THE DISCIPLINARY CULTURE OF CITATION IN SCIENTIFIC ARTICLES IN THE HUMANITIES

LA CULTURA DISCIPLINARIA DE LAS CITAS EN LOS ARTÍCULOS CIENTÍFICOS DE HUMANIDADES

A CULTURA DISCIPLINAR DA CITAÇÃO EM ARTIGOS CIENTÍFICOS NAS HUMANIDADES

Fabiana Komesu
Universidade Estadual Paulista "Júlio de Mesquita Filho" (Unesp), Brasil
fabiana.komesu@unesp.br
https://orcid.org/0000-0002-3820-1559

Juliana Alves Assis
Pontifícia Universidade Católica de Minas Gerais (PUC Minas), Brasil
juassis@pucminas.br
https://orcid.org/0000-0002-9383-4850

Christiane Donahue
Dartmouth College, Estados Unidos
christiane.K.Donahue@dartmouth.edu
https://orcid.org/0000-0001-8738-3275

Abstract:
This article aims to identify and analyze aspects that are not dealt with or cannot be visualized in the bibliometric indexes of journals and articles in the Humanities, specifically in the field of Linguistics. This study is based on a data set of the "citing texts" (114) of three articles of the field of Linguistics that come from a Brazilian journal, a Chilean journal and a South African journal, among the most cited articles of the journals Scielo basis (Google Scholar data) in the month of September, 2021. The methodological procedures take into account the criteria (i) year of publication of the citing texts, (ii) language of publication of the citing texts, (iii) discourse genre of the citing texts, and (iv) location of the institution to which the author(s) of the citing texts are affiliated. The results problematize a culture restricted to bibliometric indexes.

Keywords: Literacy, Writing, Citation indexes, Linguistics.

Resumen:
Este artículo tiene como objetivo identificar y analizar aspectos que no son tratados o no pueden ser visualizados en los índices bibliométricos de revistas y artículos de Humanidades, específicamente en el campo de la Lingüística. Este estudio se basa en un conjunto de datos de los "textos a citar" (114) de tres artículos del campo de la Lingüística que provienen de una revista brasileña, una revista chilena y una revista sudafricana, entre los artículos más citados de las revistas base Scielo (datos de Google Scholar) en el mes de septiembre de 2021. Los procedimientos metodológicos tienen en cuenta los criterios (i)

año de publicación de los textos citados, (ii) lengua de publicación de los textos citados, (iii) género del discurso de los textos citados y (iv) localización de la institución a la que pertenecen el autor o los autores de los textos citados. Los resultados cuestionan una cultura restringida a los índices bibliométricos.

Palabras clave: Literacidad, Escritura, Índices de citas, Lingüística.

Resumo:
Este artigo tem o objetivo de identificar e analisar aspectos que não são tratados ou não podem ser visualizados em índices bibliométricos de periódicos e artigos na área de Humanidades, especificamente no campo da Linguística. Este estudo é baseado em um conjunto de dados da "textos citantes" (114) de três artigos da área de Linguística, oriundos de uma revista brasileira, de uma revista chilena e de uma revista sul-africana, entre os artigos mais citados da base Scielo (dados do Google Scholar) no mês de setembro de 2021. Os procedimentos metodológicos levam em consideração os critérios (i) ano de publicação dos textos citantes, (ii) língua da publicação dos textos citantes, (iii) gênero do discurso dos textos citantes e (iv) localização da instituição à qual o(s) autor(es) dos textos citantes estão filiados. Os resultados colocam em questão uma cultura restrita a índices bibliométricos.

Palavras-chave: Letramento, Escrita, Índices de citação, Linguística.

