., ..., F, F, Q, Q,, Q., P, P., P?, VOX, X, DOG:, ENV:, SMOKING, DRINKING, SM, WHISTLE, HI, HI., HI,, SLAPPING, SING, FOODWH, YWN, TALK, FOOTSTEPS, PARP, PPAR, SNAP.

The last Regex on Table 3.2.2 was then used to erase all the double spaces and empty lines. The names of the participants in the conversations in SBCSAE were not erased to assist in interpreting the queries in this study. Figure 3.2.2 shows a preprocessed file from SBCSAE.

Figure 3.3: LancsBox X main screen overview



Source: Elaborated by the author, 2024.

One of the reasons that influenced the choice of this tool in this study is the Semantic Tagger available on it. As we are dealing with the senses of usage of SVs in a semantical sphere, using a tool that contributes to the ease of the analysis is both of paramount importance and time-saving. Therefore, both corpora were added to LancsBox X to be queried.

The figure below shows the screen of LancsBox X when adding a corpus to it. As observed, the options for Grammatical and Semantic tagging are selected.

Figure 3.3.1: LancsBox X Add Corpora screen

Source: Elaborated by the author, 2024.

LancsBox X provides its users with the following tools for corpus analyses: the Key Word in Context (KWIC) tool, GraphColl, Words Tool, and Text Tool. The first and second tools were used in this study; the last two were not necessary. However, the following lines provide information about the functionality of each one.

3.3.1 KWIC Tool (Key Word in Context)

With this tool, it is possible to generate a comprehensive list of all the occurrences of a particular term in the corpus. In addition, it can be used to:

- Determine the frequency of words or phrases.
- Identify the frequency of different word classes.
- Locate specific linguistic structures such as passive voice and split infinitives.
- Organize concordance lines.
- Conduct side-by-side comparisons of multiple analyses.

Figure 3.3.1.1: LancsBox X KWIC screen

The screenshot shows the LancsBox X 3.0.0 interface. At the top, there is a search bar with the query 'love' and a dropdown menu set to 'whole corpus' with 103K results. Below this, a summary bar indicates 'Hits: 171 (1.665.66)' and 'Texts: 40/50'. The main area is a table with four columns: 'File', 'Left', 'Node', and 'Right'. The 'Node' column consistently contains the word 'love'. The 'Left' and 'Right' columns show the surrounding text from various source files, such as 'BR0042.txt', 'BR0031.txt', etc. The bottom of the window shows a status bar with 'Searched KWIC for "love"' and '2 novas notificações'.

File	Left	Node	Right
BR0042.txt	is greener and well yeah basically is that and I	love	that country I I hope to go back someday maybe
BR0031.txt	beautiful woman h= he painted a beautiful woman and: she	love	it and she said oh now now you're talking
BR0001.txt	family they were the oh my gosh I'm in	love	with them and we are still talking
BR0019.txt	B> he manages the life with of very very	love	of his fa= son and he had a problem with
BR0019.txt	I I like so much to dance yes I I	love	the the club it's a place that only more
BR0047.txt	's most it's easier than the German but I	love	most the: the German one German language I had I
BR0047.txt	the: the German one German language I had I I	love	so much the the German literature especially the Thomas Mann
BR0047.txt	yes <A> yeah yeah I	love ...	wow very inspired man yeah nice <
BR0047.txt	specially the German people seduce me for one side I	love	so much the: what ha= had been produced since for
BR0003.txt	nice <A> it's great I	love	interpretation and translation I just love it
BR0003.txt	> it's great I love interpretation and translation I just	love	it <A> so I will
BR0042.txt	R Tolkien is my favourite I mean I I I	love	yeah I love lord of the rings I've well
BR0042.txt	my favourite I mean I I I love yeah I	love	lord of the rings I've well the first time
BR0042.txt	was ten: to twelve something between between these and I	love	it I mean I I'm I'm a sort
BR0042.txt	yeah comic books and all those things so I still	love	those things I still collect comic books and that that
BR0042.txt	n't be so so good you know and well I	love	the the the old school as well you know
BR0042.txt	do a free class I I do role plays I	love	doing you know the students love it as well I
BR0042.txt	do role plays I love doing you know the students	love	it as well I think it's a good very

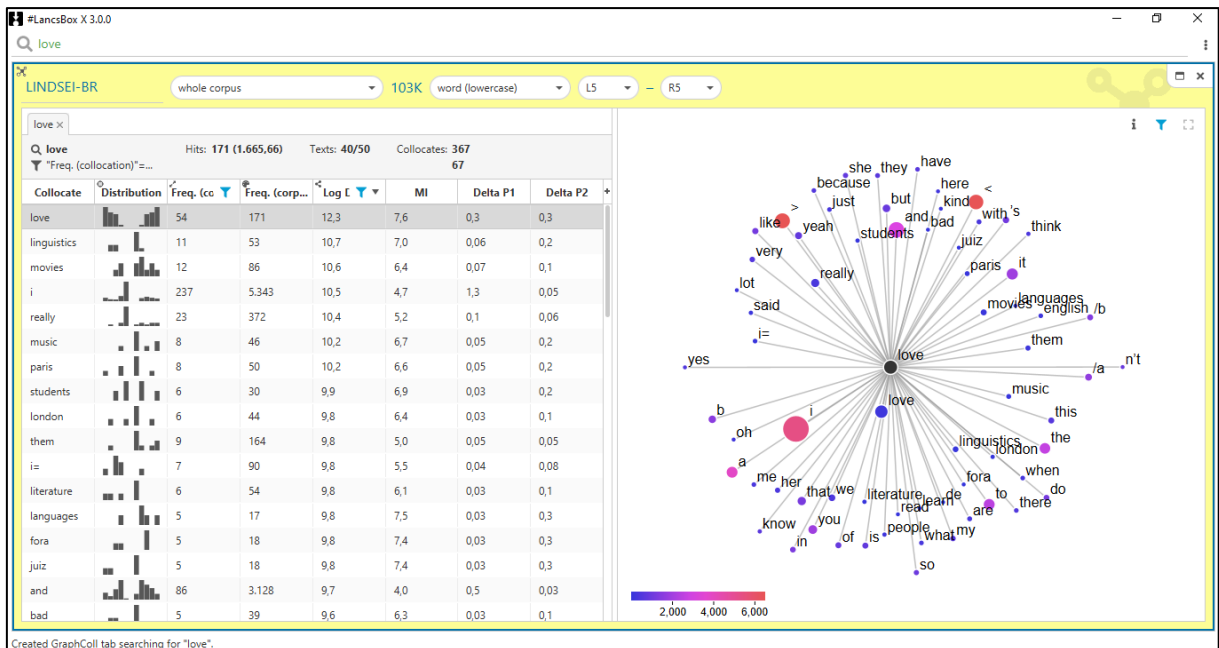
Source: elaborated by the author, 2024

3.3.2 GraphColl

GraphColl is a versatile tool that enables users to identify collocations within a given corpus, presenting the results in graphical and tabular formats. It can be employed for a variety of purposes, such as:

- detecting collocates for a particular word or phrase,
- uncovering colligations (co-occurrence of grammatical categories),
- visualizing collocations and colligations and identifying shared common collocates between words or phrases.
- Summarizing the central themes within discourse.

Figure 3.3.2: LancsBox X GraphColl screen



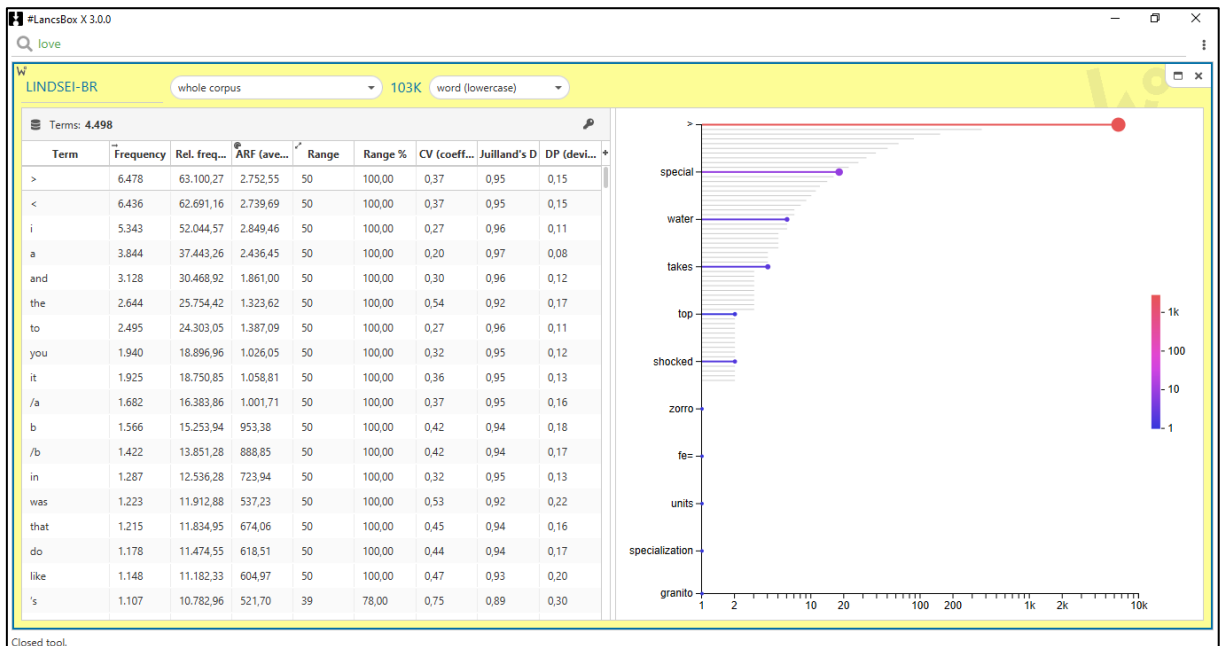
Source: elaborated by the author, 2024

3.3.3 Words Tool

This tool is designed to analyze word frequencies and semantic and grammatical categories in-depth. With this tool, it is also possible to compare different corpora using the keyword technique. Some of the things that can be done with it include:

- Computing dispersion metrics and frequency.
- Creating visual representations of frequency and dispersion.
- Comparing different corpora using the keyword method.

Figure 3.3.3: LancsBox X Words Tool screen



Source: elaborated by the author, 2024.

3.3.4 Text Tool

The Text tool is an incredibly versatile feature that offers a comprehensive overview of all files within a corpus, including their size and lexical diversity. With its full-view mode, it provides a detailed analysis of individual texts. Additionally, the Text tool allows for text searching and generates a summary table of frequencies and relative frequencies for each file. It also offers the ability to highlight search terms within individual texts. The Text tool is helpful for a variety of purposes, including:

- Exploring corpora and their texts prior to analysis.
- Gaining insights into the distribution of corpus files in terms of size, lexical diversity, and linguistic feature frequencies.
- Conducting qualitative text analysis.

Figure 3.4.4: LancsBox X Text Tool screen



Source: elaborated by the author, 2024.

3.3.5 Searching in LancsBox X

The searches in LancsBox X can be done at distinct levels of corpus annotation using:

- i) simple searches,
- ii) wildcard searches,
- iii) smart searches,
- iv) Corpus Query Language (CQL) searches.

In this study, the level adopted was the CQL because different forms of the same verb needed to be analyzed. Thus, it was more convenient to change just the POS tag and then type all the different forms of the verb investigated. Although useful, the Word and Text tools were not employed in this study as the former focuses more on all the words in a corpus and the latter on information about each text separately.

3.4 Querying LINDSEI-BR and SBCSAE

Unfortunately, it was not possible to find any study that has dealt with the frequency of SVs usage in spoken communication. Therefore, it was necessary to create a list containing five verbs within each semantic category of SVs, as explained in Chapter 2, section 2.3.1. Based on the author's experience as a learner and as an English teacher, the verbs listed below were chosen to be investigated and grouped into the following semantic categories:

- i) Verbs expressing mental states, thoughts or opinions: believe, know, prefer, think, understand.
- ii) Verbs expressing perception or senses: hear, listen, notice, see, watch.
- iii) Verbs expressing physical states: appear, associate, exist, look, seem.
- iv) Verbs expressing possession: belong, have, hold, keep, own.
- v) Verbs expressing relationship and emotions: dislike, hate, like, love, trust.
- vi) Verb "Be".

Each of these verbs was queried to obtain information to describe their usage considering frequency, form, and main collocates. As the verb "be" falls into many different subcategories of SVs, it was analyzed separately, considering its richness of usage. Table 3.3.6 shows the POS tags used to query each form of the verbs selected and mentioned above.

Table 3.3.6: POS tags

POS Tag	Query for
"VB"	Base form
"VBD"	Past form
"VBG"	Gerund or Present Participle forms
"VBN"	Past Participle Form
"VBP"	Non-third-person singular Present forms
"VBZ"	3 rd person singular Present forms

Source: elaborated by the author, 2024.

Then, each of the POS tags above was applied in the following CQL.

[hw= "verb" pos="tag"]

For example, to query for all the occurrences of the verb Be, the CQLs below were used. To query for all the other verbs, after "hw=", the word "be" was replaced by the new verb, and the "pos=" was followed by the required form.

- [hw="be" pos= "VB"] ► for base form.
- [hw="be" pos= "VBD"] ► for past form.
- [hw="be" pos= "VBG"] ► for gerund or present participle.
- [hw="be" pos= "VBN"] ► for past participle.
- [hw="be" pos= "VBP"] ► for non-3rd person singular present.
- [hw="be" pos= "VBZ"] ► for 3rd person singular present.

The context size selected was ten, and after the results were shown, the semantic filter was enabled to check which semantic tag was associated with the verbs, thus helping in the analysis. All the occurrences were saved to the computer in TXT format and then transferred to a spreadsheet. Figures 3.3.6 and 3.3.6.1 below show the results of the query for [hw="be" pos= "VB"] in LINDSEI-BR with and without the semantic tag details.

Figure 3.3.6: Verb “be” VB query.

File	Left	Node	Right
BR0020.txt	changing talking using Messenger Skype or other stuff that would	be	great if I could have if I I could
BR0020.txt	great if I could have if I I I could	be	in contact with them but I'm sorry that I
BR0020.txt	in this case in this kind of trip it would	be	different but special anyway
BR0020.txt	tell me about your future plans you're studying to	be	an English teacher or you're in a bachelor <
BR0020.txt	and and or with as an English teacher it would	be	great I I think it would be very good but
BR0020.txt	teacher it would be great I I think it would	be	very good but I'm still worried about to complish=
BR003.txt	this English notion and: my English didn't have to	be	as good as my Portuguese to translate so: it= it
BR003.txt	really want to know England so: I think it will	be	another experience another different experience because: it's an European
BR003.txt	to <A> you want to	be	there as an exchange student or <
BR003.txt	year actually and: I'm= my post graduation is gonna	be	in Rio de Janeiro so: afterwards make my master <
BR003.txt	be= that's a dream actually working like Government conferences would	be	the top I'll get there someday but I still have
BR0030.txt	really and I don't know if I want to	be	a teacher for the rest of my life but you
BR0030.txt	in the next years I know I'm going to	be	a teacher and I= I want to be a really
BR0030.txt	going to be a teacher and I= I want to	be	a really good teacher <A>
BR0030.txt	I don't know if it= this is going to	be	possible so I'll try try both <
BR0030.txt	things to do so now for now I want to	be	an engineer but maybe I can change plans I do
BR0030.txt	so I don't know if this is going to	be	possible <A> come on there
BR0030.txt	I'm twenty two already I had to to to	be	started so I need to hurry <

Source: elaborated by the author, 2024.

As seen in Figure 3.3.6.1, the semantic tag associated with the verb Be is A3+, indicating the state of being.

