

A corpus-based study of reporting verbs in academic Portuguese

Marine Laísa Matte – Elisa Marchioro Stumpf
Federal University of Ríó Grande do Sul / Brazil

Abstract – Referring to other sources is a cornerstone in academic writing and one way of framing someone else’s ideas is through reporting verbs. There is little research on this phenomenon in academic Portuguese. Most of these studies analyze reporting practices without focusing on linguistic aspects (Bessa 2011; Hoffnagel 2010), with few studies on reporting verbs (Souza and Mendes 2012). The aim of this paper is to analyze how reporting verbs are used in the *Corpus of Portuguese for Academic Purposes* (CoPEP; Kuhn and Ferreira 2020), a corpus of research articles in Brazilian and European Portuguese. CoPEP was divided into two subcorpora: one with texts related to Hard Science (engineering, exact-earth science, and health science), and another with texts related to Soft Science (applied social science and humanities). *Sketch Engine* (Kilgarriff *et al.* 2014) was used to extract the verbs that are used before and after the lemma *autor* ‘author’. Results indicate that texts in Hard Science have a slightly higher frequency of reporting verbs than texts in Soft Science, but both rely on similar reporting verbs to cite the voice of others. There is preference for the present tense in comparison with past and future, for the active voice in detriment of the passive voice, and for the order ‘author + verb’.

Keywords – reporting verbs; academic Portuguese; citation practices; disciplinary variation

1. INTRODUCTION

In the past two decades, institutions of higher education in Brazil have witnessed a growth in the number of new campuses, courses, and students. This is largely due to public policies such as the *Program for the Restructuring and Expansion of Federal Universities* (REUNI) and affirmative action programs, started by *Law of Social Quota* in 2012. Portugal, in turn, has also registered an increase in the number of international student enrollment in the last 15 years (Oliveira *et al.* 2015). Moreover, there has been an influx of students who may not be used to features of academic discourse: non-traditional entrants in higher education in Brazil and an increasing number of foreign students in Portugal. This calls attention to the need for research into Portuguese for academic purposes, so as to help students face the demands of coursework at university.



In spite of recent severe budget cuts in science and technology,¹ Brazil is one of the top producers of scientific knowledge in Latin America (Kowaltowski *et al.* 2021) with twelve percent of its researchers publishing articles in Portuguese,² together with 3 percent from Portugal (Hernández Bonilla 2021).

The language of publication seems not to be an either-or matter, meaning that scholars may choose to publish in more than one language, considering different purposes, genres, and audiences (Pérez-Llantada 2021). Besides, there is a strong link between languages of publication and disciplinary areas, with scholars from harder and health science speakers of Portuguese as L1 tending to publish their work mostly in English. In a study on language choice in scholarly publication, Solovova *et al.* (2018) analyzed the choice between English and Portuguese in articles from three disciplines (linguistics, information and library sciences, and pharmacology and pharmacy) written between 1998 and 2017. The authors state that

a comparison between Portuguese-written and English-written articles during a 20-year period divided in two decades (1998–2007 and 2008–2017) shows a rise in *both* languages within the Social Sciences and Humanities. Overall figures are substantially higher in English, but *relative* figures indicate the comparatively higher rise in Portuguese articles in the second decade (Solovova *et al.* 2018: 12, authors' italics).

Despite the budget cuts, there is a body of research being published in Portuguese, this meaning that both students and researchers need support in their publishing endeavors. However, research on academic Portuguese is still scarce, corroborating Kuhn's (2017) argument that not only Portuguese is less researched when compared to English and other languages, but also that most of the research tends to focus on text and discourse features, with few lexico-grammatical descriptions.

Furthermore, the few studies on academic Portuguese tend to focus on teaching, although didactic materials and teaching resources are still scarce (Stumpf 2021). This indicates the need for more research on academic Portuguese and shows that reported speech is a relevant feature in this discipline. Among the conventions of academic discourse, successfully integrating quotations, that is, using sources and citing the work

¹ More details about the situation can be found in Kowaltowski *et al.* (2021) and Quintans-Júnior *et al.* (2021).

² Although we acknowledge that Portuguese is a pluricentric language spoken in nine different countries spread over four continents, we bring data related to education and research from Portugal and Brazil, which are countries with a larger number of higher education institutions and journals.

of others is paramount (Coffin *et al.* 2005). It seems clear that academic texts, whether written or oral, rely on external sources to build arguments and link them to certain fields of knowledge whose citation practices can differ substantially. Hence, when incorporating other sources into their own writing, authors reveal their identity, and work towards belonging to specific discourse academic communities (Hoffnagel 2010).

This study presents the initial findings of a larger research on reporting practices in academic Portuguese, more precisely, on the use of reporting verbs in the *Corpus of Portuguese from Academic Journals* (CoPEP; Kuhn and Ferreira 2020). The paper addresses two research questions: 1) What are the reporting verbs that are mostly used to cite the work of others in Hard Science and Soft Science? and 2) To what extent are there differences and similarities in relation to how both registers use reporting verbs?

It is worth mentioning that our motivation to carry out this research was mostly based on the perceived needs of our students, who used a somewhat limited number of structures to report the work of other authors. Our intention was to find different patterns so that our students could expand their repertoire and improve their writing skills by mastering this particularly important feature of academic texts. Charles (2006: 327), in her study of phraseological patterns of citations, highlights the pedagogical applications of such a research and states that bringing the patterns to the classroom is “beneficial in raising student awareness of contextual factors and in enhancing their understanding of what lies behind the language choices evident on the page.”

The corpus, containing 9,900 texts from academic journals in both Brazilian and European Portuguese, was divided into two subcorpora: one accounting for texts related to Hard Science and another for texts related to Soft Science. It should be borne in mind, however, that there is a fine line between what is considered Hard or Soft Science, even more so in the age of interdisciplinary research. Moreover, Soft Science has been considered inferior to Hard Science historically (Smith *et al.* 2000). In addition, aspects such as verifiability, replicability and more methodological rigor have been associated to Hard Science, making Soft Science seem less robust and scientific. For our purposes, however, we consider a traditional classification of the different disciplines as belonging to those areas, similar to the way Kuhn and Ferreira (2020) organized CoPEP. Kuhn and Ferreira (2020) follow the division proposed by the *Coordination for the Improvement of Higher Education Personnel* (CAPES) in Brazil and classify the texts into three main disciplines: College of Life Sciences (Biology,

Agrarian and Health Sciences), College of Humanities (Humanities, Applied Social Sciences and Linguistics, Literature and Arts) and College of Exact Sciences, Technology and Multidisciplinary (Earth and Exact Sciences, Engineering and Multidisciplinary).

The paper is structured as follows. First, we briefly discuss some aspects related to reporting practices and, more specifically, reporting verbs in Portuguese (Section 2). Then we present our methodology (Section 3) and discuss our results, comparing them to other studies of reporting verbs in academic written language (Section 4). We conclude the paper with a summary and some final remarks highlighting limitations and suggestions for follow-up investigations (Section 5).