Recibido: 30 de octubre de 2022 **Aceptado**: 13 de junio de 2023

1. Introduction

In recent years, there has been an increasing number of papers and other initiatives aimed at discussing different aspects related to the uses and effects of bibliometric indexes in gauging the quality of a journal, an article, or even a scholar (Angermuller and Hamann 2019; Araújo and Sardinha 2011; Ampudia de Haro 2017; among others). International visibility and impact are factors directly linked to the systems of metrics of publication with which the researchers have come to live in recent years, and which are increasingly integrated into scientific culture (Santos and Kobashi 2009; Purvis 2008). This does not mean, however, that we are talking about something assumed as a consensus in all fields of knowledge or even within the same field. In this scenario, a discourse that questions the relevance of these indexes has been gaining more and more strength among researchers from different fields of knowledge (Marcovitch 2019; Marques 2013). Undesirable consequences have been discussed, such as (i) stronger competition among researchers to

publish, which is driven by metrics, research funding, and career demands. (Hyland and Jiang 2019), (ii) search for gimmicks to expand the impact of published articles, leaving other important goals of scientific activity in the background (Marques 2013) and (iii) standardization of scientific quality assessment systems, to the detriment of the specificities of each disciplinary culture. The logic of evaluation (and awarding) by citation indexes favors certain fields/subfields of knowledge, but not exactly the Humanities and its subfields.

In an attempt to contribute to this debate, this article aims to identify and analyze aspects that are not dealt with or cannot be visualized in the bibliometric indexes of journals and articles in the Humanities, specifically in the field of Linguistics. This study is based on a data set of the "citing texts" of three articles in the field of Linguistics that come from a Brazilian journal, a Chilean journal and a South African journal, that are among the most cited articles of the journals' Scielo basis (Google Scholar data) in the month of September, 2021. "Citing texts" are the texts that cite each of these three articles. In total, the citing texts add up to 114 texts. This corpus is a selection from a larger set collected as part of two international research projects on scientific articles published in high-impact journals in different fields of knowledge and in different languages.[1]

The specific objectives are to investigate: (i) the year of publication of the citing texts; (ii) the language of publication of the citing texts; (iii) the genres of the citing texts; (iv) the location of the institutions with which the authors of the citing texts are affiliated. The investigation of these different aspects seeks to problematize the functioning of the disciplinary culture of citation in scientific articles in the Humanities, using the field of Linguistics as a specific case example. In dialogue with Angermuller and Hamann (2019), this study assumes that citation indexes should be treated as cautiously as any other social metric, because they not only reflect but also construct realities as those experienced by researchers in a productivist logic of evaluation.

[1] This study is embedded in a broader investigation, carried out in the framework of two research projects funded by Brazilian government agencies: (1) the International Research Network "Literacies in Different Fields of Knowledge", from the CAPES-PrInt-Unesp program (Capes/Auxpe-Brazil process 88881.310711/2018-01), under the coordination of the first author (ongoing since 2019); (2) "Escrita acadêmica/escrita científica: das formas de presença do autor, do outro, das áreas de conhecimento e seus domínios disciplinares" / "Academic writing/scientific writing: forms of presence of the author, of the other, of the areas of knowledge and their disciplinary fields" (Universal Call MCTIC/CNPq-Brazil, 2018, process 434902/2018-7), under the coordination of the second author (2019-2022).

The definition of research fields as humanities[2] or social sciences is governed by epistemological, but also institutional, political, cultural parameters. To this extent, the classification may vary by country or institution. While Linguistics is often considered a social science, we treat it here as part of the Humanities, because its subject matter is language and discourse, and it is highly interdisciplinary with interfaces built with Education, Social Sciences and Computing. We consider that, given the motivations of the present work, the choice of the field of Linguistics is pertinent, mainly for its potential to encompass realities related to the other fields of the Humanities, or even to give rise to studies that refine specificities of each one of them.

After this introduction, the paper has three other main parts. In the first one, we present a review of the literature on bibliometric indices and their impacts on the scientific field, taking into account different disciplinary cultures and observing the objectives of our study. Then, in the second part, we describe the process of data collection and organization for the study of citations and cited texts that we carried out in order to better understand the issues with bibliometric measures. The data analysis is presented in the third part. We conclude with some notes on discussion of the study.