Figure 3.3.6.1: Verb “be” VB query with semantic tag

File	Left	Node	Right
BR0031.txt	A_Z5>_Z99 ... so_Z5 as_Z5 a_Z5: a_Z5 plan_X7+ you_Z8mf you_Z8mfIl_Z99	be_A3+	a_Z5 translator_Q3/i3.2/S2mf yeah_Z4 it_Z8's_A3+ it_Z8's_A3+ a_Z5 good_A5.1+ plan_X7+
BR0031.txt	I_Z8mf can_A7+ see_X3.4 in_Z5 in_Z5 some_N5 sites_M7 some_N5 like_E2+ would_A7+	be_A3+	like_Z5 original_T3- not_Z6 just_A14 download_M2 yeah_Z4 oh_Z4 you_Z8mf can_A7+ watch_X3.4
BR0031.txt	a_Z5 way_X4.2 to_Z5 talk_Q2.1 is_A3+ I_Z8mf you_Z8mf you_Z8mf want_X7+ to_Z5	be_A3+	him_Z8m as_Z5 you_Z4 see_Z4 he_Z8m's_A3+ too_N5++ great_A5.1+ he_Z8m's_A3+
BR0033.txt	/A_PUNCT>_Z99 <_Z99B_Z5>_Z99 yes_Z4 it_Z8 seems_A8 to_Z5 to_Z5	be_A3+	there_Z5 is_A3+ a_Z5 woman_S2.1f and_Z5 a_Z5 painter_C1/S2mf and_Z5 he_Z8m paints_A5.4+
BR0033.txt	very_A13.3 much_A13.3 like_Z5 her_Z8f and_Z5 she_Z8f does_A1.1.1n't_Z99 seem_A8 to_Z5	be_A3+	a_Z5 very_A13.3 beautiful_Q4.2+ woman_S2.1f so_Z5 she_Z8f struggles_X8+ at_Z5 him_Z8m and_Z5
BR0033.txt	painting_C1 to_Z5 all_N5.1+ her_Z8f her_Z8f girlfriends_S3.2/S2.1f yeah_Z4 she_Z8f seems_A8 to_Z5	be_A3+	very_A13.3 proud_E4.2+/S1.2 of_Z5 it_Z8 even_A13.1 though_Z5 she_Z8f knows_A5.4+ she_A5.4+ does.
BR0015.txt	4mf concerns_E6- for_Z5 granted_A9+ ever_T1.1 I_Z8mf should_S6+ always_N6+++: you_Z4 know_Z4	be_A3+	let_A5.4+ them_A5.4+ be_A3+ aware_X2.2+ of_Z5 where_M6 I_Z8mf am_A3+ and_Z5 you_Z4
BR0015.txt	_d_A9+ ever_T1.1 I_Z8mf should_S6+ always_N6+++: you_Z4 know_Z4 be_A3+ let_A5.4+ them_A5.4+	be_A3+	aware_X2.2+ of_Z5 where_M6 I_Z8mf am_A3+ and_Z5 you_Z4 know_Z4 that_Z5 stuff_O1
BR0015.txt	did_A1.1.1n't_Z6 really_Z6 they_Z8mf never_T1/Z6 really_A13.3 asked_Q2.2 for_Z5 me_Z8mf to_Z5	be_A3+	like_Z5 oh_Z4 I_Z8mf am_A3+ going_M1 there_M6 with_Z5 someone_Z8mf and_Z5: this_M6
BR0015.txt	>_Z99 right_S7.4+ <_Z99/A_PUNCT>_Z99 if_Z7 I_Z8mf: ... disappear_A10- they_Z8mf will_T1.1.3	be_A3+	concerned_E6- <_Z99/A_PUNCT>_Z99 <_Z99B_Z5>_Z99 right_S7.4+ <_Z99/A_PUNCT
BR0015.txt	learnable_Z99 so_Z5: I_Z8mf=Z51_Z8mf learned_X2.3+ how_Z5 to_Z5 walk_M1 and_Z5:	be_A3+	careful_A1.3+ and_Z5 <_Z99B_PUNCT>_Z99 <_Z99A_Z5>_Z99 You_Z8mf can_A7+
BR0031.txt	day_T1.1.3 and_Z5: he_Z8m was_A3+ too_N5++ good_A5.1+ you_Z8mf you_Z8mf ca_A7+n't_Z99	be_A3+	that_A13.3 good_A5.1+ in_Z5 one_T1.1.3 day_T1.1.3 but_Z5 I_Z8mf do_S1.1.3+n't_Z99 know_A5.4+
BR0031.txt	/B_PUNCT>_Z99 <_Z99A_Z5>_Z99 ... a_Z5 collaboration_S1.1.3+ ... you_Z8mf want_X7+ to_Z5	be_A3+	him_Z8m ... and_Z5 you_Z8mf believe_X2.1 yeah_Z4 the_Z5 true_A5.2+ ... a_Z5 great_A5.1+ fan_E2+/S2
BR0031.txt	say_Q2.1 no_Z4 I_Z8mf do_S1.1.3+n't_Z99 know_X2.2+ if_Z7 we_Z8 want_X7+ to_Z5	be_A3+	friends_S3.1/S2mf of_Z5 her_Z8f anymore_Z5 I_Z8mf ca_A7+n't_Z99 see_A5.4+ this_A5.4+ and_Z5:
BR0019.txt	future_T1.1.3 plans_X7+ what_Z8 do_S1.1.3+ you_Z8mf think_A5.4+ it_A5.4+'s_A3+ gon_T1.1.3na_Z5	be_A3+	a_Z5 happy_E4.1+ life_L1+ to_Z5 you_Z8mf after_Z5 graduation_P1 <_Z99/A_PUNCT>_Z99
BR001.txt	rest_X5.2+ in_Z5 your_Z8 life_L1+ for_Z5 about_A13.4 fifteen_N1 minutes_T1.3 you_Z8mf should_S6+	be_A3+	to_Z5 be_A3+ able_X9.1+ to_Z5 talk_Q2.1 for_Z5 three_N1 or_Z5 five_N1 minutes_T1.3
BR001.txt	your_Z8 life_L1+ for_Z5 about_A13.4 fifteen_N1 minutes_T1.3 you_Z8mf should_S6+ be_A3+ to_Z5	be_A3+	able_X9.1+ to_Z5 talk_Q2.1 for_Z5 three_N1 or_Z5 five_N1 minutes_T1.3 about_Z5 one_N1
BR001.txt	s_Z5 mother_S4f yet_T1.1.2 but_Z5 someday_T1.1.3 I'll=Z99 I_Z8mf hopefully_X2.6+ I_Z8mf will_T1.1.3	be_A3+	so_Z5 things_O2 that_A13.3 I_Z8mf should_S6+ do_A1.1.1 and_Z5 things_O2 that_A13.3 I_Z8mf

Source: elaborated by the author, 2024.

The exact process was done for the collocate; Regex querying for the base form of the SVs occurring in both corpora was applied to find the collocates calculated by the Log Dice, and the five strongest collocates (if available) were then analyzed.

This chapter presented the two corpora utilized in this thesis and the software employed to query them. Initially, it described how LINDSEI-BR and SBCSAE were compiled and their availability. Following this, it detailed the preprocessing of the texts in the corpora and how they were queried. The following chapter is dedicated to presenting the results obtained and discussing them.

CHAPTER 4

RESULTS AND DISCUSSION

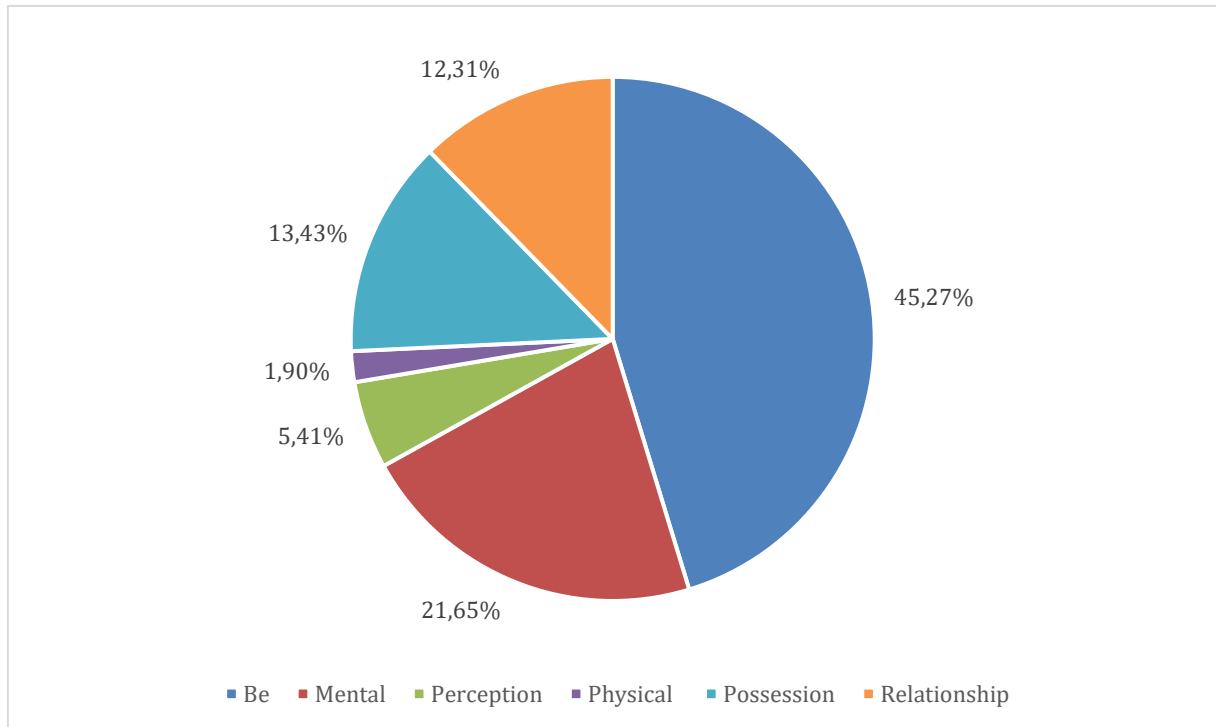
This chapter presents the total number of occurrences of the verb “be” and each SV within the investigated semantic categories (mental, perception, physical, possession, and relationship)¹³. The second part of this chapter deals with the semantic contexts in which these verbs were used according to the semantic tags attributed to the SVs with the help of LancsBox X. Following, this chapter presents the collocates to the common SVs’ base forms in both corpora. The last part of the chapter brings some final remarks about the results and compares them to the information found in the literature review of this thesis.

4.1 LINDSEI-BR SVs occurrences

Among the SVs selected to be studied in this thesis, there were 6,096 occurrences. Not surprisingly, as shown in Graph 4.1, almost 50% of such occurrences belong to the verb “be,” as this verb has a significant role in the English language for being used as an auxiliary and main verb. Related to the semantic categories adopted in this thesis, mental verbs were responsible for 21,65% of the occurrences, followed by possession and relationship verbs. These two categories had remarkably close occurrences. The category of verbs with the two lowest usages were perception and physical, 5,41% and 1,90%, respectively. Therefore, we can assume that those categories may also present the lowest frequency of usage of SVs in the progressive form. Let us analyze each category in the following sections and determine whether such a hypothesis is confirmed.

¹³ The data analyzed are available in Appendices A and B.

Graph 4.1: LINDSEI-BR SVs occurrences



Source: elaborated by the author, 2024

4.1.1 LINDSEI-BR verb "be" occurrences

Table 4.1.1 shows that the verb "be" was observed in 277 instances of the base form (VB), 1,403 instances of past form (VBD), 17 instances of gerund or present participle forms (VBG), 51 instances of past participle form (VBN), 940 instances of non-3rd person singular present forms (VBP), and 72 instances of third person singular present forms (VBZ). As mentioned before, this was unsurprising, given that "be" is among the most frequently used verbs in the English language and appears in all its forms. Interestingly, this verb did not present many occurrences of the progressive form (VBG), which might be due to the nature of the corpus, which is semi-spontaneous speech. Therefore, the participants may not have had enough opportunity to explore using such forms. Curiously, the participants in this corpus showed more preference for using this verb in the past form (VBD), which is still the highest even when we put all the occurrences of the present tense together.

Table 4.1.1: LINDSEI-BR verb “be” occurrences.

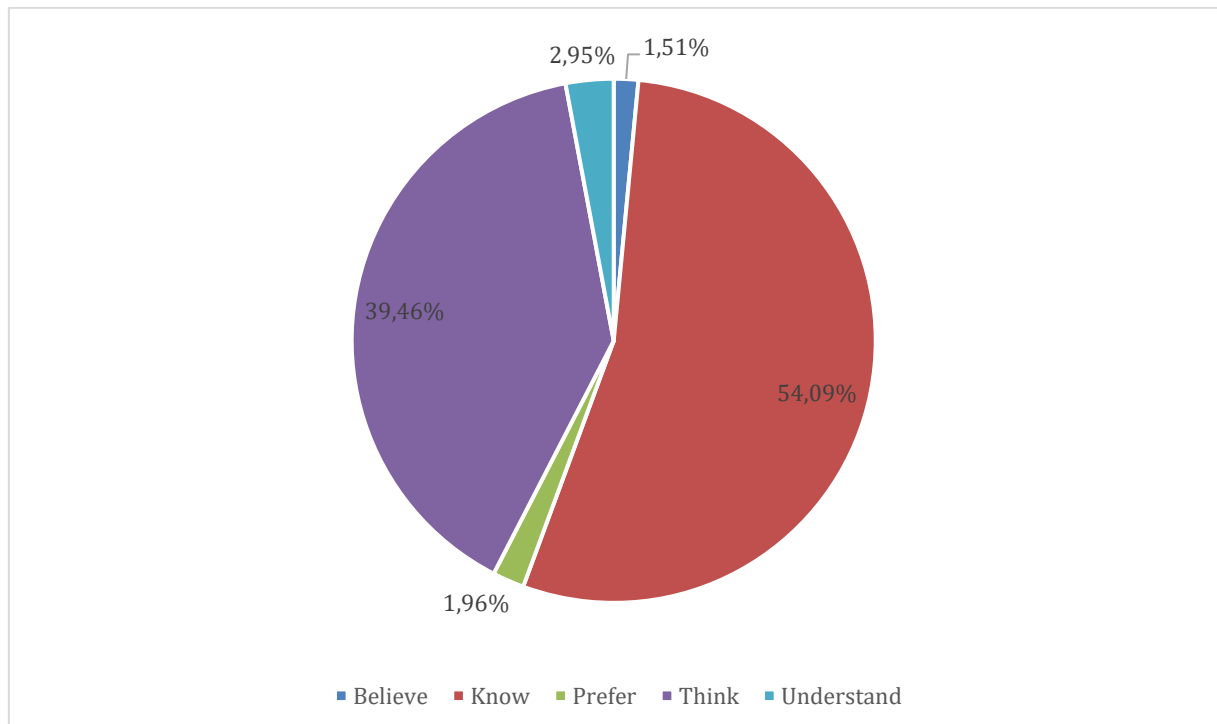
BE	
VB	277
VBD	1,403
VBG	17
VBN	51
VBP	940
VBZ	72

Source: elaborated by the author, 2024

4.1.2 LINDSEI-BR mental verb occurrences

Considering all the occurrences of the five mental verbs selected for this thesis (believe, know, prefer, think, and understand), 1,320 instances were analyzed. As shown in Graph 4.1.2, more than 50% of these occurrences belong to the verb know, followed by the verb believe. The other three verbs left had extremely low occurrences.

Graph 4.1.2: LINDSEI-BR mental verb occurrences



Source: elaborated by author, 2024

As shown in Table 4.1.2, the two verbs with the highest occurrences (know and think) are the verbs that also have stances of all forms found. They are also the verbs that had more progressive (VBG) occurrences. The verbs believe, think, and understand are mental verbs with low frequency in general and progressive usage. It may suggest that SVs with low-frequency usage in general in English might be the same verbs with low frequency in the progressive form. Thus, general usage frequency may impact whether an SV will be used in the progressive form.

Table 4.1.2: LINDSEI-BR mental verb tense occurrences

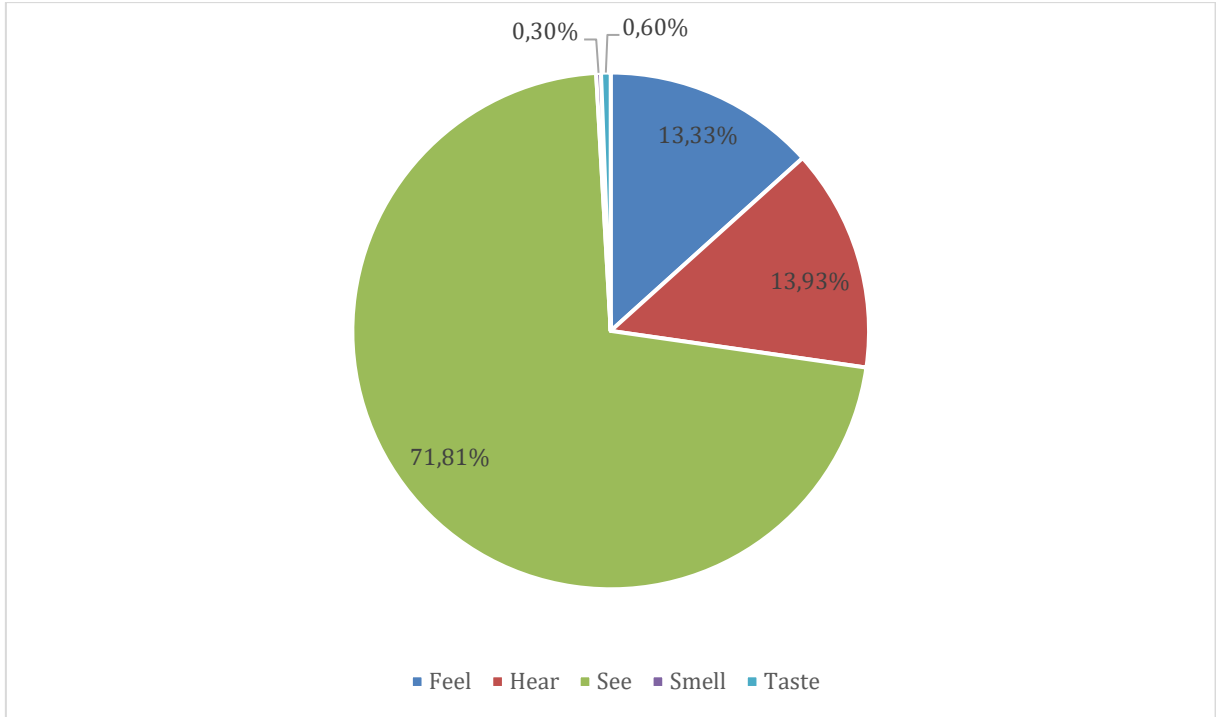
	Believe	Know	Prefer	Think	Understand
VB	9	403	3	77	27
VBD	2	17	0	55	3
VBG	0	4	0	29	1
VCN	1	2	0	2	0
VBP	6	278	21	350	1
VBZ	2	10	2	8	0

Source: elaborated by the author, 2024.

4.1.3 LINDSEI-BR perception verb occurrences

In the group of perception verbs (feel, hear, see, smell, taste), 330 occurrences were found. As seen in Graph 4.1.3, out % of these 330 occurrences, 71,81% belong to the verb “see.” It might be because, in one of the tasks in LINDSEI-BR, the participants were asked to describe a sequence of four images, which may have influenced such a high frequency of usage. The verbs hear and feel presented similar occurrences, 13,93% and 13,33%, respectively. On the other hand, smell and taste presented low frequency, showing less than 1% of occurrences.

Graph 4.1.3: LINDSEI-BR perception verb occurrences



Source: elaborated by the author, 2024.

When we analyze the forms in which these verbs are used, as shown in Table 4.1.3, again, the verbs with the highest frequency of usage are also the verbs with the highest frequency of usage in the progressive (VBG). Once more, this may be due to the nature of the corpus's tasks; thus, the results found in the analysis of this group in the SBCSAE will help to reinforce this hypothesis.