2. LITERATURE REVIEW

2.1. *Citation practices*

Research on academic language is more commonly carried out in English. Nevertheless, works such as those of Hyland (1999, 2002) can be useful for other languages, such as Portuguese. Considering these aspects, some conventions of academic practices are spread along different languages, as the language itself is a means of communication. Thus, in academic settings, language choices are shaped, among other factors, according to the specificities of particular academic communities that follow certain conventions.

Citing is a common academic practice that helps the writer be part of a research community by creating a rhetorical space (Hoffnagel 2010). According to Hyland (1999: 341), “one of the most important realizations of the research writer’s concern for audience is that of reporting, or reference to prior research,” which, in practice, happens with the use of citations. Swales (1990) argues that it is a way of indicating to which field of knowledge writers belong, as they contribute to the production of knowledge by exploring and explaining specific topics of their area and thus bringing the voice of other authors. Swales’ (1990) taxonomy includes two types of citations: integral citations and non-integral citations. The present paper focuses on integral citations which, according to Thompson (2005: 312) are “placed within the sentence and play an explicit role within the syntax of the sentence.” Hyland and Jiang (2017) show how preference for non-integral forms of citation has increased since the 1960s in four disciplines (applied linguistics, biology, engineering, and sociology), which points to a

phenomenon where importance is given to the facts and contributions from previous work without the focus on the authors.

Another way of classifying citations is by focusing on the reporting verbs. Hyland (1999, 2002) offers a typology that divides them into ‘research (real-world) acts’, ‘cognition acts’ and ‘discourse acts’. Verbs indicating research acts refer to activities and processes that take place in the real world, such as *observe*, *discover*, *analyze*, and *calculate*. Verbs representing cognition acts are those related to mental actions of the researcher, such as *believe*, *assume*, and *view*. Discourse acts are related to the verbal expression of either cognitive or research acts, such as *discuss*, *report*, and *state*. In some cases, however, these categories are not clear-cut and may overlap.

In an analysis of academic texts produced in Portuguese and published in anthropology and psychology Brazilian journals, Hoffnagel (2010) indicates that in integral citations the writer introduces the discourse being cited with the use of reporting verbs, which are one of the various aspects that make up the text. It is worth noting that verb choice is also rhetorical, suggesting that certain verbs are linked to disciplinary practices. Thus, the selection of specific reporting verbs in detriment of others is not random.

In English, this can be clearly seen in Hyland’s (1999) results where writers use more reporting verbs in philosophy than in physics. There is also a prominence of verbs such as *say*, *argue*, *think*, and *suggest* in the humanities, while harder sciences favor *use*, *report*, *describe*, and *show*. Accordingly, Soft Science tends to use more verbs expressing discourse acts, while texts related to engineering and science adopt verbs related to research acts.

2.2. *Reported speech in Portuguese*

A fair amount of research on reported speech and on reporting verbs in Portuguese focuses on journalistic (Corbari and Ramos 2018) or literary registers (Saburi Costa and Freitas 2017), and there has been little research on this phenomenon in academic language. Most studies in academic Portuguese analyze reporting practices more broadly (Hoffnagel 2010; Bessa 2011), with few studies focusing on reporting verbs other than some isolated hints here and there (Souza and Mendes 2012).

Bessa (2011) discusses the use of reporting verbs as a mandatory practice in academia, as an academic piece of writing is only valid when including arguments and theories discussed by other authors. From a dialogical perspective, Bessa (2011) puts forward the idea that following writing manuals on how to cite in academic articles is not enough to master this aspect of academic writing. Using someone else's voice is much more complex than simply reporting their ideas mechanically. As Bessa (2011: 426) argues, there are eight main reasons why writers cite:

- (i) introducing a point of view, (ii) signaling belonging to a framework, a school of thought, (iii) referring to previous works, to trace the state of a problem, (iv) supporting a definition; (v) substantiating an assertion; (vi) discussing an assertion, moving away from a position; (vii) justifying a behavior; and (viii) introducing a new idea.³

As Bessa (2011) aptly notices, understanding citation in academic texts should not be restricted to technical features; it should also encompass an enunciative dimension, for the author's positioning comes into play. As highlighted earlier, this positioning is key to the development of an authorial voice, since authors can tell apart what has been studied by others from what they are doing. Likewise, being able to properly quote the work of others helps frame the author as an insider in the field, whereby they demonstrate their knowledge of references.

In a study dealing with a theoretical and pedagogical reflection about text production in academic settings, Motta-Roth and Hedges (2010) use four academic genres as the basis of their discussion, namely, reviews, research projects, academic articles, and abstracts. Based on Swales's (1990) socio-rhetorical framework, they translate the verbs into Portuguese and analyze these academic genres in terms of organization, structure, and linguistic features in relation to academic practices accepted in academia. It is worth mentioning that although one chapter of the book is centered on different types of citation and verb classification, it is not clear whether the reporting verbs that came up as the result of their analysis are used in integral or non-integral citations. Nevertheless, Motta-Roth and Hedges' (2010) results can serve as a possible framework to meet the goals of our own study.

³ Our translation. Original version: "(i) introduzir um ponto de vista; (ii) marcar o pertencimento a uma corrente, a uma escola; (iii) referir-se a trabalhos anteriores, para traçar o estado de uma problemática, (iv) sustentar uma definição; (v) fundamentar uma afirmação; (vi) discutir uma afirmação, se afastar de uma posição; (vii) justificar um comportamento; e (viii) introduzir uma ideia nova." (Bessa 2011: 426).

Table 1 presents the verbs that are frequently used in the subjects that we consider part of Hard Science and Soft Science, respectively.

Hard Science		Soft Science	
Biology, physics, electrical engineering, mechanical engineering, epidemiology, nursing and medicine		Marketing, applied linguistics, psychology, sociology, education, philosophy	
1	<i>Descreever</i> ‘describe’	1	<i>Sugerir</i> ‘suggest’
2	<i>Desenvolver</i> ‘develop’	2	<i>Descobrir</i> ‘discover/to find out’
3	<i>Propor</i> ‘propose’	3	<i>Argumentar</i> ‘argue’
4	<i>Descobrir</i> ‘discover/find out’	4	<i>Dizer</i> ‘say’
5	<i>Mostrar</i> ‘show’	5	<i>Mostrar</i> ‘show’
6	<i>Reportar</i> ‘report’	6	<i>Descreever</i> ‘describe’
7	<i>Usar</i> ‘use’	7	<i>Notar</i> ‘notice’
8	<i>Sugerir</i> ‘suggest’	8	<i>Explicar</i> ‘explain’
9	<i>Estudar</i> ‘study’	9	<i>Reportar</i> ‘report’
10	<i>Demonstrar</i> ‘demonstrate’	10	<i>Alegar</i> ‘claim’
11	<i>Discutir</i> ‘discuss’	11	<i>Propor</i> ‘propose’
12	<i>Identificar</i> ‘identify’	12	<i>Demonstrar</i> ‘demonstrate’
13	<i>Observar</i> ‘observe’	13	<i>Analisar</i> ‘analyze’
14	<i>Expandir</i> ‘expand’	14	<i>Destacar</i> ‘highlight’
15	<i>Publicar</i> ‘publish’	15	<i>Enfocar</i> ‘focus’
16	<i>Dar</i> ‘give’	16	<i>Discutir</i> ‘discuss’
17	<i>Examinar</i> ‘examine’	17	<i>Fornecer</i> ‘provide’
18	<i>Indicar</i> ‘indicate/point out’	18	<i>Pensar</i> ‘think’

Table 1: Verbs used in Hard and Soft Science (adapted from Motta-Roth and Hedges 2010: 99)

Hoffnagel (2010) analyzed citations in 16 articles dealing with psychology, with 1,292 citations, and 16 articles dealing with anthropology, with 1,025 citations. According to Hoffnagel (2010), there is an enormous variety of reporting verbs in both genres: 135 verbs in anthropology and 90 verbs in psychology. It is worth mentioning, however, that around 50 percent of these verbs were used only once in the corpus. The top five verbs used in texts dealing with anthropology are: *dizer* ‘say’, *afirmar* ‘claim’, *citar* ‘quote’, *apontar* ‘point out’, and *mostrar* ‘show’, while in psychology they are: *realizar* ‘make/do’, *observar* ‘observe’, *propor* ‘propose’, *sugerir* ‘suggest’, and *apontar* ‘point out’.