2. Literature Review

Bibliometric indexes are associated with a set of efforts and investments, made in recent years by different actors: researchers, universities, research funding agencies, and journals, all seeking to increase the international visibility and impact of science. This leads to international cooperation, which would result in co-authorship with foreign researchers as well as publishing in foreign languages, especially in English, without necessarily considering the local or regional impact of the research.

Assuming communication as an essential activity for the progress and the very permanence of science, we cannot fail to reflect on the effects that the technological

[2] As exemplified by Hammarfelt (116), "The Organization for Economic Co-operation and Development (OECD) lists history, archaeology, genealogy, literature, languages, philosophy, arts, history of arts, religion and theology (OECD 68) while The European Reference Index for the Humanities (ERIH) distinguishes fifteen fields in the humanities (including educational research as well as gender studies and psychology). In the United States, however, the Humanities Resources Center includes eleven fields (Leydesdorff et al. 2011)".

paraphernalia devices currently adopted for measuring the visibility and impact of science, that is, the bibliometric indicators, have on the systems of academic production and evaluation. One of them seems to us well defined by Salgado and Clares (30), when associating today's articles with "an algorithmic function recoverable on the web, which boosts postgraduate programs, teaching careers, and international projection".

In the Brazilian context, we mention, for example, actions of the Assessment Advisory of Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/ Coordination for the Improvement of Higher-Level Personnel), a foundation linked to the Ministry of Education of Brazil. Such actions, released in July 2019[3], are aimed at improving the evaluation process of the intellectual production of graduate studies in the country. In the case of the evaluation of journals, what was sought by the group, in addition to adopting more objective criteria, was to avoid distortions, that is, the same journal receiving a different classification according to the field. The powerful influence of the bibliometric indexes also showed itself there, since the new Qualis Periodicals methodology disclosed by Capes proposes a reference classification, given through the combined use of bibliometric indicators—the CiteScore (from the Scopus database), the Impact Factor (from the Web of Science database) and the h5 index (from Google Scholar). According to results reported by Martín and Martín (2021), the Google Scholar database captures citations in higher percentage for all major fields of knowledge than all other databases.

The "h-index" was proposed in 2005 by Argentine physicist Jorge Hirsch, professor at the University of California (USA), as a tool that would have the power to combine quantity and quality in measuring the academic production of a researcher through the number of citations received. Over time, it was adopted as a parameter for evaluation and ranking of journals and institutions in some fields (Droescher and Silva 2014). Therefore, there is, on the one hand, a growing chorus of appreciation of this index expressed in scientific articles from different disciplinary cultures (Araújo and Sardinha 2011), but on the other hand, a questioning of the effectiveness or exclusivity of this resource for the

[3] Retrieved 16 October 2022, from https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/ORIENTAES_PROCESSO_AVALIATIVO_INFORMATIVO_1.pdf.

preparation of rankings or for the foundation of criteria for granting research investment (Ampudia de Haro 2017).

Regarding critical approaches on the issue, especially those emanating from the Social Sciences and Humanities, one of the strong arguments often highlighted is the specificities of the fields of knowledge and the differences in the modes of production and circulation of new knowledge. We can point out that books and book chapters continue to be a fundamental vehicle for the dissemination of research in some fields. On the other hand, we must recognize the strength of the discourse of the knowledge economy and impact science (Ampudia de Haro 2017), an orientation that tends towards homogenization by establishing common criteria across different disciplinary cultures.

Considering the disciplinary culture in the reflection on academic writing practices leads us to problematize the weight of bibliometric indexes as indicators of the quality of an article, a scholar, or a journal. Disciplines "are sites where differences in worldview or language use intersect as a result of the myriad backgrounds and overlapping membership of the participants" (Hyland, *Disciplinary* 20). They are composed, therefore, of views, values, and beliefs that mark the disciplinary field. This condition is affected by the universalistic and globalizing paradigm that governs the logic of evaluation by bibliometric indexes, which gives the discipline a transnational character.

To better illustrate this discussion about the literature review, we propose a brief comparison between h5 indexes from publications of different fields of knowledge. This comparison reveals clear differences between disciplinary cultures with respect to the number of citations received, as we can see, for example, in the comparison between the metrics of publications listed by Google Scholar[4] in Tables 1, 2 and 3 below.