Table 4.1.3: LINDSEI-BR perception verb forms occurrences

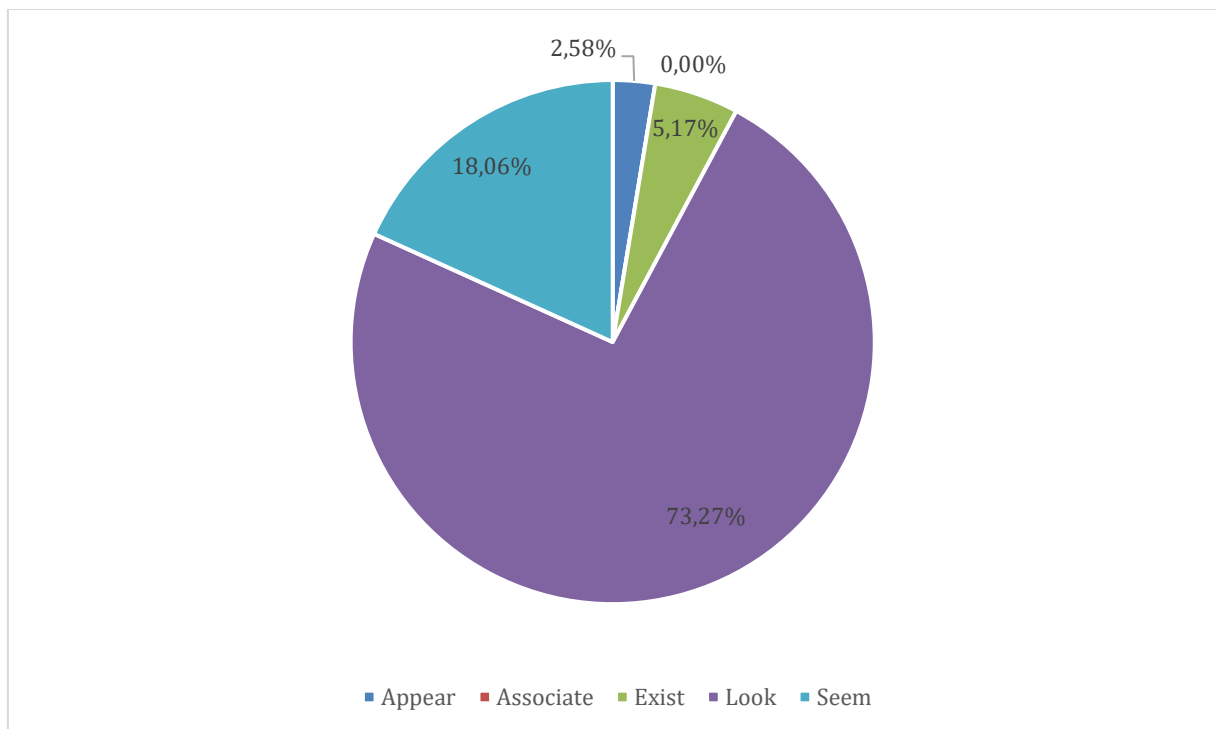
	Feel	Hear	See	Smell	Taste
VB	14	17	118	0	1
VBD	5	15	56	0	1
VBG	3	5	5	1	0
VCN	0	8	14	0	0
VBP	21	1	40	0	0
VBZ	1	0	4	0	0

Source: elaborated by the author, 2024

4.1.4 LINDSEI-BR physical verb occurrences

Only 116 occurrences of the physical verbs (appear, associate, exist, look, seem) selected here were found. So far, this is the group with the lowest occurrences, and like the group of perception verbs, one verb, here "look", was responsible for most of the occurrences found (73.27%), followed by "seem" (18.06%), "exist" (5.17%) and "appear" (2.58%). There were no occurrences of the verb "associate", as shown in the graph below.

Graph 4.1.4: LINDSEI-BR physical verb occurrences



Source: elaborated by the author, 2024.

Like the other groups analyzed so far, as shown in Table 4.1.4, the verbs with the lowest frequency in widespread usage are the verbs that do not have the occurrences spread through all the forms studied here. Again, the verb with the highest frequency (look) is the verb that presented more usage in the progressive. The verb "exist" is the only verb besides "look" that presented the progressive in its use. Besides these two verbs, no more use in the progressive form was found.

Table 4.1.4: LINDEI-BR physical verb forms occurrences

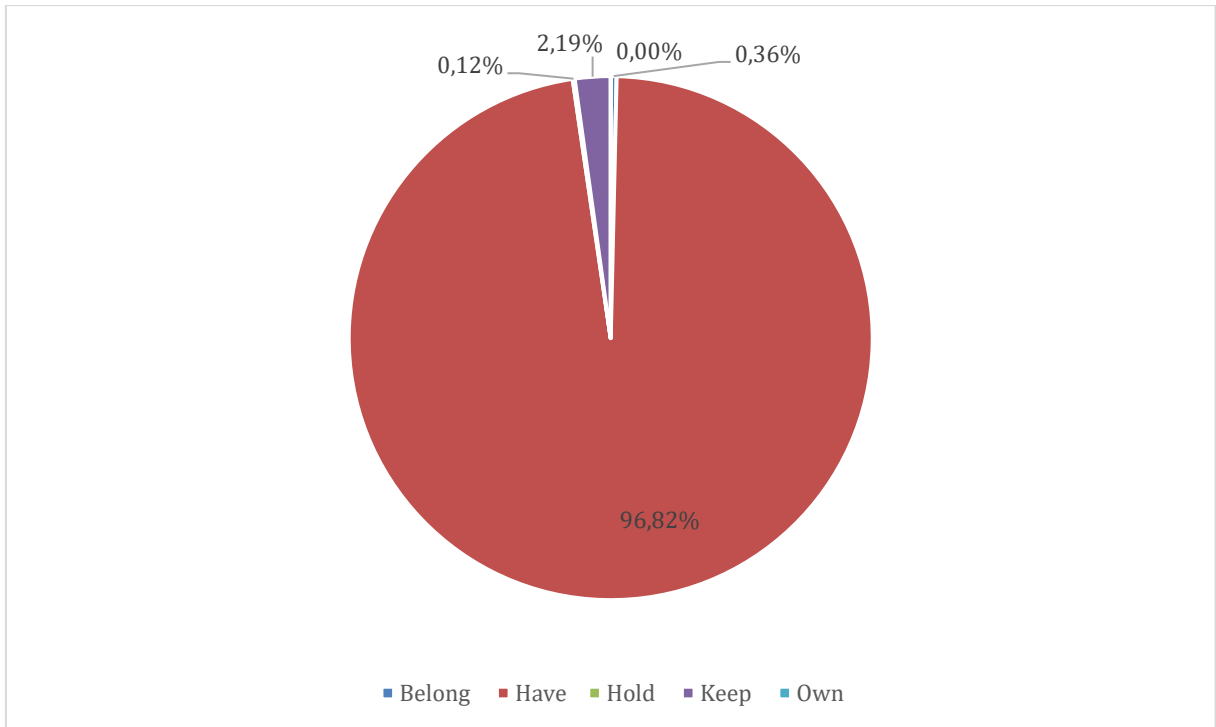
	Appear	Associate	Exist	Look	Seem
VB	1	0	4	24	4
VBD	0	0	0	15	1
VBG	0	0	1	21	0
VBN	0	0	0	1	0
VBP	1	0	0	11	1
VBZ	1	0	1	13	16

Source: elaborated by the author, 2024.

4.1.5 LINDSEI-BR possession verb occurrences

Possession verbs in the LINDSEI-BR corpus presented 819 occurrences, and as shown in Graph 4.1.5, almost one hundred percent of these occurrences belong to the verb “have” (96,82%). There were no occurrences for the verb “own,” and the occurrences of “belong” and “hold” represent less than one percent of the total. The second verb with the second-highest occurrences is “hold.” However, when compared to “have”, the frequency is still low.

Graph 4.1.5: LINDSEI-BR possession verb occurrences



Source: elaborated by the author, 2024

Table 4.1.5 shows that the verb “have” has occurrences spread in all forms analyzed. Except for “hold,” no other verb was used in the progressive form. This group would have had the lowest occurrences if the verb “have” did not belong to this semantic category. Later in this chapter, we will discuss the results found in SBCSAE and compare them to those presented here.

Table 4.1.5: LINDSEI-BR possession verb forms occurrences

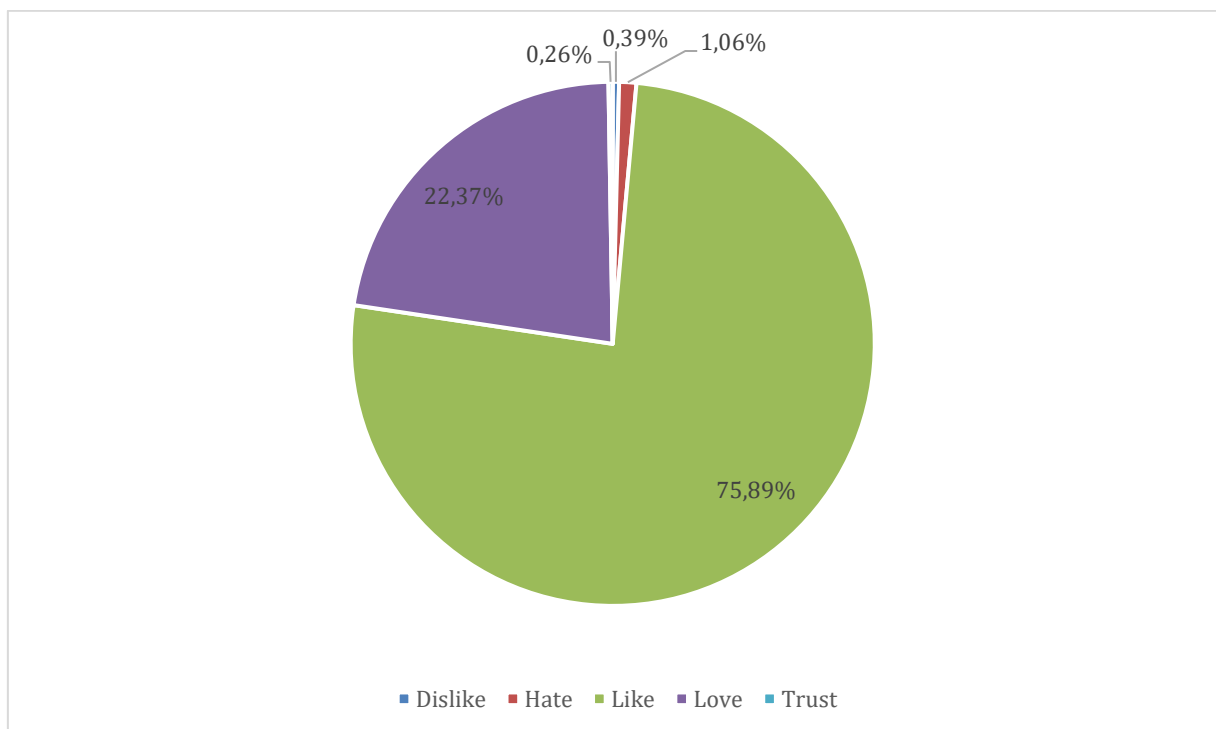
	Belong	Have	Hold	Keep	Own
VB	0	186	0	12	0
VBD	0	14	0	0	0
VBG	0	14	1	0	0
VBN	0	14	0	0	0
VBP	3	351	0	6	0
VBZ	0	60	0	0	0

Source: elaborated by the author, 2024

4.1.6 LINDSEI-BR relationship verb occurrences

The last group of SVs analyzed in LINDSEI-BR, the group of relationship verbs, had 751 occurrences. Like the group of possession verbs, almost one hundred percent of the occurrences in this group belong to one verb only, the verb “like”. The verb “love” comes second, with 22,37% of the occurrences. The other three verbs did not present high-frequency usage.

Graph 4.1.6: LINDSEI-BR relationship verb occurrences



Source: elaborated by the author, 2024

Considering the verb forms, the verb “dislike” presented two VBD occurrences and one VBP. The verb “hate” also presented two occurrences of VBD, five of VBP and one of VBZ for this verb. The following verb, “like”, had 226 VB occurrences, 43 VBD, one VBG, 290 VBP, and 10 VBZ; no occurrence of VBN was found for this verb. The verb “love” had eight VB occurrences, 20 VBD, one VBG, no VBN occurrences, 135 VBP, and four VBZ. The verb “trust” presented one VB occurrence and one VBD. No occurrences were found for the other forms.

Table 4.1.6: LINDSEI-BR relationship verb forms occurrences

	Dislike	Hate	Like	Love	Trust
VB	0	0	226	8	1
VBD	2	2	43	20	1
VBG	0	0	1	1	0
VBN	0	0	0	0	0
VBP	1	5	290	135	0
VBZ	0	1	10	4	0

Source: elaborated by the author, 2024

4.2 SBCSAE SVs occurrences

In SBCSAE, 18,338 occurrences of the selected SVs were found. This number is much bigger compared to the number of occurrences found in LINDSEI-BR corpus. However, as we are not directly comparing the number of occurrences in each corpus but rather the usage pattern, this considerable number does not impact the analysis.

As shown in Graph 4.2, almost half of the form occurrences were related to the verb "be". When we analyze the distribution of the semantic categories, we can see that the results are remarkably similar to those found in the analysis of the LINDSEI-BR corpus. Mental verbs represent the highest occurrences, followed by possession verbs, as in LINDSEI-BR. So far, the results found in SBCSAE concerning the verb "be", mental verbs, and possession verbs do not differ among learners and native speakers.

The following category in SBCSAE is the category of perception verbs, responsible for 6,68% of the occurrences. Here, we can notice the first difference in usage among the two groups compared, as the results found in LINDSEI-BR corpus show that the third group of verbs with the highest frequency of usage is the group of relationship verbs.

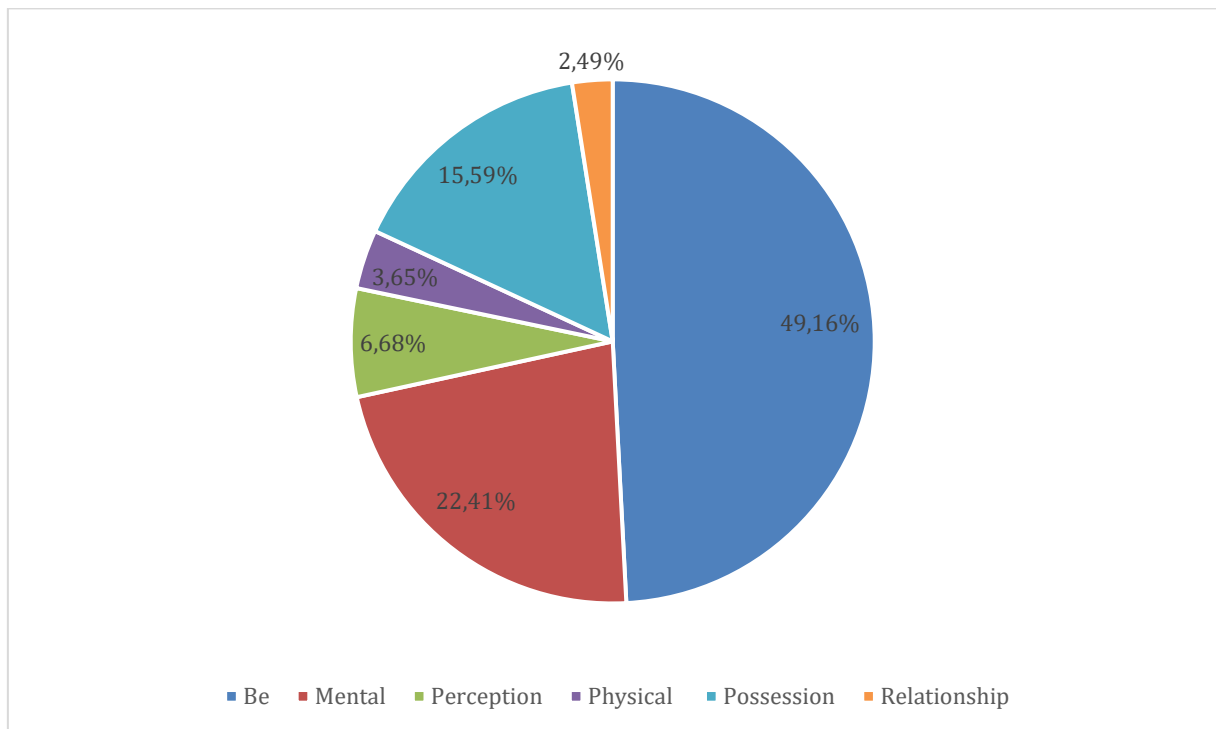
The groups with the lowest frequency of usage of SBCSAE are the groups of physical and relationship verbs, with 3,65% and 2,49% of the occurrences, respectively. This is interesting because, due to the nature of the interactions in this corpus, people would be more likely to talk about their feelings.

The results found in this first section of SV occurrences in the SBCSAE indicate that the most significant difference in usage between native speakers and learners lies

in relationship and perception verbs. Besides that, the results are the same: high frequency of usage of “be”, mental verbs, and possession verbs, and low frequency of usage of physical verbs.

Let us now analyze the distribution of the forms and frequency of the progressive for the verb “be” and the verbs in each semantic category.

Graph 4.2: SBCSAE SVs occurrences



Source: elaborated by the author, 2024.

4.2.1 SBCSAE verb “be” occurrences

As shown in Table 4.2.1, the verb “be” can be found in all the presented forms. When we compare all the forms, the use of the progressive represents the lowest occurrence, and a similar pattern was observed in the results found in LINDSEI-BR. Thus, we can assume that the verb “be” usage among these two groups is not different. A similar pattern can also be observed for the other forms when we compare the two corpora.

Table 4.2.1: SBCSAE verb “be” forms occurrences.

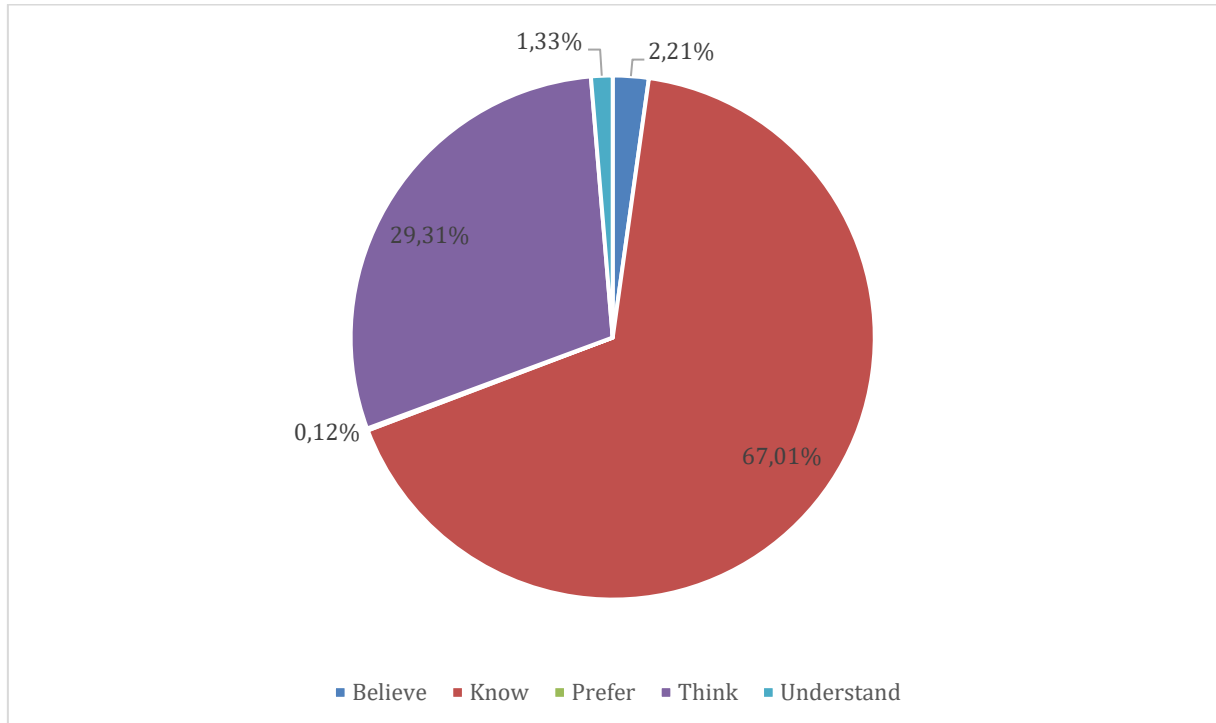
BE	
VB	1,170
VBD	3,666
VBG	118
VCN	409
VBP	3,208
VBZ	444

Source: elaborated by the author, 2024

4.2.2 SBCSAE mental verb occurrences

For the mental verbs, the results found in the LINDSEI-BR corpus showed that “know” and “think” are the verbs responsible for most of the occurrences, followed by “understand”, “prefer”, and “believe”. Comparable results can be observed in the findings in SBCSAE, as shown in graph 4.2.2. The two verbs with the highest occurrences are “know” and “think”, respectively, followed by “believe”, “understand”, and “prefer”. Although these last three verbs have a low frequency of usage, as in LINDSEI-BR, if we compare the two corpora, the order of frequency will be different. While in LINDSEI-BR, the order is “understand”, “prefer”, and “believe”, in SBCSAE, we have “believe”, “understand”, and “prefer”. Thus, native speakers use the verb “believe” more than Brazilian learners and show no preference for the verb “prefer” when speaking.