While this literature highlights the importance of the rhetorical function of citations and the choice of reporting verbs, it must be said that our work focuses on a single type of citation in order to find patterns distributed across the two large areas of Hard Science and Soft Science. Thus, due to the number of excerpts, we decided to explore the forms that were found in the corpora and relate them to what has been already published in the field by focusing on the aspects that match the purposes of our study.

3. DATA AND METHODS

3.1. The corpus

The corpus used in this investigation is CoPEP⁴ (Kuhn and Ferreira 2020), which contains 9,900 texts from academic journals balanced in both Brazilian and European Portuguese. These academic journals are all indexed in the *Scientific Electronic Library On-line* (SciELO).⁵ In order to meet the goals of the study, CoPEP was divided into two subcorpora, one subcorpus (Hard Science) containing texts from engineering, exact-earth science, and health science, and another subcorpus (Soft Science) with texts from applied social science and humanities. Table 2 provides information on the number of tokens and texts in both subcorpora.

	Words	Number of texts	Average number of words per text
Soft Science	25,744,456	4,636	5,553.2
Hard Science	14,678,555	5,264	2,788.48
Total	40,423,011	9,900	8,3411.68

Table 2: Structure of CoPEP

3.2. Methodological procedures

In order to answer our research questions, four main steps were undertaken in both the Soft Science and Hard Science subcorpora. First, we determined that our analysis would be based on the verbs that go together with the lemma *autor* ‘author’, as the focus of this investigation is on reporting verbs and how external author’s ideas are framed. Thus, as illustrated in Figures 1 and 2, we have used the *Word Sketch* tool in *Sketch Engine* (Kilgarriff *et al.* 2014) to generate the list of verbs that collocate before and after *autor* ‘author’ in both subcorpora.

⁴ CoPEP is available on *Sketch Engine* and is balanced in terms of fields of knowledge and language variety, since it includes texts published in Brazilian and European Portuguese. For more information regarding the corpus metadata and compilation, please, refer to Kuhn and Ferreira (2020).

⁵ <https://scielo.org/es/>

autor + verbo			verbo + autor		
ser	258	4.0 ...	ter	90	4.4 ...
ter	244	5.7 ...	haver	39	4.8 ...
considerar	231	8.7 ...	levar	29	6.1 ...
concluir	223	9.6 ...	referir	17	6.4 ...
referir	199	8.2 ...	existir	11	4.2 ...
apresentar	189	6.9 ...	corroborar	10	7.4 ...
sugerir	156	8.5 ...	sugerir	10	6.1 ...
defender	142	8.9 ...	dizer	10	4.7 ...
declarar	131	9.1 ...	afirmar	9	6.3 ...
afirmar	130	8.4 ...	citar	7	6.6 ...
observar	118	8.1 ...	diferir	7	6.4 ...
poder	114	4.6 ...	variar	7	5.5 ...
encontrar	113	7.1 ...	agradecer	6	7.7 ...
apontar	110	8.0 ...	concordar	6	6.8 ...
descrever	109	8.4 ...	identificar	6	4.3 ...
verificar	106	7.7 ...	acrescentar	5	6.5 ...
ir	95	3.8 ...	destacar	5	5.2 ...
propor	83	8.0 ...	conduzir	5	5.0 ...
recomendar	80	8.5 ...	envolver	5	4.1 ...
relatar	75	8.2 ...	partir	5	2.7 ...
destacar	71	7.7 ...	fazer	5	2.4 ...

Figure 1: *Word Sketch* results for the most frequent verbs in Hard Science

autor + verbo			verbo + autor		
ser	968	5.9 ...	ter	132	5.0 ...
ter	530	6.7 ...	levar	100	7.8 ...
considerar	300	8.5 ...	dizer	70	7.3 ...
defender	283	9.0 ...	afirmar	59	8.6 ...
afirmar	281	8.8 ...	referir	58	7.9 ...
referir	273	8.2 ...	haver	51	5.1 ...
ir	268	5.2 ...	fazer	36	5.2 ...
fazer	251	7.5 ...	entender	33	7.6 ...
poder	240	5.6 ...	permitir	33	6.3 ...
concluir	227	8.8 ...	defender	31	7.6 ...
apresentar	209	6.9 ...	sublinhar	25	8.1 ...
estar	185	5.8 ...	apontar	24	6.5 ...
propor	176	8.3 ...	argumentar	21	8.2 ...
procurar	167	7.9 ...	mostrar	20	5.8 ...
sugerir	160	7.9 ...	salientar	19	7.4 ...
analisar	157	8.2 ...	partir	18	4.5 ...
apontar	153	7.9 ...	apresentar	17	3.7 ...
destacar	151	8.0 ...	conduzir	16	6.3 ...
chamar	140	8.0 ...	encontrar	16	5.2 ...
mostrar	136	7.1 ...	assinalar	15	7.4 ...
argumentar	133	8.2 ...	concluir	15	7.3 ...

Figure 2: *Word Sketch* results for the most frequent verbs in Soft Science

Next, we searched for the top 15 verbs that collocate with *autor* ‘author’ by using *Corpus Query Language* (CQL) in order to have access to the concordance lines of these verbs combined with *autor* (‘author’) in a 5-word window. The following CQL queries were used (cf. Table 3).⁶

Hard Science	Soft Science
Query 1: [lemma="autor"] [] {0,5} [lemma="considerar concluir referir apresentar sugerir defender afirmar observar encontrar apontar descrever verificar propor recomendar relatar"]	Query 1: [lemma="autor"] [] {0,5} [lemma="considerar defender afirmar referir fazer concluir apresentar propor procurar sugerir analisar apontar destacar chamar mostrar"]
Query 2: [lemma="referir corroborar sugerir dizer afirmar citar diferir concordar identificar acrescentar destacar conduzir fazer"] [] {0,5} [lemma="autor"]	Query 2: [lemma="ter dizer afirmar referir fazer entender permitir defender sublinhar apontar argumentar mostrar salientar apresentar conduzir"] [] {0,5} [lemma="autor"]

Table 3: *Corpus Query Language*

Based on these two steps, we realized that the verbs *ter* ‘have’, *ser* ‘be’, *haver* ‘there is/there are’, and *fazer* ‘do/make’ presented interesting behaviors. Hence, we decided to run a new CQL search and analyze them separately, as they can be used as auxiliary verbs for compound tenses and on verb phrases. We decided to consider valid cases where *ter*, *ser*, and *haver* were used as auxiliary verbs (and not as the main verb) and *fazer* was used as the main verb to indicate something that was done by the author(s), excluding idioms and cases like those illustrated in examples (1) to (3).