[4] Clarifications about the metrics presented by Google Scholar: "Google Scholar Metrics provide an easy way for authors to quickly gauge the visibility and influence of recent articles in scholarly publications. Scholar Metrics summarize recent citations to many publications, to help authors as they consider where to publish their new research. To get started, you can browse the top 100 publications in several languages, ordered by their five-year h-index and h-median metrics. To see which articles in a publication were cited the most and who cited them, click on its h-index number to view the articles as well as the citations underlying the metrics. You can also explore publications in research fields of your interest. To browse publications in a broad field of research, select one of the fields in the left column. For example: Engineering & Computer Science or Health & Medical Sciences. To explore specific research fields, select one of the broad fields, click on the 'Subcategories' link and then select one of the options. For example: Databases & Information Systems or Development Economics. Browsing by research field is, as yet, available only for English publications. You can, of course, search for specific publications in all languages by words in their titles. Scholar Metrics

Top Publications	h5-index[5]	h5-median[6]
Nature	444	667
The New England Journal of Medicine	432	780
Science	401	614
IEEE/CVF Conference on Computer Vision and Pattern Recognition	389	627
The Lancet	354	635

Table 1 – Metrics of journals with highest h-indexes (Top publications - Scholar Metrics)
Source: Data released by Google Scholar. Retrieved 16 October 2022, from https://scholar.google.com.br/citations?view_op=top_venues&hl=pt-BR&vq=en

In Table 1, all the "Top publications – Scholar Metrics" are international, published in English, concentrated in the United Kingdom and the United States, and restricted to three fields specified by Google Scholar: Health and Medical Sciences, Life and Earth Sciences, and Engineering and Computer Science. In these fields are the publications with the highest impact factor, according to the consulted metrics.

The comparison of these indexes with those obtained by publications of other nationalities with higher H-indexes, also considering the numbers brought by Google Scholar, leads us to see a significant difference in terms of the behavior of such indexes, according to the origin of the publications. Let us turn to Table 2 to view the results found for one journal of Brazilian, Spanish, German and French origin, respectively, with the highest metrics, according to Google Scholar.

---

are currently based on our index as it was in July 2021". Retrieved 16 October 2022, from https://scholar.google.com.br/intl/pt-BR/scholar/metrics.html.

[5] According to information from Google Scholar, "The h5 index is the h-index of articles published in the past five years. This is the highest h number in a publication, where h articles published from 2016 to 2020 have been cited at least h times each." Retrieved 16 October 2022, from https://scholar.google.com.br/citations?view_op=top_venues&hl=pt-BR&vq=en. For example, if a journal has an h5-index equal to 13, it means that it has at least 13 articles that have received, in the last 5 years, at least 13 citations.

[6] According to information from Google Scholar, "The median h5 of a publication consists of the average number of citations for the articles that make up its h5 index." Retrieved 16 October 2022, from https://scholar.google.com.br/citations?view_op=top_venues&hl=pt-BR&vq=en. For example, if a journal has a median h5 equal to 20 and an h5-index equal to 13, it means that the median of the citations of the 13 articles taken for the h5-index is 20.

Top Publications Brazilian journals	**h5-index**	**h5-median**
Ciência & Saúde Coletiva	72	97
Cadernos de Saúde Pública	65	94
Top Publications Spanish journals	**h5-index**	**h5-median**
El Profesional de la Información	50	77
Revista Panamericana de Salud Pública	44	70
Top Publications German journals	**h5-index**	**h5-median**
Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz	26	31
Zeitschrift für Erziehungswissenschaft	25	34
Top Publications French journals	**h5-index**	**h5-median**
Cahiers Agricultures	17	26
Reseaux	17	26