Graph 4.2.2: SBCSAE mental verb occurrences



Source: elaborated by the author, 2024

Regarding the verb forms for each of the verbs in this group, as seen in table 4.2.2, again, the verbs with the highest frequency are also the verbs that present the most usage of the progressive, the verbs “think” and “know”. Whereas in LINDSEI-BR, the other verb with the most progressive usage is the verb “understand”, in SBCSAE, we have the verb “believe”, and in both corpora were not found occurrences of the progressive for “prefer”. Therefore, the results indicate that high-frequency verbs can also be associated with the most progressive usage.

Table 4.2.2: SBCSAE mental verb forms occurrences

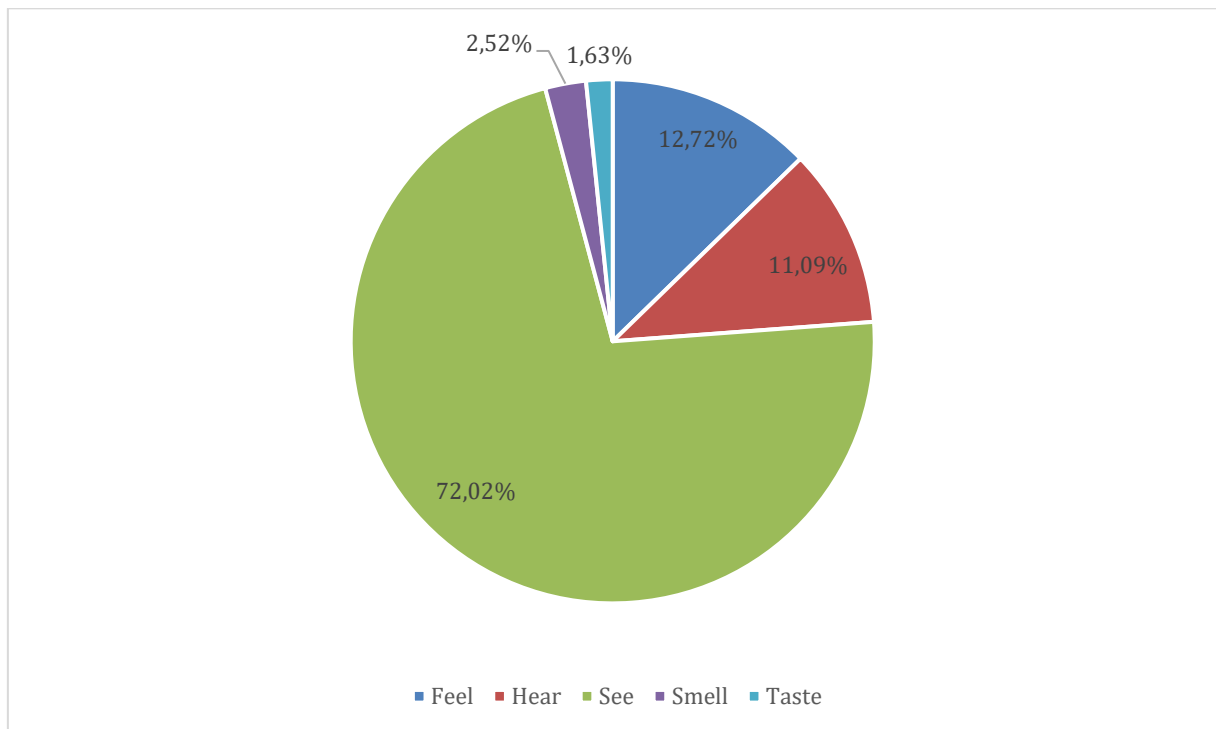
	Believe	Know	Prefer	Think	Understand
VB	63	676	2	251	35
VBD	2	107	1	255	5
VBG	4	8	0	70	0
VBN	0	14	0	13	2
VBP	21	1,921	0	596	11
VBZ	1	29	2	20	2

Source: elaborated by the author, 2024

4.2.3 SBCSAE perception verb occurrences

For the perception verbs group, a total of 1,226 occurrences was found in the SBCSAE. Graph 4.2.3 shows that “see” is the verb with most occurrences, more than 70%. The same result was observed in LINDSEI-BR. Following the verb “see”, the verbs “feel” and “hear” had the most occurrences, 12,72% and 11,09%, respectively. The same pattern was found in LINDSEI-BR. However, the learners used the verb “feel” more than the verb “see”. In both corpora, the verbs “smell” and “taste” had a low frequency of usage, however, in an inverted order. In LINDSEI-BR, “taste” had more occurrences than “smell”, and in SBCSAE, “smell” had more occurrences when compared to “taste.”

Graph 4.2.3: SBCSAE perception verb occurrences



Source: elaborated by the author: 2024

As shown in Table 4.2.3, again, the verbs with the highest occurrences (feel, hear, see) are the verbs with the most usage of the progressive. There is only a slight difference when comparing the results found in LINDSEI-BR. While in SBCSAE, the verb “feel” had more progressive occurrences than the verb “hear,” in LINDSEI-BR, “hear” had more occurrences. The verbs “smell” and “taste” had the same results.

In section 4.1.3, it was mentioned that the results found in LINDSEI-BR for the perception verbs might have been influenced by the nature of the tasks used to compile the corpus and that the results found in this group's query in SBCSAE would help determine whether it was true or not. It seems that it is not, as there were no significant differences between the usage of these verbs among the two groups of participants studied here.

Table 4.2.3: SBCSAE perception verb forms occurrences

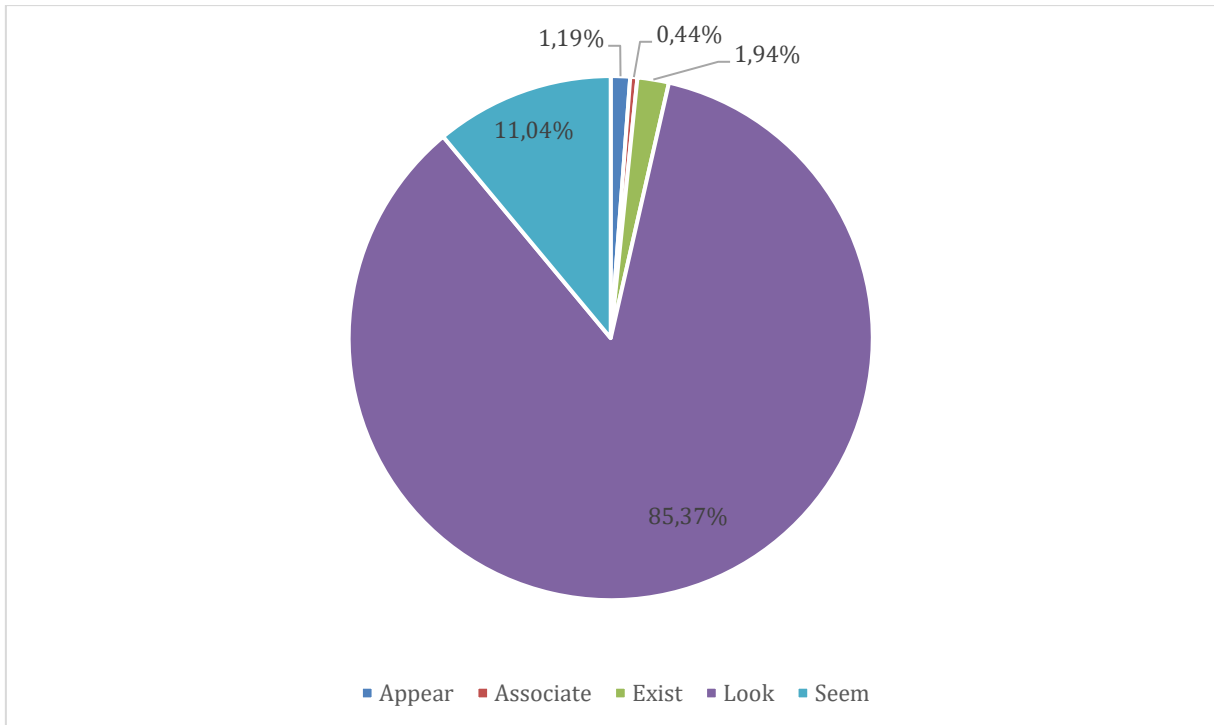
	Feel	Hear	See	Smell	Taste
VB	45	49	537	9	8
VBD	35	34	104	0	2
VBG	16	6	12	1	0
VBN	2	29	64	0	0
VBP	45	16	160	3	1
VBZ	13	2	6	18	9

Source: elaborated by the author, 2024.

4.2.4 SBCSAE physical verb occurrences

The total number of physical verb occurrences in SBCSAE was 670. As seen in graph 4.2.4, the verb “look” had the most occurrences, followed by the verbs “seem,” “exist,” “appear,” and “associate,” respectively. LINDSEI-BR showed the same sequence; however, in the later corpus, the verb “associate” had no occurrences. Therefore, the native speakers and the learners present the same patterns when using physical verbs.

Graph 4.2.4: SBCSAE physical verb occurrences



Source: elaborated by the author, 2024.

When analyzing the forms for each verb in this group, table 4.2.4 shows that the only verb with occurrences in all forms is the verb “look”. The same can be observed in LINDSEI-BR. Both corpora also presented occurrences of the verb “exist” used in the progressive, and no occurrences of the progressive were found for the verbs “appear” and “seem.” The verb “associate”, although with a low frequency of usage in SBCSAE, three in total, was used once in the progressive. As mentioned, no occurrences were found in LINDSEI-BR for this last verb.

This group deviates from the results found in the other groups already analyzed. Although presenting equivalent results compared to LINDSEI-BR, physical verbs are verbs of lower frequency of usage. This includes verbs used in the progressive that, isolated, did not have many occurrences in the corpus, for instance, the verbs “associate” and “exist”.

Table 4.2.4: SBCSAE physical verb forms occurrences

	Appear	Associate	Exist	Look	Seem
VB	3	0	3	234	10
VBD	3	0	0	94	9
VBG	0	1	8	100	0
VBN	0	2	0	6	0
VBP	0	0	1	53	10
VBZ	2	0	1	85	45

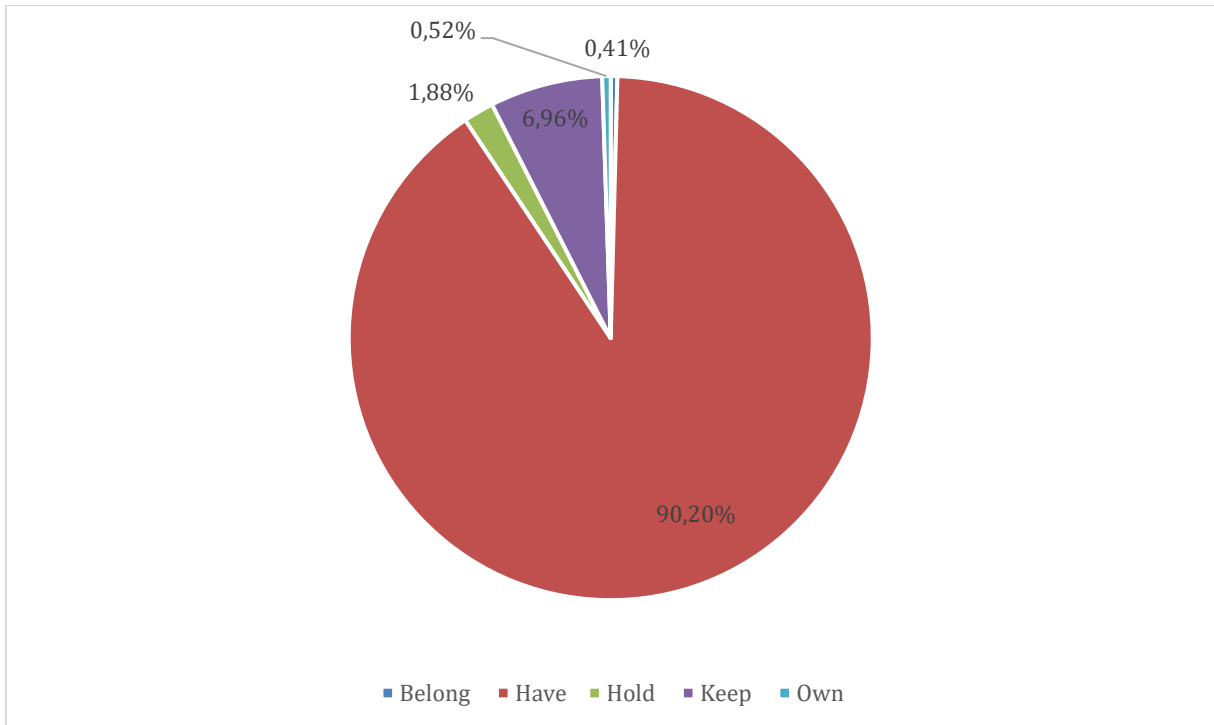
Source: elaborated by the author, 2024.

4.2.5 SBCSAE possession verb occurrences

The possession verb group had 2,859 occurrences, and graph 4.2.5 indicates that “have” is the verb accountable for almost one hundred percent of the occurrences. This was also observed in LINDSEI-BR. As mentioned in section 4.1.5, if the verb “have” did not belong to this category in LINDSEI-BR, this group would also have had the lowest number of occurrences in the SBCSAE.

Both corpora had “keep” as the second verb with the most occurrences, and the results differ for the other verbs when comparing the two corpora. The learners in LINDSEI-BR had “belong” as the third most used verb, whereas in SBCSAE, “belong” was the least used verb. Although not in the same order, the verb “hold” showed a low frequency of usage in both corpora. No occurrences of “own” were found in LINDSEI-BR, and this verb was the fourth most used in the SBCSAE.

Graph 4.2.5: SBCSAE possession verb occurrences



Source: elaborated by the author, 2024.

Table 4.2.5 shows that the selected possession verbs usage is considerably spread in the forms investigated. Except for the verb “belong”, all the other verbs were used in the progressive. Comparing the results found here to those found in LINDSEI-BR corpus, they differ significantly. Possession verb usage among the learners of LINDSEI-BR showed little use of the progressive (have and hold, only), and not many occurrences spread throughout the forms.

Table 4.2.5: SBCSAE possession verb forms occurrences

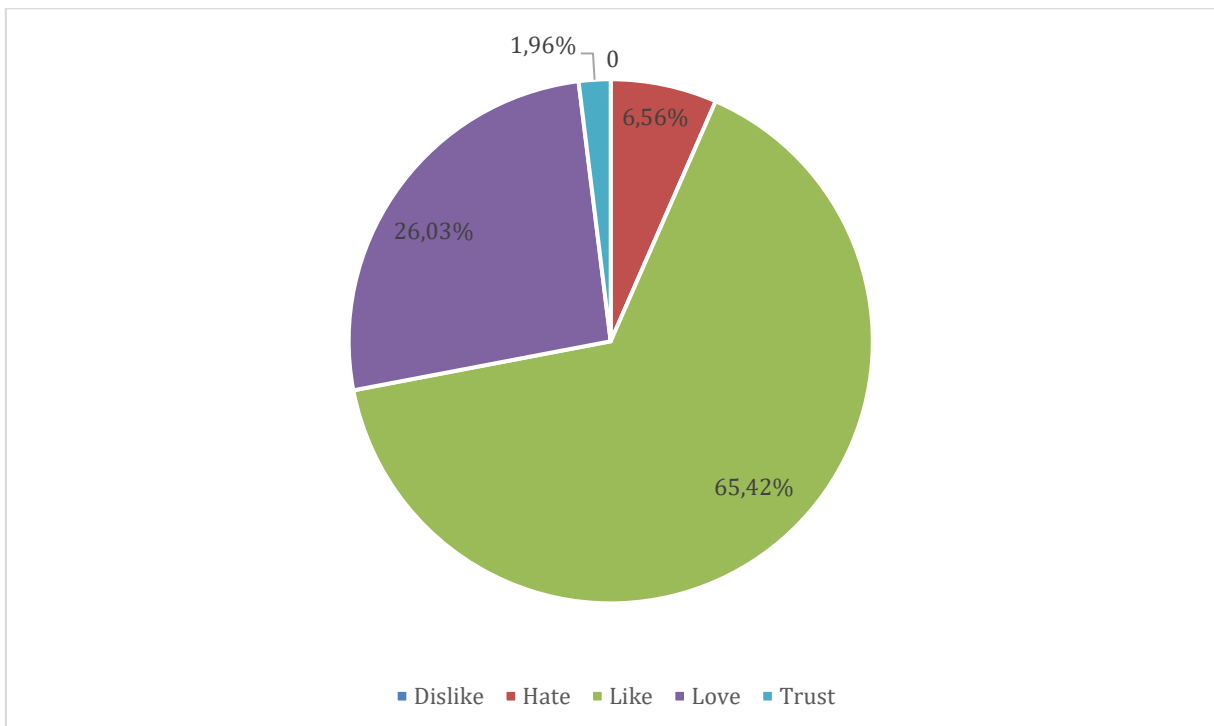
	Belong	Have	Hold	Keep	Own
VB	6	737	32	81	1
VBD	1	684	6	49	3
VBG	0	87	3	15	1
VBN	0	67	4	7	5
VBP	2	790	2	30	5
VBZ	3	214	7	17	0

Source: elaborated by the author, 2024.

4.2.6 SBCSAE relationship verb occurrences

The selected verbs of the semantic category of relationship verbs had a total of 457 occurrences in the SBCSAE, of which 65,42% belonged to the verb “like”. The same result was found in LINDSEI-BR, and both corpora had the verb “love” as the second most used. Contrary to LINDSEI-BR findings for the verb “dislike”, SBCSAE had no occurrences for this verb. The verbs “hate” and “trust” did not show a high frequency of usage in either corpus. In LINDSEI-BR, the verb “hate” was used more than the verb “trust” in the SBCSAE.