- (1) Ao refletir sobre a dinâmica regional da economia brasileira, diferentes autores **fazem** uso de importantes ressalvas para pensar o processo de desconcentração produtiva verificado a partir da Região. ‘Reflecting on the regional dynamic of the Brazilian economy, different authors **make** use of important caveats to think about the process of deconcentration verified from the region’.
- (2) Este autor não **fazia** parte do seletto grupo dos intelectuais vinculados à academia. ‘This author did not **make** part of a select group of intellectuals linked to academia / This author was not part of a select group of intellectuals linked to academia’.

⁶ Translation of the verbs to English:

Hard Science query 1: ‘consider’, ‘conclude’, ‘refer’, ‘present’, ‘suggest’, ‘defend’, ‘state’, ‘observe’, ‘find’, ‘point out’, ‘describe’, ‘verify’, ‘propose’, ‘recommend’, ‘report’.

Hard Science query 2: ‘refer’, ‘corroborate’, ‘suggest’, ‘say’, ‘state’, ‘cite’, ‘differ’, ‘agree’, ‘identify’, ‘add’, ‘highlight’, ‘conduct’, ‘do/make’.

Soft Science query 1: ‘consider’, ‘defend’, ‘state’, ‘refer’, ‘do/make’, ‘conclude’, ‘present’, ‘propose’, ‘intend’, ‘suggest’, ‘analyze’, ‘point out’, ‘highlight’, ‘call, to show’.

Soft Science query 2: ‘have’, ‘say’, ‘state’, ‘refer’, ‘do/make’, ‘understand’, ‘allow’, ‘defend’, ‘underline’, ‘point out’, ‘argue’, ‘show’, ‘stress’, ‘present’, ‘conduct’.

- (3) Este autor **tinha** como objetivo desenvolver um site onde os próprios usuários poderiam gerar conteúdo. ‘This author **had** as objective developing a site where the users could generate content’.

Other verbs that were initially excluded are *declarar* ‘declare’ and *agradecer* ‘thank’, since all of them occurred in formulaic expressions as in, for instance, *the author(s) declare(s) that there is no conflict of interest* and *we thank the editor and two anonymous reviewers*. The verb *levar* ‘lead, take’ was also excluded since it was mainly used in sentences such as *this led authors to state/defend*, and we focused on verbs coming afterwards. Finally, *existir* ‘exist’ and *partir* ‘leave’ were also excluded since they do not function as reporting verbs.

Valid occurrences were then classified according to: 1) the discipline in which they occurred, 2) voice (passive or active), 3) number (singular or plural of ‘author’), 4) order (‘verb + author’ or ‘author + verb’), 5) tense, aspect and mood or non-finite verb forms (converb, past participle, or infinitive). Since we aimed at finding patterns of use, we also classified the reporting verb according to Hyland’s (1999, 2002) typology, which considers the type of activity that the verbs refer to. Besides, the classification of the verbs was partly based on Shaw (1992) and Hyland and Jiang (2017), whose studies account for the tense and the aspect of verbs. In cases where two valid verbs were used, the sentence was classified twice, once for each verb, as in (4), below. Likewise, verb phrases denoting time or modality were classified according to the first verb, as in (5)–(6).

- (4) A autora **citando** Bourdieu (1983) **afirma** (*citando* ‘citing’): converb, *afirma* ‘states’: 3rd person singular simple present
 ‘The author, **citing** Bourdieu (1983), **states** (...)’
- (5) A autora **continua referindo** (*continua* ‘continues’): 3rd person singular simple present
 ‘The author **continues referring** (...)’
- (6) Os autores **deverão referir-se** (*deverão* ‘should’: 3rd person plural future simple)
 ‘The authors **should refer themselves** (...)’

4. RESULTS AND DISCUSSION

In this section, we first present the overall frequency of reporting verbs distributed in the corpus. Normed counts per 10,000 words are presented between brackets. As stated earlier, the aim is to find the most frequent reporting verbs in the registers related to Hard and Soft Science, together with their patterns of use, and to discuss possible differences and similarities between the registers. Thus, the quantitative results shown in the tables are followed by the discussion of the data. When possible, we refer to other studies conducted in English to try to support and provide the motivation for our findings.

As Table 4 shows, 6,103 occurrences of reporting verbs used before or after the word *autor* ‘author’ in a five-word window were valid. Of these, 3,716 (normed frequency 1.44) are attested in the texts related to Soft Science and 2,387 (normed frequency 1.62) in the texts related to Hard Science. These frequencies indicate a higher frequency of reporting verbs with *autor* ‘author’ in Hard Science. Although “softer disciplines tend to employ more citations” (Hyland 1999: 346), the differences in the normed counts could be partially explained by the preference in Soft Science to use proper names to refer to authors, instead of using the lemma ‘author’ in a more general way.

Excerpts with reporting verbs	
Soft Science	3,716 (1.44)
Hard Science	2,387 (1.62)
Total	6,103

Table 4: Frequency of excerpts with reporting verbs in each register

Table 5, below, provides information on voice, order, and number of reporting verbs in each subcorpus. The data include raw frequencies and the percentages between brackets. Overall, there is a clear preference for the active voice and the order *autor* + *verbo* ‘author + verb’ in both registers, with 87 percent of the occurrences in the active voice in Soft Science and 90 percent in Hard Science. As regards percentages, the order *autor* + *verbo* ‘author + verb’ is more frequent in both registers (80% of the cases in Soft Science and 90% of the cases in Hard Science) when compared to the order *verbo* + *autor* ‘verb + author’. Soft Science exhibits a slightly higher variation in terms of order, showing more excerpts with *verbo* + *autor* ‘verb + author’ order, as shown in Table 5. Finally, when it comes to number, in Soft Science, the choice for the singular

or the plural is balanced: 53 percent of the cases are attested in the singular form and 47 percent in the plural form. By contrast, in Hard Science, the preference is for the plural form (84% of instances) to the detriment of the singular form (16% of instances). The preference for the plural form in Hard Science might be related to the fact that in this register publications with multiple of authors are common.

	Excerpts with reporting verbs	Active voice ⁷	Passive voice	<i>Autor</i> + verb order	Verb + <i>autor</i> order	Singular form	Plural form
Soft Science	3,716	3,261 (87%)	92 (2.4%)	2,992 (80%)	725 (20%)	2,003 (53%)	1,714 (47%)
Hard Science	2,387	2,156 (90%)	48 (2%)	2,162 (90%)	226 (10%)	392 (16%)	1,995 (84%)
Total	6,103	5,417	140	5,154	951	2,395	3,709

Table 5: Number of excerpts with reporting verbs by voice, order, and number

The results on tense, aspect, and mood are shown in Table 6.