Graph 4.2.6: SBCSAE relationship verb occurrences



Source: elaborated by the author, 2024.

The results in Table 4.2.6 show that only the verb “like” was used in the progressive in SBCSAE. It was expected that the verb “love” would have been used in the progressive based on the results found by Martínez-Vásquez (2018) that indicated an increase in the usage of this verb in the progressive form. These two verbs are also the ones with more use in different forms, and the same was observed in LINDSEI-BR. As mentioned before, there were no occurrences of “dislike” in SBCSAE, and the other two verbs left did not present a high frequency of usage and division among the forms.

Table 4.2.6: SBCSAE relationship verb forms occurrences

	Dislike	Hate	Like	Love	Trust
VB	0	2	155	37	6
VBD	0	6	28	19	0
VBG	0	0	2	0	0
VBN	0	0	1	2	0
VBP	0	22	89	47	2
VBZ	0	0	24	14	1

Source: elaborated by the author, 2024.

4.3 Semantic Tags

PyMUSAS is the semantic tagger available in LancsBox X. It was first developed by Paul Rayson in C only for English and later created in Python by Andrew Moore. The tagger uses the UCREL Semantic Analysis System (USAS) as a framework. This derives from the framework first outlined in Tom McArthur's Longman Lexicon of Contemporary English (McArthur, 1981) and contains a multi-layered architecture that includes 21 primary discourse domains (illustrated below). These are further broken down and, in some instances, allow for even more detailed sub-classifications. A full description of all the USAS category systems can be seen in Appendix C.¹⁴

¹⁴ For a full comprehension read <https://ucrel.lancs.ac.uk/usas/usas%20guide.pdf>.

Figure 4.3: USAS discourse domains

A general and abstract terms	B the body and the individual	C arts and crafts	E emotion
F food and farming	G government and public	H architecture, housing and the home	I money and commerce in industry
K entertainment, sports and games	L life and living things	M movement, location, travel and transport	N numbers and measurement
O substances, materials, objects and equipment	P education	Q language and communication	S social actions, states and processes
T Time	W world and environment	X psychological actions, states and processes	Y science and technology
Z names and grammar			

Source: available at <https://ucrel.lancs.ac.uk/usas/>. Accessed on May, 2024.

The following sections present the semantic tags found in the two corpora here analyzed. I do not intend to quantify the tags or deepen the discussion on this topic; I intend only to present the discourse domains found.

4.3.1 Semantic Tags for Be in LINDSEI-BR

Primarily, the tag set associated with using "be" in all instances found in LINDSEI-BR is associated with the idea of existence (A3+). The verb "Be" is also used to convey the idea of evaluation (A5.4+) and kinship (S4), indicating relationships between family members. Additionally, the verb "be" has one occurrence where it is tagged as a grammatical term (Z5).

4.3.2 Semantic Tags for Mental Verbs in LINDSEI-BR

The first verb in the group of mental verbs in LINDSEI-BR is "believe". According to the results, this verb is used to indicate thought and belief (X2.1), evaluation (A5.4+), and as a discourse marker and emphatic term (Z4). The semantic tags associated with the verb "know" indicate evaluation (A5.4+), knowledge/perception/retrospection (X2.2+), and discourse markers and emphatic terms (Z4). The verb "prefer" is used to indicate fondness/affection/partiality/attachment or the lack thereof (E2++) and evaluation (A5.4+). "Think" is used to express evaluation (A5.4+), thought, and belief (X2.1) and as a discourse marker and emphatic term (Z4). "Understand" is used to express evaluation (A5.4+) and understanding/comprehension (X2.5). Most occurrences of mental verbs used by the learners in LINDSEI-BR are associated with mental states. Curiously, PyMUSAS also tags some of these as discourse markers, which requires further investigation to understand and explain this phenomenon better.

4.3.3 Semantic Tags for Perception Verbs in LINDSEI-BR

The following category is that of perception verbs. "Feel" is the first verb in this group, and it is used to express evaluation (A5.4+), reasoning/thinking, level of belief/skepticism (X2.1), and desire/aspiration (X7+). "Hear" is used to express evaluation (A5.4+) and sensory terms relating to sound (X3.2). The verb "see" is used to express evaluation (A5.4+), terms related to crime/criminal activities and the legal system (G2.1), sensory terms relating to sight (X3.4+), and discourse markers, emphatic communication terms (Z4). The only occurrence of "smell" is used to express sensory terms relating to smell (X3.5). The occurrences of "taste" are used to express evaluation (A5.4+) and sensory terms relating to taste (X3.1).

4.3.4 Semantic Tags for Physical Verbs in LINDSEI-BR

In the next group of verbs used to describe physical states, the first verb, "appear," indicates abstract terms relating to appearance/impression (A8). The verb "exist" has all its occurrences associated with a tag indicating abstract terms relating to existence (A3+). The verb "look" is used to express evaluation (A5.4+) and mainly to refer to abstract terms relating to appearance/impression (A8), similar to the verb

"appear." The last verb in this group, "seem", is used in the same manner as "appear" and "look" to evaluate something/someone (A5.4+) and to refer to abstract terms relating to appearance/impression (A8).

4.3.5 Semantic Tags for Possession Verbs in LINDSEI-BR

In the category of possession verbs, the only tag used with the verb "belong" indicates the speaker is talking about general/abstract terms relating to allocating/relinquishing/acquiring/receiving, etc. (A9+). The verb "have" in LINDSEI-BR is used to evaluate something/someone (A5.4+), to discuss general/abstract terms denoting causal relationships, or lack thereof (A2.2), to talk about general/abstract terms relating to allocating/relinquishing/acquiring/receiving, etc. (A9+), with terms depicting fondness/affection/partiality/attachment, or lack thereof (E2+), with terms relating to moral principles/accepted moral practices, or lack thereof (G2.1), and as discourse markers, emphatic communication terms (Z4). The tags associated with the verb "hold" refer to terms depicting putting/taking/pulling/pushing movements/activities (M2). The verb "keep" is used with evaluative terms depicting authenticity (A5.4+) and general/abstract terms relating to allocating/relinquishing/acquiring/receiving, etc. (A9+).

4.3.6 Semantic Tags for Relationship Verbs in LINDSEI-BR

Among the relationship verbs, the only tag associated with "dislike" indicates that this verb is used evaluatively (A5.4+). The verb "hate" is used to evaluate a term (A5.4+) and with terms depicting fondness/affection/partiality/attachment or the lack thereof (E2+). The verb "like" has its occurrences tagged to indicate the speaker is evaluating a term (A5.4+), discussing terms depicting fondness/affection/partiality/attachment, or the lack thereof (E2+), abstract terms relating to appearance/impression (A8), and discourse markers, emphatic communication terms (Z4). The verb "love" is used to evaluate (A5.4+) and to depict fondness/affection/partiality/attachment or the lack thereof (E2+). Finally, the verb "trust" is used with evaluative terms depicting authenticity (A5.4+).

In conclusion, we can see that the meanings attributed to the SVs used by the learners in LINDSEI-BR are mostly associated with the states they inherently describe. Now, let us proceed to the results found in SBCSAE.

4.3.7 Semantic Tags for “be” in SBCSAE

The tags for "be" in SBCSAE are similar to those found in LINDSEI-BR; most are associated with the idea of existence (A3+). Both corpora present the verb "be" used with the idea of evaluation (A5.4+). Unlike LINDSEI-BR, which had occurrences of "be" associated with kinship (S4), indicating relationships between family members, and instances where it was tagged as a grammatical term (Z5), SBCSAE includes tags indicating general/abstract terms relating to allocating/relinquishing/acquiring/receiving, etc. (A9+), abstract terms denoting importance/significance (A11.1+), and terms relating to crime/criminal activities and the legal system (G2.1).

4.3.8 Semantic Tags for Mental Verbs in SBCSAE

With the verb “believe,” the semantic tags used are the same as those in LINDSEI-BR. They indicate thought and belief (X2.1), the idea of evaluation (A5.4+), and discourse markers and emphatic terms (Z4). The verb “know” in SBCSAE has tags similar to those associated with “know” in LINDSEI-BR. They indicate the idea of evaluation (A5.4+), terms relating to the level of knowledge/perception/retrospection (X2.2+), and discourse markers and emphatic terms (Z4). However, two tags not associated with the verb “know” in LINDSEI-BR are present in SBCSAE; they indicate terms relating to crime/criminal activities and the legal system (G2.1) and abstract terms denoting importance/significance (A11.1+). The verb “prefer” is used with terms depicting fondness/affection/partiality/attachment, or the lack thereof (E2++), the same result found in LINDSEI-BR. The verb “think” in both corpora is used to express the idea of evaluation (A5.4+), thought and belief (X2.1), and as discourse markers and emphatic terms (Z4). However, one tag in SBCSAE uses terms depicting the position of/point of reference for X (M6). The verb “understand” is also used to express the idea of evaluation (A5.4+) and with terms depicting the level of understanding/comprehension (X2.5). The same was observed in LINDSEI-BR.

4.3.9 Semantic Tags for Perception Verbs in SBCSAE

As in LINDSEI-BR, the verb "feel" is associated with the idea of evaluation (A5.4+), terms relating to reasoning/thinking, level of belief/skepticism (X2.1), and terms depicting the level of desire/aspiration (X7+). One tag found in SBCSAE but not present in LINDSEI-BR is used with terms depicting the level of permission/consent/authorization (S7.4+). With the verb "hear," the tags found are the same as those in LINDSEI-BR, expressing evaluation (A5.4+) and sensory terms relating to sound (X3.2). The verb "see" in SBCSAE shows only one difference compared to the results found in LINDSEI-BR. The tags used to express evaluation (A5.4+), sensory terms relating to sight (X3.4+), and as discourse markers, emphatic communication terms (Z4) are present in both corpora. However, in LINDSEI-BR, the tag G2.1, used to discuss terms related to crime/criminal activities and the legal system, was not found in SBCSAE. The verb "smell," as shown in the section that quantified the occurrences found in each corpus, had more occurrences in SBCSAE than in LINDSEI-BR. The tag X3.5, which indicates sensory terms relating to smell, was found in both corpora. However, in SBCSAE, this verb was also used to evaluate (A5.4+). As in LINDSEI-BR, the verb "taste" in SBCSAE was used to express evaluation (A5.4+) and sensory terms relating to taste (X3.1).

4.3.10 Semantic Tags for Physical Verbs in SBCSAE

All occurrences of "appear" in SBCSAE indicate abstract terms relating to appearance/impression (A8), consistent with the observations in LINDSEI-BR. There were no occurrences of the verb "associate" in LINDSEI-BR. In SBCSAE, the verb "associate" is used to evaluate (A5.4+), denote comparative terms indicating similarity/difference (A6.1+), and relate to mental actions and processes in general (X2.1). The verb "exist" in both corpora presents the same tag associated with abstract terms relating to appearance/impression (A3+). The verb "look" in SBCSAE, as in LINDSEI-BR, is used to express the idea of evaluation (A5.4+) and primarily to discuss abstract terms relating to appearance/impression (A8). However, in SBCSAE, more tags are associated with "look," indicating terms depicting the position of/point of reference for X (M6), descriptive terms relating to the appearance/look of X (O4.2+),

and terms depicting commencement/completion (T2). The verb "seem" is used in the same manner in both corpora to evaluate something/someone (A5.4+) and to discuss abstract terms relating to appearance/impression (A8).

4.3.11 Semantic Tags for Possession Verbs in SBCSAE

As in LINDSEI-BR, "belong" had only one tag associated with its usage, indicating general/abstract terms relating to allocating/relinquishing/acquiring/receiving, etc. (A9+). Most of the tags found for "have" in LINDSEI-BR were also found in SBCSAE. This verb was used to evaluate something/someone (A5.4+), to talk about general/abstract terms denoting causal relationships or lack thereof (A2.2), to talk about general/abstract terms relating to allocating/relinquishing/acquiring/receiving, etc. (A9+), and with terms relating to moral principles/accepted moral practices, or lack thereof (G2.1). Tags indicating terms relating to life and death (L1), abstract terms denoting importance/significance (A11.1+), and terms relating to relationships between family members/familiars (S4) were found only in SBCSAE. The verb "hold" in LINDSEI-BR had only one tag associated with its use, referring to terms depicting putting/taking/pulling/pushing movements/activities (M2). SBCSAE also had this tag for "hold," along with other tags indicating evaluation (A5.4+) and terms depicting the position of/point of reference for X (M6). As in LINDSEI-BR, the verb "keep" in SBCSAE was used with evaluative terms depicting authenticity (A5.4+) and general/abstract terms relating to allocating/relinquishing/acquiring/receiving, etc. (A9+). However, two other tags were associated with using this verb, indicating general/abstract terms relating to the level of openness/concealment/exposure, etc. (A10), and terms depicting the level of help/hindrance (S8). The verb "own" had no occurrences in LINDSEI-BR, and all occurrences found in SBCSAE were tagged as general/abstract terms relating to allocating/relinquishing/acquiring/receiving, etc. (A9+), indicating that all occurrences refer solely to possession.

4.3.12 Semantic Tags for Relationship Verbs in SBCSAE

To finish this section, we now deal with the verbs used to describe/talk about relationships. There are no occurrences of "dislike" in SBCSAE. In both corpora, the verb "hate" was used to evaluate a term (A5.4+) and with terms depicting fondness/affection/partiality/attachment or the lack thereof (E2+). The occurrences of the verb "like" in SBCSAE indicate the speaker was evaluating a term (A5.4+), talking about terms depicting fondness/affection/partiality/attachment, or the lack thereof (E2+), and discussing abstract terms denoting importance/significance (A11.1). The latter was not found in LINDSEI-BR. Two other tags used with "like" in LINDSEI-BR, relating to abstract terms concerning appearance/impression (A8) and discourse markers, emphatic communication terms (Z4), were not found in SBCSAE.

The verb "love" presented the same tags in both corpora, being used to evaluate (A5.4+) and with terms depicting fondness/affection/partiality/attachment, or the lack thereof (E2+). In LINDSEI-BR, "trust" was used only with evaluative terms depicting authenticity (A5.4+). This tag was also found in SBCSAE, but there is also the occurrence of a tag associated with terms relating to reasoning/thinking and the level of belief/skepticism (X2.1).

As shown in this section, the tags found in both corpora do not satisfactorily differ, and most of them keep their mental idea. There were some differences between the use made by learners and natives, indicating that the native speakers applied different nuances to these SVs; however, there were no significant differences.

The following section deals with the collocates for each BV in this study.

4.4 Collocations

This section compares the five strongest collocates for the base form of the SVs found in both corpora. Some SVs were left out because they were not high in frequency of usage. Thus, few occurrences were found, and LancsBox X could not find collocates for such verbs. Another group of verbs that were left out was the one that did not have occurrences in both corpora. As we compare the occurrences in both corpora, such verbs would not be relevant in this section.

4.4.1 Be Collocates

The analysis of the collocates associated with the verb "be" in both LINDSEI-BR and SBCSAE reveals some interesting patterns in language use among learners and native speakers. In LINDSEI-BR, the presence of "would," "teach," "to," "want," and "an" as collocates suggest a focus on desires and future aspirations, likely reflecting the academic and aspirational context of the learners, who are predominantly studying to become teachers. This is particularly evident from collocates like "teach" and "want," notably absent in SBCSAE.

In contrast, SBCSAE participants prefer colloquial forms such as "gonna" and contractions like "'ll," indicating a more informal and spontaneous use of language. The occurrence of "gonna" and "'ll" as strong collocates in SBCSAE but not in LINDSEI-BR underscores the more casual speech patterns of native speakers than the formal usage by learners in LINDSEI-BR. This distinction might be attributed to the quasi-spontaneous nature of speech in LINDSEI-BR, where participants, despite being in a less controlled setting than a structured interview, still operate within an academic setting that influences their language use.

These findings highlight the impact of context on language use. Learners tend to adopt a more formal register, possibly due to their educational settings and future career goals. At the same time, native speakers exhibit a more relaxed and varied use of language in everyday contexts. This analysis provides insights into how different environments and purposes for language use can shape linguistic patterns, including the choice and frequency of certain collocates with common verbs like "be".

- LINDSEI-BR "be" collocates: Would, Teach, To, Want, An

Examples with "would":

Nation's Pride yeah **would be** Stolz der Nation yeah in
no idea what that will be but I **would be** glad
culture maybe I **would** like **be** fascinated and want to stay
'm I'm studying to **be** a translator I **would** like
there so I I **would be** more I don't know

Examples with “teach”:

comfortable and learn how to **be a teacher** you know because

> great chance to to **be a good teacher** </A

fiancée a girlfriend ... you wanna **be a: an English teacher** ... you

after graduation do you wanna **be an English teacher:** or translator:

so you are studying to **be: an English teacher** I suppose

Examples with “to”:

yes it seems **to to be** there is a woman and

know if I want **to be** a teacher for the rest.

of impossible but that would **be** something I'd like **to**

s= she asks him **to be** in a big square because

they didn't use **to be** my friends you know because

Examples with “want”:

know if we **want to be** friends of her anymore I

I don't **want to be** like I don't want

they **want** the candidate to **be** on the show they press

culture maybe I would like **be** fascinated and **want** to stay

way I **want** them to **be** I can survive emotionally ow

Examples with “an”:

like his dream: is to **be an** artist and he's

fiancée a girlfriend ... you wanna **be a: an** English teacher ... you

after graduation do you wanna **be an** English teacher: or translator:

student do you want to **be** a Spanish teacher **an** English

so you are studying to **be: an** English teacher I suppose

- SBCSAE “be” collocates: would, gon, 'll, to, na (gonna)

Examples with “would”:

but, ALICE: **Wouldn't it be** funny if she taught English,

said, that it **would** probably, **be** brought up, at board retreat,

we thought maybe registers **would be** important. And I'd, I

FRAN: And what happened **would be**, rather than have it, things you could do, that **would be** just really helpful? STEVEN: Say

Examples with “gonna”:

CURT: It's **gonna be** another wonderful Saturday. KITTY: YELL know. There's **gonna be** tasks. BRAD: Mhm. PHIL: BRAD:

I'm not **gonna be** on Cops. Cause that's

LARRY: SETH: You **gonna be** putting anything there? LARRY: I 're **gonna** need to **be**, reallocated to this area. And

Examples with “ll”:

at, GENEVIEVE: Honey you'll **be** away from parents anyway. STEPHANIE: occasion, and you'd all **be** welcome, I'll leave some

and, you'll st- still **be** able to enjoy the the

DORIS: Lazix. but I'll **be** taking the generic drug. Of it's not displaced. should be pretty stable. It'll **be**

Examples with “to”:

having to learn how **to be** more assertive. Otherwise his show

imagine you'd, it would be tough **to Be** like watching

waitlisted. And Williams used **to be** really lib- liberal whatever. But

I just think, they wanna **be** able **to** just kind of

in demo fashion, you'll **be** so low **to** the ground,

Examples with “na”:

the street are gonna **be** coming after them. You know?

What're you gonna **be**. STEPHANIE: I'm gonna

I'm not gonna **be** here on my birthday. ALICE:

I'm not gonna **be** there until January. CAROLYN: Have- for, there're gonna **be** a lot of community resources,

4.4.2 Mental Verbs Collocates

In both corpora, common collocates for mental verbs such as "think" and "know" include "do," "n't," and "about," indicating a shared use of these verbs to express negation and abstraction or reflection. Unique to LINDSEI-BR are collocates like "take" and "much," which may suggest a more deliberate or measured way of expressing mental processes. In contrast, SBCSAE introduces "why" and "but," which could imply a more interrogative or contrasting mode of expression, and a higher use of the personal pronoun "I," reflecting a more subjective or personal discourse.

LINDSEI-BR Mental Verbs collocates examples:

- Believe: That, You

Examples with "that":

interesting yeah he made you **believe that** it's possible at
 good movies I just can **believe that** you like bad movies
 finish he she could not **believe that** she looked like that
 was there to make then **believe that** they had to move
 my God I can't **believe** she is doing **that** you

Examples with "you":

interesting yeah he made **you believe** that it's possible at
 so much fear do **you believe** <B
 good movies I just can **believe** that **you** like bad movies
 > I can I can **believe** in **you**
 my God I can't **believe** she is doing that **you**

- Know: n't (not), do, I, If, What

Examples with "n't":

brother they I **don't know** I don't know what
 saw once I **don't know** if I remember the: name
 Bissau: or I **don't know** but he got married in

I think parents **don't know** how to raise the children
 A> I **don't know** this people

Examples with "do":

B>I **don't know** <B

both possibilities I **don't know** if it's possible one inside
 don't know I **don't know** if that's the pronunciation in
 believe and I **don't know** I think it was very
 > yes I **don't know** the the English version but

Examples with "I":

until today I **don't know** I really don't know
 > and: I **don't know** ... I don't know and
 's another I **don't know** it's different right this
 I'm very happy to **know** that we have a chance
 I think parents **don't know** how to raise the children

Examples with "if":

say no I **don't know if** we want to be
 this painter I **don't know if** it's the proper
 very different I **don't know if if** an exchanged= exchanged
 is from I **don't know if** it's from Guiné
 way cause I **don't know if** is only my opinion

Examples with "what":

know I I **don't know what** I I want to
 nice because you **don't know exactly what** is happening and
 go out: but **don't know what** I= doing th= tonight:
 and: then I **don't know what** I have to you
 to Norway I **don't know what** I'm gonna do

- Think: About, Do, Take, N't, Much

Examples with “about”:

when I entered I: didn't **think about** teaching I was thinking
 right now all I can **think about** is my dissertation because
 that much what do you **think about** the film </A
 no I I don't **think about** it yet I I
 there's so much to to **think about** this because I I

Examples with “do”:

beautiful and you **don't think** I'm beautiful and he
 A> but **do** you **think** that no= how do you
 so yes I **don't think** so <
 B> I **don't think** so of course things are
 <A> **do** you **think** that we are open and:

Examples with “take”:

take your time you can **think** about you're gonna
 can **take** your time to **think** about your narrative </A
 <A> you can think **about** it **take** your time
 in English I don't **think** I I: I could **take**
 > ok **take** your time **think** about <

Examples with “n't”

's French I **don't think** they have any vocals and
 much they I **don't think** they like to talk a
 to do to make I **think** I **don't** know a:
 in Portuguese I **don't think** I would be able to
 I I I **don't think** so ... okay I don't

Examples with “much”:

much they I don't **think** they like to talk a
much things to do and **think** about <
 that **much** what do you **think** about the film </A
 same director this person will **think** it's not that **much** but
 there's so **much** to to **think** about this because I I

- Understand: Did, What, N't, Do, You

Examples with "did":

little strange I **didn't understand** nothing for me it was
 why I I **didn't understand** anything everyone talking was I
 hamburger I I **didn't understand** the beginning till my mom
 Spanish and they **didn't understand** my Portuguese actually I would
 sorry singers I **didn't understand** okay gospel gospel yes but

Examples with "what":

I had no difficulties to **understand what** people s= say because
 to concentrate really hard to **understand what** they were trying to
 word I was able to **understand** so this **what** I what
 but then no one would **understand what** I was talking about
 because you can you can **understand what** she is passing through

Examples with "n't":

little strange I **didn't understand** nothing for me it was
 because I **can't** like **understand** properly the things you know
 time there I **didn't understand** their Spanish and they did
 sorry singers I **didn't understand** okay gospel gospel yes but
 do you **understand** doesn't call my attention

Examples with "do"

about the salary **do** you **understand** okay I **don't understand** can you repeat </B
 in Paris I I **do understand** French from people outside Paris
 more than it **do** you **understand** and I I don't
 do you **understand** doesn't call my attention

Examples with "you"

okay you un= do **you understand** and <
you able to: comprehend or **understand** something about this </A
 words that **you** don't **understand** it can be hard you

 do you **understand** doesn't call my attention
because you can **you** can **understand** what she is passing through

SBCSAE mental verbs collocates examples:

- Believe: can, could, n't, Tom, does

Examples with "can":

Rivers Stadium. CAM: **Can't believe** I didn't know that.
DAN: JENNIFER: I **can't believe** you just won. I was
TOM_: TOM_: you **can't believe** this. TOM_: TOM_: I TOM_:
man. Oh I **can't believe** he said that MANY: WALT:
guy , I **can't** really **believe** that guy's her husband.

Examples with "could":

Bette Davis. I **couldn't believe**. These two old, I do
Jeez. I just **couldn't believe** that. PHIL: BRAD: When I
four thousand. I **couldn't believe** it. TOM_: TOM_: I said
know. MILES: I **couldn't believe** it. This woman I knew
too you know. **Couldn't believe** it. She had like on

Examples with "n't":

off, KENDRA: Mom **doesn't believe** I didn't spend the
DAN: JENNIFER: I **can't believe** you just won. I was
four thousand. I **couldn't believe** it. TOM_: TOM_: I said
who **doesn't believe** - not believe, but doesn't & MELISSA: Agree
saying that, I **don't believe** a forum, is being presented

Examples with "Tom":

Now that I don't **believe**. TOM_: And just is a
incredibly upset I **couldn't believe** this. TOM_: TOM_: Because TOM_:
four thousand. I **couldn't believe** it. TOM_: TOM_: I said
TOM_: TOM_: you can't **believe** this. TOM_: TOM_: I TOM_:
reading Jane Austin. Could you **believe** that? TOM_: She's reading

Examples with “does”:

off, KENDRA: Mom **doesn't believe** I didn't spend the
with anyone who **doesn't believe** - not believe, but doesn't &
who **doesn't believe** - not believe, but doesn't & MELISSA: Agree

- Know: Do, N't, I, How, did

Examples with “do”:

Yeah, PATTY: I **don't know** which way's better for
ALICE: Yeah. I **don't know**. ANNETTE: varsity probably starts at
cold start problem? **Do** you **know** uh, if it's been
So you **don't** really **know**, if there's a LARRY:
do this. You have to **know** that whatever God has given

Examples with “n't”:

guys? LUCY: I **don't know** the neighbors. DAN: LUCY: I
two old, I **don't know**, I wanna say windbags. Ha-
SETH: So we **don't know** if there's any particular
NICK: DARREN: I **don't know** what the hell I'm
you? ARIANNA: I **don't know**. He really bugs me. He

Examples with “I”:

church. Somewhere I **don't know**. WENDY: Were they going to
track really well, let me **know**, and I'll We'll
still I still **don't know** MARK: JUDGE: that may have
playing, RICHARD: I **don't know** FRED: because last week, the
n't know , I **don't know**. TIM: They show the uh-,

Examples with “how”:

But what he doesn't **know how** to separate is, he
attic. So I **don't know how** old it is. MARCI:
KEVIN: **How** long, do you **know**? MARIE: Uh his fever? KEVIN:
seventy-seven. I **don't know how** many owners she's

watch you, I don't **know how** many Americans have been

Examples with "did":

ALICE: Well I **didn't know**. ANNETTE: That's sad about time. PATRICK: I don't **know**. **Did** you wanna do something

PETE: MILES: I **didn't know** this, but apparently in Brazil,

MANY: WOOD: I **didn't know** whether I was or not,

the- SHERRY: They **didn't know** what he had. He had

- Think: Do, N't, Would, About, I

Examples with "do":

book? KAREN: Yeah. **Do** you **think** you want to read that

No, GILBERT: I **don't think** MARIA: the other guy. GILBERT:

the freezer, what **do** you **think** would happen to the water.

aren't. I **don't think** it's everything. JIM: not

KEN: No. I **don't think** so. JOANNE: Isn't Nicaragua

Examples with "n't":

LAJUAN: I I **don't think** I don't care what

the left, I **don't think** so. KEVIN: Oh, okay. KENDRA:

doing? SHARON: I **don't think** we can have pie till

nh. RON: I **don't think** so. FRANK: and, recorded the

goldfish, and I **don't think** I'd ever seen it

Examples with "would":

probably, I I **wouldn't think** he would be more than

the freezer, what do you **think would** happen to the water.

an un- -believable concept? to **think** that maybe I **would** be

you know, I don't **think** he **would** do anything, when

KEN: JOANNE: that I **would think** that they would have, a

Examples with “about”:

but I'll have to **think about** it He's real
 anyway, what it made me **think about** was, JILL: JEFF: the
 to see what people would **think about** you? JAN: No. FRANK:
 Oh okay I don't **think** I know **about** that project.
 And Sharon doesn't even **think about** what she's saying.

Examples with “I”:

Right. LAJUAN: I don't **think** so. And and if they
 You know, I don't **think** they really liked my answer,
 opening. And I we may **think** about Gail, PHIL: Well Gail is very
 guy. GILBERT: I don't **think** there's as much trust,
 uh SETH: Yeah. Do you **think**, SETH: well, I guess you

- Understand: Why, n't, did, do, but

Examples with “why”:

Which I, I couldn't **understand**. JOHN: Yeah. LUCY: **Why** they
 's **why** I don't **understand** where uh they come up
 Well then I don't **understand why** you like Rita. PATRICK:
 we get into it, to **understand, why** is it, including ourselves,
 of Chicanos don't even **understand why** we don't vote,

Examples with “n't”:

ANDREW: CINDY: I **can't understand** you. DARLENE: I don't
 your English, you **didn't understand** what he was saying, LENORE:
 and I **don't** even **understand** the definition. DARRYL: The definition
 cards though. I **don't understand** that. JENNIFER: Every time you
 Because they MRC **didn't understand** MRC. And they were no

Examples with “did”:

but if you **didn't understand** your English, you didn't
 were terrified. They **didn't understand**. What was going on. And
 Because they MRC **didn't understand** MRC. And they were no

be careful. I don't **understand** that. MARIE: Because I **did**
The thing I **didn't understand** is, LINDA: I don't

Examples with "do":

ANDREW: CINDY: I can't **understand** you. DARLENE: I **don't**
's why I **don't understand** where uh they come up
and I **don't** even **understand** the definition. DARRYL: The definition
just KEVIN: You **don't understand** MARIE: So we just give
of this. I **don't understand** what not taking his lunch

Examples with "but":

But, Cathy and Jawahar don't **understand** how to mix their friends
but if you didn't **understand** your English, you didn't
but a small measure to **understand**, where we're really at.
cutting cutting edge, can be **understand understood**
intellectually, MICHAEL: **but** those
but you just can't **understand** em? No, he's okay

4.4.3 Perception Verbs Collocates

Both corpora display collocates "can" and "could" for "hear" and "see," pointing to modal verbs' role in discussing potential and abilities in perception. SBCSAE uniquely includes negations like "n't" for "feel" and "hear," absent in LINDSEI-BR. This might indicate a greater tendency among native speakers to discuss the absence or negation of sensory experiences. LINDSEI-BR employs more definitive terms like "I" and "so," suggesting a more direct or assured expression of perception.

LINDSEI-BR perception verbs collocate examples:

- Feel: So, And, I

Examples with "so":

for me I didn't **feel** it **so** much and then
 classes just **so** I could **feel** more confident and come back
 one had ma= made me **feel so so**: so inside the

Examples with "and":

and: they: also made me **feel** really welcome they showed me
 for me I didn't **feel** it so much **and** then
 classes just so I could **feel** more confident **and** come back
 a way that she would **feel** happy **and** she would feel
and: you know you can **feel** the city next everything **and**:

Examples with "I":

for me I didn't **feel** it so much and then
 classes just so I could **feel** more confident and come back
 my my dad don't **feel** happy about it but I
 okay but I: didn't **feel** like you I I **feel**
 but the ending I can: **feel** more connected with the the

- Hear: Can, Some, So, To, Was

Examples with "can":

can speak something I **can hear** I can talk about su=
 that you **can hear** the language so when I:
 internet too so you **can hear** all radio in United States
 B> no I **can hear** but it doesn't call.
 my attention but I **can hear** to: to practice my my

Examples with "some":

radio and begin to: to **hear** to listening **some some some**
 understand better when I am **hear some some** joke or: maybe:

Examples with "so":

that you can you can **hear** the language **so** when I:
 internet too **so** you can **hear** all radio in United States
so: I was I was: **hear** that what they was speaking
 begin to **so** you: you: **hear** <A
 people I really wanted to **hear** them talking **so** I did

Examples with "to":

I I like **to to hear** music wondering what when I
 to them you refer **to hear** them than go to was
 radio and begin to: **to hear** to listening some some some
 people I really wanted **to hear** them talking so I did
 my attention but I can **hear to: to** practice my my

Examples with "was":

to them you refer to **hear** them than go to **was**
 so: I **was** I **was**: **hear** that what they was speaking
 it **was** really beautiful to **hear** them and what more <

- See: Let, Can, How, Could, See

Examples with "let":

right now but **let's see** how that goes but academically
 let's see she's here now she
 people what else **let me see** <A
 what to do **let's see** <A
 let me see another dream

Examples with "can":

girls so that they **can see** what the what he s=

was and as we **can see** she was not very happy
 I can see I **can see** I can learn another lesson
 can read clearly I **can see** that is an good English
 Viçosa they are trying to: **see** if they **can** take that

Examples with "how":

city in but just to **see how** would it be nowadays
 to see the culture to **see how** people dress and what
 it was good to to **see how** especially the parents how
 right now but let's **see how** that goes but academically
 and then she went to **see how** it was and it

Examples with "could"

because and I also **could see** that Brazilian people are very
 results because she **couldn't** that she was not what
 was Danish but: I **could see** yeah almost the same but:
 to the scenes you **could see** that a character was in
 because she she **couldn't see** herself that way it's ok

Examples with "see":

to **see** that and to **see** how different the the the
 to **see** the culture to **see** how people dress and what
 I can **see** I can **see** I can
 B> okay can I **see** it again can I **see**
 can **see** anything can't **see** anything and I I go

SBCSAE perception verbs collocates examples:

- Feel: n't, like, did, do, but

Examples with "n't":

and then I **didn't feel** that great. So, ANNETTE: Mhm.
 cause they, they **didn't feel** comfortable with us living together
 Mhm. And I **don't feel** like measuring. : JAN: Something does

so, I **don't** wanna **feel** discouraged , cause like I'm
in there. I **didn't feel** anything in his belly, but,

Examples with “like”:

fill? KATHY: Just makes me **feel like** I ate a candy
Mhm. And I don't **feel like** measuring. : JAN: Something does
's okay Did your knee **feel like** it had some play
cool, cause I don't **feel like** going anywhere. NANCY: Yeah.
so, I don't wanna **feel** discouraged , cause **like** I'm

Examples with “did”:

and then I **didn't feel** that great. So, ANNETTE: Mhm.
cause they, they **didn't feel** comfortable with us living together
Yeah but, I **didn't feel** like she taught it any
birds, but, they **didn't feel** that they had enough eggs,
's okay **Did** your knee **feel** like it had some play

Examples with “do”:

So, but uh, **don't feel** bad about keeping Horizon on
don't I **don't feel** comfortable here. GAIL: I do
mean, so they **don't feel** singled out? SHARON: Now I
don't know, you could **feel** something that wouldn't be
DANA: but I **don't feel** like making it. KELLY: Think

Examples with “but”:

So, **but** uh, don't **feel** bad about keeping Horizon on
then what- how would you **feel**. **But** Cathlene's like, you
late. **But** I don't **feel** any of the symptoms. But
but they wouldn't necessarily **feel**, RICKIE: JUNE: REBECCA: scared or
in there. I didn't **feel** anything in his belly, **but**,

- Hear: could, did, can, about, me

Examples with “could”:

hear you. ANDREW: You **could hear** me? DARLENE: I heard you, talk to Hector, I **could hear** him a little bit, but it was amazing, I **could hear** people there, but they would Uh, RICHARD: And you **could hear** me, uh, have the conversation NICOLE: I know, **could you hear** it? AMY: NICOLE: How can

Examples with “did”:

you very much. **Did you hear** that? um, RANDY: The problem AUD: CYNTHIA: coyote **didn't hear** them. Because he was lazy, 're racist. **Did you ever hear** this? CORINNA: No. PATRICK: Did fat Yeah, where **did we hear** that. On TV? ROY: It shit, really? I **didn't hear** about it. ALICE: Yeah. Ron

Examples with “can”:

Call me, when I **can hear** you It was funny. That ALLEN: LUCY: you **can still hear** the planes. JOHN: We may good said tiger. **Can I hear** you say that? MANY: That MANY: Yeah. WALT: **Can you hear** me okay? AUD: Yeah. AUD: NICOLE: I know, could you **hear** it? AMY: NICOLE: How **can**

Examples with “about”:

on duty CORINNA: Did you **hear about** that cop in Milwaukee? this. MARILYN: Didn't you **hear about** him? PETE: No. MARILYN: AUD_: I would like to, **hear something about**, how she came shit, really? I didn't **hear about** it. ALICE: Yeah. Ron ROSEMARY: Oh I didn't **hear anything about** it. BETH: ? No,

Examples with “me”:

Call **me**, when I can **hear** you It was funny. That because they want **me to hear** it. TOM_: Okay, TOM_: and- Uh, RICHARD: And you could **hear me**, uh, have the conversation Yes, so that he would **hear me**. LISA: Oh. MARIE: And Alright Mom. SHERI: Did you **hear me** okay? STEVEN: Yeah. SHERI:

- See: let, could, can, if, how

Examples with “let”:

not sure. And **let's see**. Uh what did we get
 this ball here, **let's see** if they bounce as before.
 my production card. **Let's see**. The day before yesterday. I
 This is my only straightedge. **See let's** say that's
 market ROY: Well **let's see**, a three-person salad bowl.