		Hard Science	Soft Science
Pretérito 'past'	<i>Pretérito perfeito</i> (preterite perfect)	699 (0.48)	370 (0.14)
	<i>Pretérito imperfeito</i> (preterite imperfect)	10 (0.01)	40 (0.02)
	<i>Pretérito imperfeito contínuo</i> (preterite imperfect continuous)	0 (0)	2 (0.001)
	<i>Pretérito mais-que-perfeito</i> (past perfect)	3 (0.002)	5 (0.002)
	<i>Pretérito mais-que-perfeito composto</i> (compound past perfect)	0 (0)	6 (0.002)
	<i>Pretérito perfeito composto</i> (compound preterite perfect)	54 (0.04)	93 (0.04)
		766 (0.52)	516 (0.2)
Presente 'present'	<i>Presente simples</i> (simple present)	1,428 (0.97)	2,817 (1.09)
	<i>Presente contínuo</i> (present continuous)	2 (0.0013)	2 (0.001)
		1,430 (0.97)	2,819 (1.09)
Futuro 'future'	<i>Futuro perifrástico</i> (compound future)	0 (0)	6 (0.002)
	<i>Futuro perifrástico contínuo</i> (compound future continuous)	0 (0)	1 (0.0004)
	<i>Futuro do presente</i> (simple future)	2 (0.001)	4 (0.002)
	<i>Futuro do pretérito</i> (conditional tense - would)	2 (0.001)	2 (0.001)
		4 (0.003)	13 (0.005)

Table 6: List of tenses, moods, and aspects with raw and normed frequencies

Notably, the present is the preferred tense. Hard Science makes a more frequent use of the past tense (0.52), whereas Soft Science makes a more frequent use of the present (1.09), and the future is rarely used in both registers, but slightly more frequent in Soft Science (0.005) than in Hard Science (0.003). Some examples are provided in (7)–(12).

(7) (...) a grande maioria dos autores **não observou** (*observou* ‘did not observe’: 3rd person singular preterite perfect). Hard Science.

‘(...) the vast majority of authors **did not observe** (...)’

⁷ The sum of the occurrences in the active and the passive voice does not match the number of excerpts with reporting verbs in each subcorpus because occurrences of non-finite verb forms cannot be considered either active or passive voice.

- (8) O autor **afirmava** (*afirmava* ‘stated’: 3rd person singular preterite imperfect). Hard Science.
‘The author **stated** (...)’
- (9) O autor do texto **estava fazendo** (*estava fazendo* ‘was doing’: 3rd person singular preterite imperfect continuous). Soft Science.
‘The author of the text **was doing** (...)’
- (10) O *autor* **apontara** (*apontara* ‘had pointed out’: 3rd person singular past perfect). Soft Science.
‘The author **had pointed out** (...)’
- (11) O autor **havia afirmado** (*havia afirmado* ‘had stated’: 3rd person singular compound past perfect). Soft Science.
‘The author **had stated** (...)’
- (12) (...) autoras/es feministas **têm feito** (*têm feito* ‘have done’: 3rd person plural compound preterite perfect)
‘(...) the feminist authors **have done** (...)’

When further looking at the distinct forms these three tenses may have, there is a stronger preference for the *pretérito perfeito* ‘preterite perfect’ in Hard Science (0.48) than in Soft Science (0.14). Despite these differences, it is possible to observe that, although *pretérito imperfeito contínuo* (‘preterite imperfect continuous’) and *pretérito mais-que-perfeito composto* (‘compound past perfect’) are rare tenses of reporting verbs in Soft Science (0.001 and 0.002, respectively), they are not used in Hard Science. As for the preterit perfect compound, indicating an action that started in the past but is still ongoing (akin to the present perfect tense in English), there are 54 occurrences in Hard Science and 93 in Soft Science, making it the second most frequent past tense in the corpus. According to Hyland and Jiang (2017), there has been an increasing trend in using reporting verbs in the present tense in sociology and engineering alike (which would belong to our Soft and Hard Science subcorpora, respectively), followed by the past tense. However, since Hyland and Jiang (2017) analyze four disciplines, each of them belonging to different registers, their data is difficult to compare with ours, mainly because the languages under analysis (English and Portuguese) have their own peculiarities.

In the present tense, the simple (cf. (13)) and continuous (cf. (14)) aspects are used, the latter with an extremely low frequency. *Presente simples* ‘simple present’, on the other hand, is a very common verb tense when citing the voice of others and incorporating sources. In the excerpts with reporting verbs, this tense is higher in Soft Science (normed frequency 1.09) than in Hard Science (normed frequency 0.97).

(13) (...) autor australiano **destaca** (*destaca* ‘highlights’) 3rd person singular simple present)

‘(...) the Australian author **highlights** (...)’

(14) (...) autor **está defendendo** (*está defendendo* ‘is defending’) 3rd person singular present continuous)

‘(...) the author **is defending** (...)’

Finally, there are few occurrences of future tense in both subcorpora, besides the fact that the comparison between the areas is balanced. Some of these occurrences are shown in (15)–(18) below.

(15) (...) autor **vai chamar** (*vai chamar* ‘will call’: 3rd person singular compound future). Soft Science

‘(...) the author **will call** (...)’

(16) (...) autor **vai defendendo** (*vai defendendo* ‘keeps defending’: 3rd person singular compound future continuous). Soft Science

‘(...) the author **keeps defending** (...)’

(17) (...) autores que **apresentarão** (*apresentarão* ‘will present’: 3rd person plural simple future). Hard Science

‘(...) the authors that **will present** (...)’

(18) (...) autores **defenderiam** (*defenderiam* ‘would defend’: 3rd person plural conditional tense). Soft Science.

‘(...) the authors **would defend** (...)’

The examples provided above are all in the indicative mood. Occurrences in the subjunctive mood, as in (19) and (20) are rarely attested in the corpus (65 overall), even if the subjunctive is required after some subordinating conjunctions in Portuguese, such as *embora* ‘although’. Most cases are in the present tense, followed by instances in the compound preterite perfect.

- (19) (...) autores **defendam** (*defendam* ‘defend’: 3rd person plural simple subjunctive). Hard Science
‘(...) authors **defend** (...)’
- (20) (...) autores **tenham apresentado** (*tenham apresentado* ‘had presented defend’: 3rd person singular compound preterit perfect subjunctive conditional tense). Soft Science.
‘(...) authors **had presented** (...)’

Another interesting finding in the analysis is the use of modalization. Modalization is mainly used when the writer makes a stand towards the voice of the author being cited, instead of bringing the voice of the other in a more impartial way, as in (21)–(22).

- (21) (...) autor parece **defender**. Soft Science.
‘(...) the author **seems to defend** (...)’
- (22) (...) autor **precisa considerer**. Soft Science.
‘(...) the author **needs to consider** (...)’

We also coded non-finite verb forms that appeared as dependent clauses in complex sentences with reporting verbs, if the verb was a verb in our list (cf. Table 1). As shown in Table 7, among these verb forms, the past participle (cf. (23)) is the one with the highest frequency in the corpus, with 333 occurrences of which 236 were attested in Soft Science and 97 in Hard Science. There are 135 occurrences of the infinitive (cf. (24)), evenly distributed in both subcorpora (0.03). As for converb forms, which are non-finite verb forms used in adverbial subordination (cf. (25)), these are also evenly distributed (0.01). The compound converb (cf. (26)), however, is slightly more frequently attested in Hard Science (0.005) than in Soft Science (0.002).