Examples with “could”:

arrow. HI You **couldn't see** it We get to Universal
 PAM: I wish I **could see** you in action. CAROLYN: No.
 JILL: JEFF: I **could** just **see** myself like lecturing to a
 day, you **couldn't** even **see** or anything, from, but look
 glasses on, I **couldn't see** you AUD: What is the

Examples with “can”:

now. This penstock, you **can see** on either side of us
 my personality. as you **can see**, MANY: WALT: I've not
 n't know if you **can see** it or not, but it
 Okay. BRAD: That you **can see**. TAMMY: Okay. BRAD: Okay that
 n't know if everyone **can see** him, over here in this

Examples with “if”:

the phone. SABRINA: Let me **see if** we have enough milk.
 to nine in the evening. **See**, FRED: Unhunh. RICHARD: So **if**
 stand up, and checked to **see if** there were two eggs
 go up to Dillard's, and **see if** they still have those
 'd be, up and rolling. **See if** I can find the

Examples with “how”:

this. It's okay KITTY: **See how** she does that? Check
 ta check on everybody, and **see how** everybody TOM_: You're

balloons, but, I- let's **see how** many of these balloons,
how it connects, did we **see** that down there? I'll
 take that, Ken Let's **see how** many bottles you got.

4.4.4. Physical Verbs Collocates

"Look at" is a common collocate in both, but SBCSAE incorporates "does," which is absent in LINDSEI-BR. LINDSEI-BR sticks to more generic or role-oriented language.

LINDSEI-BR physical verbs collocates examples:

- Look: At, Her, Me, She, Like

Examples with "at":

all of her friends to: **look at** her pictures and none
 is not me this doesn't **look** like me **at** all please
 I'll ask you to **look at** these four pictures <
 I'll ask you to: **look at** these: pictures study these
 stood up and she's **look at** your her= her pictures

Examples with "her":

that he would make **her look** a little bit better maybe
 because **her** hair didn't **look** very good and her face
 bit about that afterwards yeah **look** at **her** face </B
 to all **her** friends saying **look** it's me this famous
 stood up and she's **look** at your **her= her** pictures

Examples with "me":

saying see that is **me look** how beautiful I am <
 is not me this doesn't **look** like **me** at all please
 to all her friends saying **look** it's **me** this famous
 you God they didn't **look** for **me** because I suppose

Examples with “she”:

he managed to make her **look** nothing like **she** actually looks
 he said no you do **look** like that and **she** said
 stood up and **she’s look** at your her= her pictures
 she look what **she’s look** like because she have that:
 was different now was better **look she** was prettiest in the

Examples with “like”:

looks like a didn’t **look like** a fast food but
 it and it doesn’t **look like** their diet it’s
 he managed to make her **look** nothing **like** she actually looks
 is not me this doesn’t **look like** me at all please
 she look what she’s **look like** because she have that:

SBCSAE physical verbs collocates examples:

- Look: At, Look, does

Examples with “at”:

he’s probably gonna **look at** it too. Sabrina, he
 way JENNIFER: I wanna just **look at** my cards here. DAN:
 n’t know HAROLD: Don’t **look at** me. MILES: I was
 see or anything, from, but **look at** your cigarettes. Just sitting
 or, yeah. Take, look at, **look at** the NPH that you

Examples with “look”:

look out the window, and **look** at the bright. and talk.
 my opinion, GAIL: Let’s **look** at Let’s **look** at
 with you? DEB: Sure. JULIE: **Look look**, JACKIE: No. JULIE: Yeah
 TIM: Oh., LEA: Okay. JUDY: **Look** up. **Look** up babe. Wait.
 KRISTIN: Um, or, yeah. Take, **look** at, **look** at the NPH

Examples with “does”:

his face. **Doesn't** it **look** funny? LUCY: Turn this way.

SCOTT: It's reasonable. KAREN: **Look** at this. When **does** this

three. KEVIN: He **doesn't look** that LISA: He's He

measuring.: JAN: Something **doesn't look** right BRETT: whoa, where did

slave, JACKIE: Doesn't it look good? JULIE: it looks wonderful.

4.4.5 Possession Verbs Collocates

Similar structures are observed in both corpora with "do," "to," "n't," and "any," showing consistency in expressing possession or lack thereof. This similarity indicates that basic expressions of possession and absence are universally understood and used by both learner and native speaker groups, regardless of the language proficiency level.

LINDSEI-BR possession verbs collocate examples:

- Have: Any, Hobbies, Do, N't, To

Examples with “any”:

A> ok do you **have any** other hobbies

point and I don't **have any** many contacts in in

and: what else do you **have any** hobbies

A> and do you **have any** hobbies

and write poems do you **have a any** other hobby or

Examples with “hobbies”:

A> ok do you **have any other hobbies**

cinema to dance do you **have any hobbies** do you wanna

A> and do you **have any hobbies** other than cinema

> nice and: do you **have any hobbies**

nice very nice do you **have other interests any hobbies** <

Examples with "do":

subtitles I I **don't have** many idea about that because
 is different I **don't have** this hair so the man :
 because they **don't actually have** black people there so they're=
 A> and: **do** you **have** plans to study design or
 <A> **do** you **have** a job

Examples with "n't":

do that you **don't have** time you do when you
 supermarkets and: we **didn't have** money so: we bought a
 didn't I didn't have an idea how would be
 > I I **don't have** free time no no I
 not married I **don't have** s= children nothing like that

Examples with "to":

to everyone no one were **have** the courage **to** say the:
 lots of countries I could **have** the opportunity **to to** share
 need time to do **to have** all those kids so time
 yeah yeah I have I **have to** venture in this <
 > <A> **to have** the degree double degree it

SBCSAE possession verbs collocates examples:

- Have: Do, To, N't, Any, Would

Examples with "do":

then supposedly, you **don't have** to um , buy any more
don't __ You can't **have** any, WENDY: It's like
 molecules , what temperature **do** you **have** people. Yes, cold, or low
 salad. SHERRY: Um, **do** you **have** ranch? JAMIE: Mhm.
 ratio. Now you always **do have** breakfast. Really. Unless you're

Examples with “to”:

JIM: in order **to** or- **have** business cards. LCL in in made. I'd like **to have** it back. SABRINA: Okay. I be okay, but I'll **have to** think about it He Yeah. So this used **to have** asbestos. Hunh? LARRY: Yeah. SETH: KELLY: DANA: I don't **have to** go to class Like,

Examples with “n't”:

grades, but she didn't **have** the political views. And, uh, well POUND. I don't **have** time right now. I'm NANCY: So we don't **have** anything DANA: We have a remember. SAM: I don't **have** any onion powder. ANGELA: And I didn't have to **have** it, there's no way.

Examples with “any”:

was that, I didn't **have any** ARTS entries, I had, of five years. Do I **have any** questions. AUD: What's fine, then we won't **have any** nightstogether. I said, I She's SHERRY: Does she **have any** already? BETH: Hm-m. **any** of us LORI: Might **have** been Nora EVELYN: Come on.

Examples with “would”:

balance was, we **would** then **have** to, if they had a uh, we **would** love to **have** you all come. There are remains, that I **would** not **have** done the work, for this Well, we **would** like to **have** that as an option. SETH: all who are married, **would have** picked up on it, the

4.4.6 Relationship Verbs Collocates

Both corpora show a preference for "would" and contractions like "'d", but SBCSAE reflects more personal sentiments with additional use of "me" and "kids". This suggests a more familiar, intimate use of language among native speakers, while LINDSEI-BR remains more formal or generalized.

LINDSEI-BR relationship verbs collocates:

- Like: 'd (would), Informally, interview, Hi, Would

Examples with “d”:

A> hi I'd **like** to interview you informally on
 A> hi I'd **like** to interview you informally on
 <A> I'd **like** to interview informally on things
 > so: hi I'd **like** to interview you informally on
 <A> I'd **like** to interview you informally on

Examples with “informally”:

A> Hi: I'd **like** to interview you **informally** on
 A> hi I'd **like** to interview you **informally** about
 A> Hi I'd **like** to interview you **informally** on
 A> hi I'd **like** to interview you **informally** on
 A> hi I'd **like** to interview you **informally** on

Examples with “interview”:

A> hi I'd **like** to **interview** you informally on
 A> hi I'd **like** to **interview** you informally on
 A> hi I'd **like** to **interview** you in informally
 A> hi I'd **like** to **interview** you informally on
 A> hi I'd **like** to **interview** you informally on

Examples with “hi”:

A> **hi** I'd **like** to interview you informally on
 A> **hi** I'd **like** to interview you in informally
 A> **hi** I'd **like** to interview you informally on
 A> **hi** I'd **like** to interview you informally on
 A> **hi** I'd **like** to interview you informally on

Examples with “would”:

my thing but I **would like** to try it more I
 because I I **wouldn't like** to I don't know
 her the way she **would like** to be and she would
 medieval times and I **would like** to to see the evolution
 be: a translator I **would like** to be a translator I

- Love: To

Examples with "to":

but when I do I love to read I love the
 country but I= I would love to do it here also
 A> you have to love your students otherwise you kill
 he teach me how to love rock and bossa nova you
 girls that I used to love but when when the years

SBCSAE relationship verbs collocates:

- Like: 'd, Would, Do, Kids, N't

Examples with "'d":

one I made. I'd **like** to have it back. SABRINA:
 still ask. LANCE: we'd **like** the left, if it becomes
 on one hand I'd **like** to take it back to
 here, you'd have to **like**, you know, pay fifty bucks
 some of it. I'd **like** to, just for a couple

Examples with "would":

in there, JO: I **would like** that. WESS: and you do
 come back. And they **would Like**, I have couple parents, who
 bouncing rubber balls. I **would like** you people in the audience
 Yes. AUD: AUD_: I **would like** to, hear something about, how
 order, you know, I **would like** a case of double-aught

Examples with “do”:

do you like. **Do** you **like** big cities, like a Chicago?

n't. KENDRA: I **don't like** ___ KEVIN: Do you like shrimp

NANCY: Mm. DANA: **Do** you **like** that kind? I love bread.

's what I **don't like** about him. Oh. That was

buildings are, I **don't like** the new buildings. ALICE: Do

Examples with “kids”:

was, like uh did she **like** having **kids**, and she said

ways. Like, **kids** can either **like**, they'll say one word,

LINDA: his **kids** didn't **like** that, JANICE: No, his kids

No, his **kids** didn't **like** it, but Henry was against

Examples with “n't”:

n't, Most people **don't** like big pieces. JO: Oh , Mm-

those. KAREN: You **didn't** like them? SCOTT: Not really. KAREN:

case, RICHARD: Or **didn't** like, FRED: and making your life

Broiling fish. **Don't** you **like** that? You know he had

DIANE: He definitely **wouldn't like** Michael EVELYN: Well, JANICE: That

- Love: love, 'd, Would, Me, do

Examples with “love”:

Love? DARRYL: PAMELA: Could I **love** you? PAMELA: Could I

you, this morning, do you **love** me. Do you really **love**

everything , yes Lord I do **love** you. Well, the saving **love**

um, but uh, I would **love** I would **love** to go.

I would **love** I would **love** to go. LENORE: Yeah. JOANNE:

Examples with “'d”:

welcome to come. We'**d love** to have you there. So

get away JILL: I'**d love** to. You know I was

actually, you know, I'**d love** to do gray water, PETE:

of life. JOANNE: I'**d love** to see it. LENORE: JOANNE:

it. LENORE: JOANNE: I'd **love** to see him chase and

Examples with "would":

week from neighbors. Who **would love** to have their house done, festival. And uh, we **would love** to have you all come.

person someday, **wouldn't they love** to see the photograph of um, but uh, I **would love I would love** to go.

I **would love I would love** to go. LENORE: Yeah. JOANNE:

Examples with "me":

mean JEFF: Don't you **love me**? JILL: Oh. JEFF: Aw.

says to Peter, do you **love me**, more than these? Simon,

son of Jonah, do you **love me**, more than these. Three

in his name. Do you **love me**. You know, Jesus did

love me. Do you really **love me**. But he asks it,

Examples with "do":

to know that I **do love** him. CAM: Mhm. LAJUAN: And

mean JEFF: **Don't you love** me? JILL: Oh. JEFF: Aw.

in his name. **Do you love** me. You know, Jesus did

KELLY: something to **do** Men **love** that stuff. DANA: ARIANNA: My

out here, **don't you love** it? FRAN: I do, I

The results found in this section reveal that the usage of verbs with their collocates varies between the two corpora, reflecting the different contexts or possible dialectal influences within the English language as used by learners and natives. Perhaps this may be the most significant difference found when considering that one corpus is of semi-spontaneous speech, and one is of spontaneous speech.

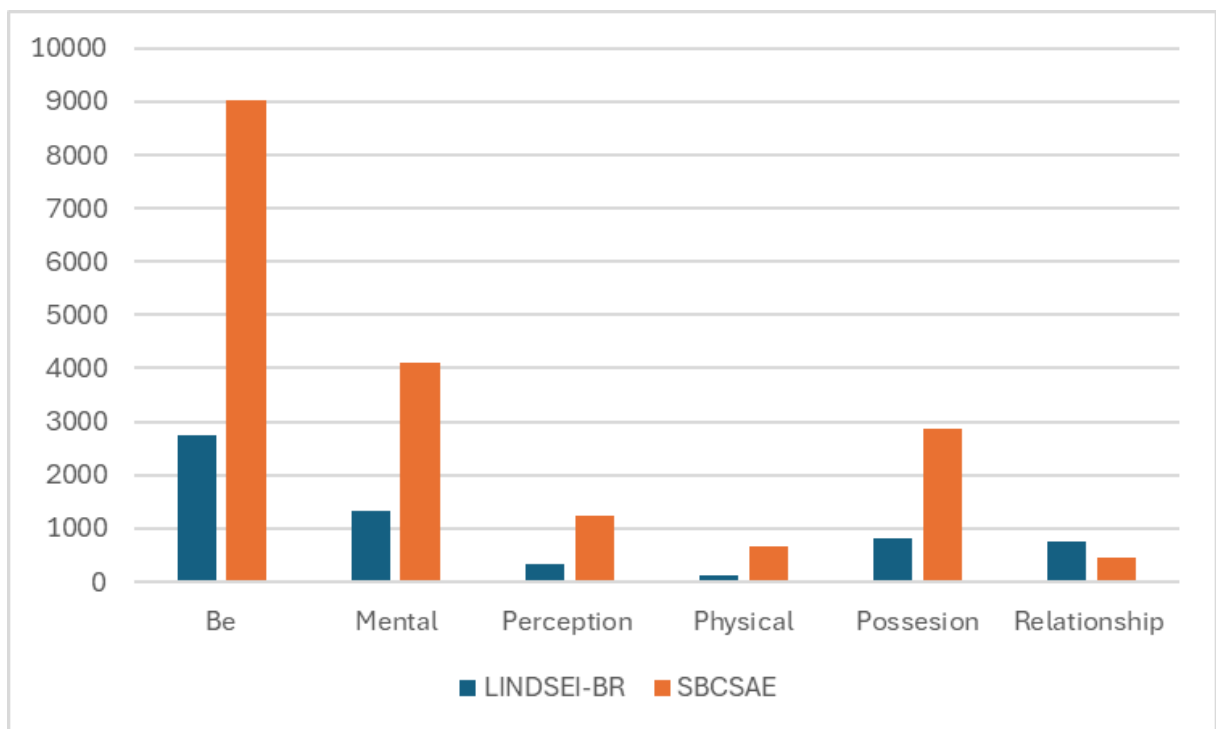
The findings revealed that the learner corpora contained numerous repeated sentences, resulting in fewer collocates. This repetition was attributed to the scripted questions posed to participants during data collection. Consequently, the context of usage was more constrained compared to the native corpus, as the language was not employed in an entirely spontaneous and accessible manner.

These analyses underscore how different contexts (learner versus native) influence linguistic patterns. Learners tend to use a more formal register, possibly influenced by academic settings, while native speakers exhibit a more relaxed and personal use of language in everyday contexts. This distinction offers insights into how language is adapted and used in varying social and cultural environments.

4.5 LINDSEI-BR and SBCSAE final remarks

As shown in the previous sections, the results found in the two corpora are incredibly similar. Graph 4.3 compares all the occurrences for the selected verbs in each semantic category. Both corpora have the verb “be” accountable for most of the occurrences, followed by the mental verbs and possession verbs, respectively.

Graph 4.5: Comparison between LINDSEI-BR and SBCSAE SVs widespread usage



Source: Elaborated by the author, 2024

The other three categories appear in distinct positions, although they are all low-frequency categories. In LINDSEI-BR, relationship verbs were used more than perception and physical verbs, whereas, in SBCSAE, perception verbs were used more than physical and relationship verbs.

When looking inside each semantic category individually without considering the number of times each SV was used, but exclusively focused on the verbs used in the progressive, both corpora had perception verbs with the most verb occurrences in the progressive. They shared the category and the same verbs within it – feel, hear, see, and smell- which were used in the progressive in both corpora. – There were no occurrences of taste in the progressive in both corpora. SBCSAE also had the possession category with four verbs occurring in the progressive (have, hold, keep, and own), compared to two verbs in LINDSEI-BR (have and hold).

The second semantic category with SV occurring in the progressive the most in both corpora is the category of mental verbs, with three verbs used in the progressive. However, different from the perception verbs, the corpora presented different verbs in the progressive. While in LINDSEI-BR, the three verbs are know, think, and understand, SBCSAE presented believe, know, and think being used in the progressive. None of the corpora had the verb prefer being used in the progressive.

The semantic categories of physical and relationship verbs had only two verbs occurring in the progressive in LINDSEI-BR, “exist” and “look” for physical verbs, and like and love for the relationship category. SBCSAE showed that within the physical category, the verbs associate, exist, and look were used in the progressive, and only the verb like was used in the progressive within the relationship category.

Considering the semantic tags associated with the verbs in each corpus, the occurrences of “be” in LINDSEI-BR are primarily associated with the idea of existence (A3+). Additionally, “Be” is used to convey the idea of evaluation (A5.4+) and kinship (S4), indicating relationships between family members. There is also one occurrence where it is tagged as a grammatical term (Z5). In comparison, the tags for “Be” in SBCSAE are similar, primarily associated with existence (A3+) and evaluation (A5.4+). However, unlike LINDSEI-BR, SBCSAE includes tags indicating general/abstract terms relating to allocating/relinquishing/acquiring/receiving (A9+), abstract terms denoting importance/significance (A11.1+), and terms relating to crime/criminal activities and the legal system (G2.1).