	Past participle	Infinitive	Converb	Compound converb
Soft Science (364)	236 (0.09)	85 (0.03)	36 (0.01)	7 (0.002)
Hard Science (184)	97 (0.06)	50 (0.03)	29 (0.01)	8 (0.005)
Total	333	135	65	15

Table 7: Number of non-finite verb forms in Soft and Hard Science

- (23) (...) no caso **apresentado** pelos autores (*apresentado* ‘presented’: past participle). Hard Science.
‘(...) in the case **presented** by the authors (...)’

(24) (...) palavras ou expressões utilizadas pelos autores para **descrever** (*descrever* ‘describe’: infinitive). Hard Science.

‘(...) words or expressions utilized by the authors to **describe** (...)’

(25) (...) Alguns autores, **observando** a formação do enfermeiro (*observando* ‘observing’: converb). Hard Science.

‘(...) Some authors, **observing** the education of the nurse (...)’

(26) (...) não **tendo** os autores **encontrado** (*tendo encontrado* ‘having found’: compound converb). Hard Science.

‘(...) not **having** authors **found** (...)’

Tables 8 and 9 provide information regarding the most frequent verbs with the lemma *autor* ‘author’. In total, 27 different reporting verbs make up 2,387 occurrences in Hard Science. Out of these, 2,196 in the *autor + verbo* ‘author + verb’ order, and 191 in the *verbo + autor* ‘verb + author’ order. Concerning Soft Science, 26 different reporting verbs make up 3,716 occurrences. Out of these, 2,686 are used in the *autor + verbo* ‘author + verb’ order, and 1,030 are used in the *verbo + autor* ‘verb + author’ order.

		Hard Science <i>Autor + verbo</i>	Hard Science <i>Verbo + autor</i>	TOTAL
1	<i>Referir</i> ‘refer’	185	83	268
2	<i>Considerar</i> ‘consider’	248	2	250
3	<i>Concluir</i> ‘conclude’	235	2	237
4	<i>Sugerir</i> ‘suggest’	164	40	204
5	<i>Apresentar</i> ‘present’	169	1	170
6	<i>Defender</i> ‘defend’	158	0	158
7	<i>Afirmar</i> ‘state’	139	11	150
8	<i>Observar</i> ‘observe’	127	1	128
9	<i>Encontrar</i> ‘find’	119	2	121
10	<i>Verificar</i> ‘verify’	110	3	113
11	<i>Apontar</i> ‘point out’	110	2	112
12	<i>Descrever</i> ‘describe’	104	1	105
13	<i>Recomendar</i> ‘recommend’	93	1	94
14	<i>Propor</i> ‘propose’	83	0	83
15	<i>Relatar</i> ‘report’	79	0	79
16	<i>Referir-se</i> ‘refer oneself/themselves’	26	0	26
17	<i>Citar</i> ‘quote’	22	1	23
18	<i>Fazer</i> ‘do/to make’	14	0	14
19	<i>Destacar</i> ‘highlight’	1	9	10
20	<i>Acrescentar</i> ‘add’	0	9	9
21	<i>Corroborar</i> ‘corroborate’	0	8	8
22	<i>Dizer</i> ‘say’	0	8	8
23	<i>Identificar</i> ‘identify’	0	7	7
24	<i>Propor-se</i> ‘propose oneself/themselves’	7	0	7
25	<i>Contribuir</i> ‘contribute’	1	0	1
26	<i>Apresentar-se</i> ‘introduce oneself/themselves’	1	0	1
27	<i>Reportar</i> ‘report’	1	0	1
	Total	2,196	191	2,387

Table 8: Frequencies according to order in Hard Science

	Soft Science <i>Autor + verbo</i>	Hard Science <i>Verbo + autor</i>	TOTAL	
1	<i>Defender</i> ‘defend’	313	105	418
2	<i>Afirmar</i> ‘state’	285	75	360
3	<i>Fazer</i> ‘do/make’	254	76	330
4	<i>Referir</i> ‘refer’	220	96	316
5	<i>Considerar</i> ‘consider’	0	305	305
6	<i>Apresentar</i> ‘present’	174	88	262
7	<i>Concluir</i> ‘conclude’	245	0	245
8	<i>Apontar</i> ‘point out’	165	71	236
9	<i>Analisar</i> ‘analyze’	181	0	181
10	<i>Sugerir</i> ‘suggest’	161	0	161
11	<i>Mostrar</i> ‘show’	136	24	160
12	<i>Propor</i> ‘propose’	150	0	150
13	<i>Destacar</i> ‘highlight’	146	0	146
14	<i>Chamar</i> ‘call’	139	0	139
15	<i>Dizer</i> ‘say’	0	75	75
16	<i>Referir-se</i> ‘refer oneself/themselves’	62	0	62
17	<i>Sublinhar</i> ‘underline’	0	40	40
18	<i>Salientar</i> ‘stress’	0	33	33
19	<i>Propor-se</i> ‘propose oneself/themselves’	31	0	31
20	<i>Argumentar</i> ‘argue’	0	29	29
21	<i>Entender</i> ‘understand’	0	13	13
22	<i>Destacar-se</i> ‘stand out’	8	0	8
23	<i>Mostrar-se</i> ‘show oneself/themselves’	8	0	8
24	<i>Apresentar-se</i> ‘introduce oneself/themselves’	5	0	5
25	<i>Encontrar</i> ‘find’	2	0	2
26	<i>Defender-se</i> ‘defend oneself/themselves’	1	0	1
Total	2,686	1,030	3,716	

Table 9: Frequencies according to order in Soft Science

When it comes to the order, some verbs tend to occur in the *verbo + autor* ‘verb + author’ order and are hardly attested in examples of *autor + verbo* ‘author + verb’. In Hard Science, this is the case for *destacar* ‘highlight’, *acrescentar* ‘add’, *corroborar* ‘corroborate’, *dizer* ‘say’, and *identificar* ‘identify’, most of them being verbs used to express discourse acts. In Soft Science, apart from *considerar* ‘consider’ and *entender* ‘understand’ (classified as verbs indicating cognition acts), the other verbs that are only used in the ‘verb + author’ order are *dizer* ‘say’, *sublinhar* ‘underline’, *salientar* ‘stress’, and *argumentar* ‘argue’ are also all expressing discourse acts.

The five most frequent reporting verbs are different in Hard and Soft Science. *Referir* ‘refer’, *considerar* ‘consider’, *concluir* ‘conclude’, *sugerir* ‘suggest’, and *apresentar* ‘present’ are the five most frequent reporting verbs in Hard Science (Table 8), whereas *defender* ‘defend’, *afirmar* ‘claim’, *fazer* ‘do/make’, *referir* ‘refer’, and *considerar* ‘consider’ are the five most frequent reporting verbs in Soft Science, as shown in Table 9.

Considering the five most frequent verbs in Soft Science, it is not surprising that *defender* ‘defend’ is the most frequent verb, since it is a verb that clearly shows a strong stance from the author’s perspective (by author, we mean the author(s) of the original text, the one being reported). According to Hyland’s framework of evaluative meaning of reported verbs, *defender* could be considered a neutral verb when related to the writer’s opinion (writer refers to the author of the text reporting other works), which, in turn, can indicate a positive view from the author’s perspective (Hyland and Jiang 2017). One could also argue that texts in Soft Science value clear positioning of authors and writers alike more than text in Hard Science do. Among the five most frequent verbs in the Hard Science corpus, the verb *sugerir* ‘suggest’ stands out, as it shows a more tentative stance from the author.

Furthermore, there are verbs that are used both in Soft and Hard Science together with *autor*, whether before or after the lemma. However, most appear exclusively in either Hard or Soft Science. The reporting verbs that are used in both Soft and Hard Science are *afirmar* ‘state’, *apontar* ‘point out’, *apresentar* ‘present’, *concluir* ‘conclude’, *considerar* ‘consider’, *defender* ‘defend’, *propor* ‘propose’, *referir* ‘refer’, and *sugerir* ‘suggest’. The reporting verbs used exclusively in Hard Science are *observar* ‘observe’, *encontrar* ‘find’, *verificar* ‘verify/check’, *descrever* ‘describe’, *recomendar* ‘recommend’, and *relatar* ‘report’. The verbs that are only used in Soft Science are *fazer* ‘make/do’, *analisar* ‘analyze’, *mostrar* ‘show’, *chamar* ‘call’, *destacar* ‘highlight’, and *procurar* ‘intend’.

These results are in line with Motta-Roth and Hedges (2010) who determined a list of the 18 reporting verbs most frequently used in subjects that correspond to what we understand as Hard Science. Out of these, only four match our results, namely *descrever* ‘describe’, *propor* ‘propose’, *sugerir* ‘suggest’, and *observar* ‘observe’. When it comes to the Soft Science subcorpus, the reporting verbs we retrieved that coincide with the Motta-Roth and Hedges’ (2010) list are five, namely, *sugerir* ‘suggest’, *mostrar* ‘show’, *propor* ‘propose’, *analisar* ‘analyze’, and *destacar* ‘highlight’. These discrepancies can be explained because we have narrowed down our analysis to the sequences *autor + verbo* ‘author + verb’ and *verbo + autor* ‘verb + author’. This was not the case in Motta-Roth and Hedges (2010), who included other types of citations in their study.

There is also partial agreement between our results and those of Hoffnagel (2010). Although our Soft Science subcorpus comprises more disciplinary registers than that of Hoffnagel, which is restricted to texts dealing with psychology and anthropology, the ten most frequent reporting verbs in both areas also appear within the first 15 positions in Soft Science, with the exception of *citar* ‘cite’ and *observar* ‘observe’, which are not attested in this subcorpus. For this analysis, we considered *realizar* and *fazer* as synonyms, since both may mean ‘do/make’.

The reporting verbs were also classified according to Hyland’s (1999, 2002) typology. The only verb that was excluded was *fazer* ‘do/make’, as it is a delexical verb whose meaning is extremely light since it is attached to the noun linked to it, as in *fazer referência* ‘make reference’ or *fazer uma discussão* ‘make a discussion’. The results are shown in Table 10.

	Hard Science	Soft Science
Research (real-world)	<i>Concluir</i> ‘conclude’ <i>Apresentar</i> ‘present’ <i>Observar</i> ‘observe’ <i>Encontrar</i> ‘find’ <i>Verificar</i> ‘verify’ <i>Apresentar-se</i> ‘introduce oneself’	<i>Apresentar</i> ‘present’ <i>Concluir</i> ‘conclude’ <i>Analizar</i> ‘analyze’ <i>Mostrar-se</i> ‘show oneself’ <i>Apresentar-se</i> ‘introduce oneself’ <i>Encontrar</i> ‘find’
Cognition	<i>Referir</i> ‘refer’ <i>Considerar</i> ‘consider’ <i>Defender</i> ‘defend’ <i>Recomendar</i> ‘recommend’ <i>Corroborar</i> ‘corroborate’	<i>Defender</i> ‘defend’ <i>Referir</i> ‘refer’ <i>Considerar</i> ‘consider’ <i>Referir-se</i> ‘refer oneself’ <i>Entender</i> ‘understand’ <i>Defender-se</i> ‘defend oneself’
Discourse	<i>Sugerir</i> ‘suggest’ <i>Afirmar</i> ‘state’ <i>Apontar</i> ‘point out’ <i>Descrever</i> ‘describe’ <i>Propor</i> ‘propose’ <i>Relatar</i> ‘report’ <i>Referir-se</i> ‘refer oneself’ <i>Citar</i> ‘quote’ <i>Destacar</i> ‘highlight’ <i>Acrescentar</i> ‘add’ <i>Dizer</i> ‘say’ <i>Propor-se</i> ‘propose oneself’ <i>Contribuir</i> ‘contribute’ <i>Reportar</i> ‘report’	<i>Afirmar</i> ‘state’ <i>Apontar</i> ‘point out’ <i>Sugerir</i> ‘suggest’ <i>Propor</i> ‘propose’ <i>Destacar</i> ‘highlight’ <i>Chamar</i> ‘call’ <i>Dizer</i> ‘say’ <i>Sublinhar</i> ‘underline’ <i>Salientar</i> ‘stress’ <i>Propor-se</i> ‘propose oneself’ <i>Argumentar</i> ‘argue’ <i>Destacar-se</i> ‘stand out’

Table 10: Classification and occurrences of reporting verbs based on Hyland’s (1999, 2002) typology

As shown in Table 10, the most frequent reporting verbs in the data belong to the group of discourse acts. This might be because the scope of our research on reporting verbs was restricted to the sequences ‘author + verb’ and ‘verb + author’. Thus, verbs such as

afirmar ‘state’, *apontar* ‘point out’, *descrever* ‘describe’, and *relatar* ‘report’ are relatively neutral verbs that do not convey a negative or a positive tone. According to Hyland and Jiang (2017), this phenomenon is becoming more frequent, as there is a tendency for authors to use neutral forms to refer to verbal activities.

5. SUMMARY AND FINAL REMARKS

Our research has addressed an existing gap in the literature regarding the analysis of reporting verbs in academic Portuguese, for verbs either following or preceding the lemma *autor* ‘author’. For this purpose, we used CoPEP and analyzed two subcorpora, one representing softer sciences and the other harder sciences. From the 6,103 valid occurrences of reporting verbs extracted from CoPEP, 3,716 (1.44) are used in the Soft Science corpus, while 2,387 (1.62) are used in Hard Science corpus, which shows a higher frequency of reporting verbs with *autor* ‘author’. The results showed that, from the 15 most used verbs in both Soft and Hard Science, there are nine verbs used in both corpora: *afirmar* ‘state’, *apontar* ‘point out’, *apresentar* ‘present’, *concluir* ‘conclude’, *considerar* ‘consider’, *defender* ‘defend’, *propor* ‘propose’, *referir* ‘refer’, and *sugerir* ‘suggest’, even though there is no agreement in the order in which they appear (judging by the number of occurrences).