The semantic tags for mental verbs in LINDSEI-BR show that "believe" indicates thought and belief (X2.1) and evaluation (A5.4+) and is used as a discourse marker and emphatic term (Z4). The verb "know" is associated with evaluation (A5.4+), knowledge/perception/retrospection (X2.2+), and discourse markers (Z4). "Prefer" indicates fondness/affection/partiality/attachment (E2++) and evaluation (A5.4+). "Think" expresses evaluation (A5.4+), thought, and belief (X2.1) and is used as a discourse marker and emphatic term (Z4). "Understand" is used for evaluation (A5.4+) and understanding/comprehension (X2.5). The findings for SBCSAE are similar, with "believe," "know," "prefer," "think," and "understand" carrying comparable tags. However, SBCSAE includes additional tags for "know," indicating terms relating to crime/criminal activities (G2.1) and importance/significance (A11.1+), and for "think," indicating the position of/point of reference for X (M6).

The perception verbs in LINDSEI-BR show that "feel" is used to express evaluation (A5.4+), reasoning/thinking, level of belief/skepticism (X2.1), and desire/aspiration (X7+). "Hear" is used for evaluation (A5.4+) and sensory terms relating to sound (X3.2). "See" is associated with evaluation (A5.4+), terms related to crime/criminal activities and the legal system (G2.1), sensory terms relating to sight (X3.4+), and discourse markers (Z4). The only occurrence of "smell" is used to express sensory terms relating to smell (X3.5). "Taste" is used to express evaluation (A5.4+) and sensory terms relating to taste (X3.1). In SBCSAE, "feel" carries similar tags but includes an additional tag for permission/consent/authorization (S7.4+). "Hear" and "taste" have the same tags as in LINDSEI-BR. "See" has one difference; the tag for crime/criminal activities (G2.1) is absent in SBCSAE. "Smell" has more occurrences in SBCSAE and includes an additional tag for evaluation (A5.4+).

The semantic tags for physical verbs in LINDSEI-BR show that "appear" indicates abstract terms relating to appearance/impression (A8). "Exist" is associated with abstract terms relating to existence (A3+). "Look" is used for evaluation (A5.4+) and appearance/impression (A8). "Seem" is used similarly for evaluation (A5.4+) and appearance/impression (A8). In SBCSAE, the verb "appear" has the same tag as in LINDSEI-BR. "Associate", which is not found in LINDSEI-BR, is used in SBCSAE for evaluation (A5.4+), similarity/difference (A6.1+), and mental actions/processes (X2.1). "Exist" has the same tag in both corpora. "Look" in SBCSAE includes additional tags for position/point of reference (M6), appearance/look (O4.2+), and commencement/completion (T2). "Seem" is used in the same manner in both corpora.

The possession verbs in LINDSEI-BR show that "belong" indicates general/abstract terms relating to allocating/relinquishing/acquiring/receiving (A9+). "Have" is used for evaluation (A5.4+), causal relationships (A2.2), allocating/relinquishing/acquiring/receiving (A9+), fondness/affection/partiality/attachment (E2+), moral principles (G2.1), and as discourse markers (Z4). "Hold" refers to movements/activities (M2). "Keep" is used for evaluation (A5.4+) and allocating/relinquishing/acquiring/receiving (A9+). In SBCSAE, "belong" has the same tag as in LINDSEI-BR. "Have" shares most tags with LINDSEI-BR but includes additional tags for life and death (L1), importance/significance (A11.1+), and family relationships (S4). "Hold" has the same tag for movements/activities (M2) as in LINDSEI-BR, with additional tags for evaluation (A5.4+) and position/point of reference (M6). "Keep" shares tags with LINDSEI-BR and includes additional tags for openness/concealment (A10) and help/hindrance (S8). "Own" has no occurrences in LINDSEI-BR and is consistently tagged in SBCSAE for allocating/relinquishing/acquiring/receiving (A9+).

The semantic tags for relationship verbs in LINDSEI-BR show that "dislike" is used evaluatively (A5.4+). "Hate" is used for evaluation (A5.4+) and fondness/affection/partiality/attachment (E2+). "Like" indicates evaluation (A5.4+), fondness/affection/partiality/attachment (E2+), appearance/impression (A8), and discourse markers (Z4). "Love" is used for evaluation (A5.4+) and fondness/affection/partiality/attachment (E2+). "Trust" is used for evaluation (A5.4+). In SBCSAE, "dislike" does not occur. "Hate" has the same tags as in LINDSEI-BR. "Like" shares tags for evaluation (A5.4+) and fondness/affection/partiality/attachment (E2+), with an additional tag for importance/significance (A11.1+). Tags for appearance/impression (A8) and discourse markers (Z4) found in LINDSEI-BR are absent in SBCSAE. "Love" has the same tags in both corpora. "Trust" is tagged similarly for evaluation (A5.4+) in both corpora, with SBCSAE including an additional tag for reasoning/thinking and belief/skepticism (X2.1).

Focusing on the collocates found for the base form of the SVs in the corpora, in LINDSEI-BR, the verb "be" collocates with "would," "teach," "to," "want," and "an." In contrast, SBCSAE shows "be" collocating with "would," "gonna," "ll," "to," and "na" (as in "gonna"). The results indicate that "would" and "to" are common in both corpora. However, LINDSEI-BR has "teach" and "want," which are not found in SBCSAE, likely because many participants in LINDSEI-BR were studying to become teachers. In

SBCSAE, colloquial forms such as "gonna" and contractions like "'ll" appear as strong collocates, suggesting that learners in LINDSEI-BR use English more formally than native speakers.

For mental verbs, LINDSEI-BR shows "believe" collocating with "that" and "you," while in SBCSAE, it collocates with "to," "it," "I," "you," and "that." For "know," LINDSEI-BR lists "n't," "do," "I," "if," and "what," whereas SBCSAE includes "do," "n't," "I," "how," and "did." "Prefer" has no collocates in either corpus. "Think" in LINDSEI-BR collocates with "about," "do," "take," "n't," and "much," while in SBCSAE, it collocates with "do," "n't," "would," "about," and "I." "Understand" in LINDSEI-BR has "did," "what," "n't," "do," and "you" as collocates, compared to "why," "n't," "did," "do," and "but" in SBCSAE. Common collocates like "do," "n't," and "about" are found in both corpora for "think" and "know." However, "take" and "much" appear in LINDSEI-BR but not in SBCSAE, which instead has "why" and "but" and a higher use of the personal pronoun "I."

For perception verbs, LINDSEI-BR lists "feel" collocating with "so," "and," and "I," while SBCSAE includes "n't," "like," "did," "do," and "but." "Hear" in LINDSEI-BR collocates with "can," "some," "so," "to," and "was," compared to "could," "did," "can," "about," and "me" in SBCSAE. "See" has "let," "can," "how," "could," and "see" in LINDSEI-BR, while SBCSAE lists "let," "could," "can," "if," and "how." Both corpora show "can" and "could" for "hear" and "see." However, SBCSAE includes negations like "n't" for "feel" and "hear," which are not found in LINDSEI-BR. "I" and "so" in LINDSEI-BR are replaced by more conditional or speculative terms in SBCSAE, like "if" and "about."

For physical verbs, LINDSEI-BR shows "look" collocating with "at," "her," "me," "she," and "like," while SBCSAE has "look" collocating with "at," "look," and "does." "Look at" is common in both corpora, but SBCSAE includes "does," which is not found in LINDSEI-BR.

For possession verbs, "have" in LINDSEI-BR collocates with "any," "hobbies," "do," "n't," and "to." In SBCSAE, "have" collocates with "do," "to," "n't," "any," and "would." Similar structures with "do," "to," "n't," and "any" appear in both corpora, showing consistency in expressing possession or lack thereof.

For relationship verbs, "like" in LINDSEI-BR collocates with "'d" (would), "informally," "hi," "would," and "on." In SBCSAE, "like" collocates with "'d," "would," "do," "kids," and "n't." "Love" in LINDSEI-BR collocates with "to," while in SBCSAE, it

collocates with "love," "d," "would," "me," and "do." Both corpora prefer "would" and contractions like "d". SBCSAE expresses more personal sentiments (e.g., "me," "kids"), while LINDSEI-BR includes more formal or less personal collocates like "informally" and "hi."

When the results of this study are compared to what has been previously reported in the literature about SV, they indicate that every verb can be used in the progressive form, as stated by Sawn (2017) and Biber et al. (2021). It seems that the probability of SVs happening in the progressive form is more connected to the frequency of usage.

The results found here are also aligned with those found by Rautionaho (2020), in which the author states that there is no major difference in SV usage when comparing inner-circle varieties of English to outer-circle varieties of English. When considering the general occurrences, the learners of LINDSEI-BR used such verbs similarly to the native speakers of SBCSAE, both in quantity and the semantic categories found. However, the learners in LINDSEI-BR make slightly less use of SVs in the progressive compared to native speakers. This is somehow linked to what was observed by Falhasari et al. (2012), that indicates that learners with high levels of proficiency tend to stick to the use of grammar rules.

Considering the results found by Martínez-Vázquez (2018) compared to those found here, the verb "love" was expected to occur more frequently in the progressive. However, this did not happen. Among the learners, only one occurrence of "love" in the progressive was found, and no occurrence of this verb in the progressive was found among American natives.

CHAPTER 5

CONCLUSION

Stative verbs (SVs) have been studied over the years using different approaches, and CL has contributed to a better understanding of their behavior. Results in the available literature suggest that SVs usage in the progressive has grown, and this tendency is found among learners and native speakers.

This thesis aims to investigate the usage of SVs by Brazilian learners of English as a second language and compare it with the usage by native American English speakers. The study focused on a corpus-based analysis using two spoken corpora: the Louvain International Database of Spoken English Interlanguage-Brazil (LINDSEI-BR) and the Santa Barbara Corpus of Spoken American English (SBCSAE). The main objectives were to quantify the occurrences of selected SVs, identify the most used semantic categories, analyze the usage of SVs in the progressive form, and compare the findings between the two groups.

The analysis revealed that Brazilian learners and native speakers use SVs with surprising similarities in frequency and context. Among the semantic categories, mental verbs had the highest occurrences, followed by possession verbs. Perception and physical verbs showed the lowest frequency. This distribution suggests that the general frequency of SVs in the language might influence their usage in the progressive form.

Regarding the progressive form, the study found that both groups used SVs in this aspect, albeit with some differences. Native speakers showed a slightly higher tendency to use SVs in the progressive form than Brazilian learners. This finding aligns with previous research suggesting that higher proficiency levels might lead to less flexibility in using SVs, including in the progressive form.

The semantic tagging conducted using the LancsBox X software provided further insights into the context of SV usage. The tags indicated that most occurrences of SVs fell into expected semantic categories such as mental states, perception, possession, and relationships. Notably, verbs like "know," "think," and "believe" frequently appeared with semantic tags related to cognitive processes, while verbs like "have" and "own" were tagged under possession. This categorization helped both groups understand the nuanced use of these verbs in different contexts.

Regarding verb forms, the past form (VBD) and the non-3rd person singular present form (VBP) were the most frequently used among Brazilian learners. The base form (VB) and the third-person singular present form (VBZ) also had significant occurrences. However, the gerund or present participle form (VBG) and the past participle form (VBN) were less common. This pattern was similar for native speakers, indicating a shared tendency in form usage across both groups. The lower frequency of progressive forms among Brazilian learners suggests a more cautious approach to using SVs in less traditional grammatical structures.

The specific objectives of the research were met, as the study successfully quantified the occurrences of SVs, identified the most frequently used semantic categories, compared the SVs' collocates found in both corpora, and analyzed the usage of these verbs in the progressive form. Additionally, the comparison between Brazilian learners and native speakers provided insights into the similarities and differences in SV usage, contributing to the understanding of L2 acquisition.

In revisiting the initial purposes, this research sheds light on acquiring English as a second language, particularly in stative verb usage. It highlights how Brazilian learners, even at advanced proficiency levels, might still adhere closely to the rules presented in grammar books, possibly limiting their flexibility in using SVs in non-traditional forms such as the progressive. On the other hand, native speakers exhibit a more dynamic use of these verbs, reflecting the natural evolution and variability of the English language.

Overall, this study contributes to the field of L2 research by providing empirical data on the use of stative verbs, offering insights that can inform English language teaching and material development. It underscores the importance of exposing learners to authentic language use and encouraging a more flexible approach to grammar that reflects real-world usage. Future research could further explore the factors influencing the use of SVs in progressive forms among different learner populations and proficiency levels, enhancing our understanding of the complexities involved in second language acquisition.

Future research can also address several limitations identified in this study, particularly the interaction between syntax and semantics and its impact on using stative verbs. Additionally, the investigation of ambiguities in SVs usage presents a promising avenue for further inquiry, potentially enhancing the findings of this work. Furthermore, these verbs could be examined through the lens of various grammatical

theories to provide a more comprehensive understanding of their behavior and functions.

REFERENCES

- ADOLPH, S.; KNIGHT, D. Building a spoken corpus. *The Routledge Handbook of corpus linguistics*, p. 38-52, 2010.
- ATKINS, S.; CLEAR, J.; OSTLER, N. Corpus Design Criteria. *Literary and Linguistic Computing*, (7) 1. Oxford: Oxford University Press, 1–16, 1992
- BIBER, D. *Variation across speech and writing*. New York: Cambridge University Press, 1991.
- BIBER, D. Representativeness in corpus design. *Literary and linguistic computing*, v. 8, n. 4, p. 243-257, 1993.
- BIBER, D. et al. *Longman grammar of spoken and written English*. London: Longman, 2021.
- BREZINA, V.; Platt, W. #LancsBox X [software], Lancaster University, Available in: <http://lancsbox.lancs.ac.uk>.
- DAVIS, M. Corpora: an introduction, in BIBER, D.; REPPEN, R. (Eds.). *The Cambridge handbook of English corpus linguistics*. Cambridge University Press, p. 11-32, 2015.
- DU BOIS, J. W.; et al. *Santa Barbara Corpus of Spoken American English*. 2000-2005.
- FALHASIRI, M.; YOUHANAEE, M.; BARATI, H. Second Language Acquisition of Progressive Aspect of Stative and Achievement Verbs in English. *Journal of Language Teaching and Research*, v. 3, n. 5, p. 992, 2012.
- FRANCIS, W. N. Language Corpora B.C. In SVARTIK, J. (ed.), 17–32, 1992
- FREUND, N. Recent change in the use of stative verbs in the progressive form in British English: I'm loving it. *University of Reading Language Studies Working Papers*, v. 7, p. 50-61, 2016

GILQUIN, G.; DE COCK, S.; GRANGER, S. (Ed.). Louvain International Database of Spoken English Interlanguage (LINDSEI). Première édition. Louvain-la-Neuve: Presses Universitaires de Louvain, 2010.

GRANGER, S.; HUNG, J.; PETCH-TYSON, S. Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching. Amsterdam: Benjamins, 2002.

GRANGER, S.; GILQUIN, G.; MEUNIER, F. (eds). The Cambridge handbook of learner corpus research. Cambridge University Press, 2015.

HANUM, Upi L. Forms and Meanings of Stative Verbs in Progressive Tense: a Corpus Linguistics Research. ELS Journal on Interdisciplinary Studies in Humanities, v. 1, n. 2, p. 184-190, 2018.

HUNDT, M. The Use of Progressive Aspect in Late Modern English: A Quantitative Analysis. ICAME Journal, 28, 49-67, 2004.

LINDQUIST, H.; LEVIN, M. Corpus linguistics and the description of English. Edinburgh University Press, 2018.

MARTÍNEZ-VÁZQUEZ, M. I'm loving it! A Corpus-Based Study of the Progress of love. Journal of English Linguistics, v. 46, n. 2, p. 140-166, 2018.

MCENERY, T.; WILSON, A. Corpus Linguistics. Edinburgh: Edinburgh University Press. 1996.

MCENERY, T.; XIAO, R. What corpora can offer in language teaching and learning. Handbook of research in second language teaching and learning. Routledge, p. 364-380, 2011

MCENERY, T.; HARDIE, A. Corpus linguistics: Method, theory, and practice. Cambridge University Press, 2012.

MCENERY T.; BROOKES, G. Building a written corpus: what are the basics?'. In: O'KEEFFE, A.; McCARTHY, M. (eds), *The Routledge Handbook of Corpus Linguistics* (2nd edition). London: Routledge, pp. 35–47, 2022.

MELLO, H. Methodological issues for spontaneous speech corpora compilation the case of C-ORAL-BRASIL. *Spoken Corpora and Linguistic Studies*, in Raso, TOMMASO, R.; MELLO, H. (eds). *Spoken corpora and linguistic studies*. Vol. 61. John Benjamins Publishing Company, p. 27-68, 2014.

MELLO, H. et al. LINDSEI-BR: An Oral English Interlanguage Corpus. In VII GSCP International Conference: Speech and Corpora, 85-86, 2013.

NOTEPAD++ TEAM. Notepad++ [software]. Available at <https://notepad-plus-plus.org/>. 2023.

PEARSON, J. *Terms in Context*. John Benjamins Publishing Co. 1998

RAUTIONAHO P. Revisiting the myth of stative progressives in world Englishes. *World Englishes*. Available at <https://doi.org/10.1111/weng.12520>, p, 1–24, 2020.

ROTHMAYR, A. *The structure of stative verbs*. 2009

SAEED, J. I. *Semantics*. 3. ed. Oxford: Wiley-Blackwell, 2009.

SINCLAIR, J. *Corpus, concordance, collocation*. 1991.

SINCLAIR, J. *Corpus Typology: A Framework for Classification*. EAGLES document, 1-18, 1994. Now published as SINCLAIR, J. *Corpus Typology: A Framework for Classification*. In G. Melchers & B. Warren (Eds.), *Studies in Anglistics*. Stockholm: Almqvist and Wiksell International.16, p. 17-34, 1995.

SINCLAIR, J. *Corpus and Text - Basic Principles'*, in M. Wynne (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*, Oxford: Oxbow Books, pp. 1–16, 2005.

STAPLES, S. Spoken discourse. In: BIBER, D., REPPEN, R. The Cambridge handbook of English corpus linguistics. Cambridge University Press. p. 271 – 291, 2015.

SWAN, M. Practical English Usage. 4. ed. Oxford: Oxford University Press, 2016.

VENDLER, Z. Verbs and Times. The Philosophical Review, Vol. 66, No. 2. pp. 143-160, 1957

APPENDIX A: LINDSEI-BR DATA SET

[\(Available only through electronic means\)](#)

APPENDIX B: SBSCSAE DATA SET

[\(Available only through electronic means\)](#)

APPENDIX C: USAS Semantic Subcategories Labels

[\(Available only through electronic means\)](#)