Our list partially matches the verbs mentioned in previous studies (Motta-Roth and Hedges 2010), although it is difficult to compare and contrast the data since this study differs in the way citations were collected. In our data, the reporting verbs are used mainly in the simple present and preterit perfect tenses, with a preference for the use of active voice and the order *autor + verbo* ‘author + verb’. These patterns of use might be related to the genre under analysis. Nevertheless, further research which includes genre as a variable could help support this argument. Some patterns regarding different uses of reporting verbs in English according to disciplinary areas have not been found in our corpus, making it difficult to draw comparisons between English and Portuguese. Searching for other citation patterns can help shed further light into this. For now, one could also speculate that disciplinary differences are not so marked in Portuguese, which points to more established patterns in English.

As stated earlier, the pedagogical concerns leading to this study explain why it is more related to the form of the occurrences and not so much to their rhetorical function

in the texts. The number of excerpts with reporting verbs made it difficult to code them manually with relation to this pragmatic aspect. Although it is a large number, it needs to be acknowledged that different types of citation are not included, for example, cases with the use of proper nouns or where nouns such as *researchers* or *scholars* were used, among others. This means that there is still room for more studies that encompass other forms of citation, which could then account for the disciplinary variation that can be seen in these structures. While it is understood that form, meaning, and function are intertwined, the focus on form here also helps address the lack of studies dealing with lexico-grammatical features of academic Portuguese, as pointed out by Kuhn (2017).

Other aspects that were not controlled for in our study include the section of the paper from which the reporting verb came from, which can influence the verb tense, and whether they came from the same article. Therefore, one way of continuing this study would be to broaden the search so as to encompass other forms, while gathering and coding for more information about the occurrences. Another possibility is looking into the differences regarding the language varieties represented in CoPEP. Once accounting for a representative sample with different types of reporting occurrences, more patterns can be brought to light and can then be compared and contrasted to other more widely researched languages, such as English.

Despite the limitations of this study, our findings can be useful and represent, to the best of our knowledge, a first step into the large-scale study of reporting verbs in different disciplinary areas in Portuguese. They can aid the development of much-needed pedagogical materials aimed at novice researchers and learners of academic Portuguese, whether as a first or a second language. They can also be used for studies with learner corpora, since CoPEP is a representative sample of academic language both in the Brazilian and European varieties of Portuguese.

REFERENCES

- Bessa, José Cezinaldo Rocha. 2011. (Re)pensando a citação em textos acadêmico-científicos. *Signum: Estudos Da Linguagem* 14/2: 421–439.
- Charles, Maggie. 2006. Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes* 25/3: 310–331.
- Coffin, Caroline, Mary Jane Curry, Sharon Goodman, Ann Hewings, Theresa Lillis and Joan Swann. 2005. *Teaching Academic Writing: A Toolkit for Higher Education*. New York: Routledge.

- Corbari, Alcione Tereza and Quézia Cavalheiro M. Ramos. 2018. Verbos dicendi na notícia: Pontos de um continuum argumentativo na construção da intertextualidade. *Fórum Linguístico* 15/1: 2903–2923.
- Hernández Bonilla, Juan Miguel. 2021. How to end the hegemony of English in scientific research. *El País* English Edition. <https://english.elpais.com/usa/2021-07-30/how-to-end-the-hegemony-of-english-in-scientific-research.html> (3 Jul, 2022.)
- Hoffnagel, Judith C. 2010. A prática de citação em trabalhos acadêmicos. *Cadernos de Linguagem e Sociedade* 10/1: 71–88.
- Hyland, Ken. 1999. Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics* 20/3: 341–367.
- Hyland, Ken 2002. Activity and evaluation: Reporting practices in academic writing. In John Flowerdew ed. *Academic Discourse*. London: Longman, 115–130.
- Hyland, Ken and Feng Jiang. 2017. Points of reference: Changing patterns of academic citation. *Applied Linguistics* 40/1: 64–85.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1/1: 7–36.
- Kowaltowski, Alicia, Marcus Oliveira and A. Ariel Silver Hernan Chaimovich. 2021. The push for open access is making science less inclusive. *Times Higher Education (THE)*. <https://www.timeshighereducation.com/opinion/push-open-access-making-science-less-inclusive> (3 Jul, 2022.)
- Kuhn, Tanara Zingano. 2017. *A Design Proposal of an Online Corpus-driven Dictionary of Portuguese for University Students*. Lisboa: Universidade de Lisboa dissertation.
- Kuhn, Tanara Zingano and José Pedro Ferreira. 2020. O Corpus de Português Escrito em Periódicos - CoPEP. *Documentação de Estudos em Lingüística Teórica e Aplicada* 36/2: 2020360209. <https://doi.org/10.1590/1678-460x2020360209>
- Motta-Roth, Désirée and Graciela Rabuske Hendges. 2010. *Produção Textual na Universidade*. São Paulo: Parábola Editorial.
- Oliveira, Isabel Tiago de, Madalena Ramos, Ana Ferreira and Sofia Gaspar. 2015. Estudantes estrangeiros em Portugal: Evolução e dinâmicas recentes (2005/6 a 2012/13). *Revista de Estudos Demográficos* 54: 39–56.
- Pérez Llantada, Carmen. 2021. Genres and languages in science communication: The multiple dimensions of the science policy interface. *Language & Communication* 78: 65–76.
- Quintans-Júnior, Lucindo José, George Rego Albuquerque, Sérgio Campello Oliveira and Robério Rodrigues Silva. 2020. Brazil's research budget: Endless setbacks. *EXCLI Journal* 19: 1322–1324.
- Saburi Costa, Bianca Freitas and Cláudia Freitas. 2017. Verbos de elocução em português: Um estudo descritivo com base em grandes corpora e motivado pela linguística computacional. *Fórum Linguístico* 14/3: 2266–2285.
- Smith, Laurence Daniel, Lisa A. Best, Donald Allan Stubbs, John Johnston and Andrea Bastiani Archibald. 2000. Scientific graphs and the hierarchy of the sciences: A latourian survey of inscription practices. *Social Studies of Science* 30/1: 73–94.
- Shaw, Philip. 1992. Reasons for the correlation of voice, tense, and sentence function in reporting verbs. *Applied Linguistics* 13/3: 302–319.
- Solovova, Olga Santos, Joana Vieira and Joaquim Veríssimo. 2018. Publish in English or perish in Portuguese: Struggles and constraints on the semiperiphery. *Publications* 6/2: 1–14. <https://doi.org/10.3390/publications6020025>

- Souza, Medianeira and Wellington Vieira Mendes. 2012. Uma análise sistêmico-funcional do dizer em artigos científicos de graduandos. *Documentação de Estudos em Lingüística Teórica e Aplicada* 28: 537–560.
- Stumpf, Elisa Marchioro. 2021. Portuguese as an additional language for academic purposes: Contributions from the academic literacies model to course design. *Revista Linguagem & Ensino* 24/2: 317–331.
- Swales, John. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Thompson, Paul. 2005. Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes* 4/4: 307–323.

Corresponding author

Marine Laísa Matte
Federal University of Rio Grande do Sul
Department of Modern Languages
Bento Gonçalves Avenue 9500
91540-000 Porto Alegre
Brazil
E-mail: marine.matte@ufrgs.br

received: October 2021
accepted: June 2022