Table 2 – Metrics of journals with the highest H-index from different countries
Sources: Data released by Google Scholar. Retrieved 16 October 2022, from:
https://scholar.google.com.br/citations?view_op=top_venues&hl=pt-BR&vq=pt;
https://scholar.google.com.br/citations?view_op=top_venues&hl=pt-BR&vq=es;
https://scholar.google.com.br/citations?view_op=top_venues&hl=pt-BR&vq=fr;
https://scholar.google.com.br/citations?view_op=top_venues&hl=pt-BR&vq=de

The information in Table 2 may favor interesting interpretations not only about the disciplinary cultures in question, but also about specificities of the scientific culture of each country in the Table, based on Google Scholar data. It draws attention both to the variety of the dominant field(s) of scientific knowledge and to the citation rates in each journal, as well as to the variation of rates itself. In the case of Brazil, the two journals with the highest h5-index are from the Public Health field; regarding the journals from Spain, journals from the Communication and also the Public Health fields dominate. The journals with the lowest h5-index, among the four countries listed, are from Germany and France. The German journals are divided into the fields of Public Health and Education, respectively. The French journals cover the fields of Agricultural Sciences and Communication, respectively.

As a last example, we now turn our attention to Table 3, showing the Top publications in the specific field of interest to this article: Language and Linguistics.

Journals (Major publications in Language and Linguistics)	**h5-index**	**h5-median**
Journal of Memory and Language	44	68
Language Learning	42	69
Applied Linguistics	40	77
Studies in Second Language Acquisition	37	65
Language, Cognition and Neuroscience	36	47

Table 3 – Metrics of journals with highest h-indices in the field of Language and Linguistics (Top publications in the field in question - Scholar Metrics)
Source: Data released by Google Scholar. Retrieved 16 October 2022, from https://scholar.google.com.br/citations?view_op=top_venues&hl=pt-BR&vq=hum_languagelinguistics

A brief comparative examination of the information presented in Tables 1, 2 and 3 allows us to offer some observations and reflections, and in particular questions, some of them endorsed by scholars who have been working on the subject for years (Ioannidis et al. 2014; Thomaz, Assad and Moreira 2011; among others). Part of these questions is contemplated in the recommendations expressed in the "San Francisco Statement on Research Evaluation (DORA)" written by a group of editors of scientific journals that met during the Annual Meeting of the American Society for Cell Biology (ASCB) in San Francisco (CA) in 2012 (Retrieved 16 October 2022, from https://blog.scielo.org/blog/2013/07/16/declaracao-recomenda-eliminar-o-uso-do-fator-de-impacto-na-avaliacao-de-pesquisa/#.YmHAcNrMK71).

At first, two aspects stand out: the significant difference in citation rates by field of knowledge and the clear predominance of English language journals with higher citation rates. What can we say or question about them? A first issue could be one that questions the assumption underlying the use of such indexes: are the most cited articles, in fact, the most important ones? Moreover: what do such indexes have to do with the nature of the field and with the ways of doing science in each disciplinary culture? Wouldn't the limit of years for counting citations be privileging fields that have short average citation lives, as in Health Sciences, for example, and disfavoring others, such as History and Linguistics, where citations seem to have longer lives? Wouldn't the high citation density of English-language publications have, as a counterpart, a weakening of journal articles dedicated to portraying

local science (Corrêa, 2020)? What impacts can evaluation approaches based on values originating in the natural sciences bring, in the long run, to scientific research in the Humanities? Moreover, questioning this evaluation system could be “an argument for changing the relationships in the publication networks so that we can reconstruct knowledge – and presumably conduct international relations – in more egalitarian and enriching terms” (Canagarajah 305).

Several other questions and reflections could be added to this list. However, in this article, as stated, we are interested in turning our gaze specifically to the citation indexes of journals in the large field of Humanities, seeking to give visibility to aspects that can help us to interpret these numbers from other angles. We will take as “corpus”: (i) three (03) articles from the Linguistics field among the most cited in the Scielo base (data from Google Scholar), each of them coming from three (03) journals from the Linguistics field – one Brazilian, one Chilean and one South African – also selected from the h5-index of the journals from the Linguistics field, and (ii) the set of citing texts related to these three (03) articles, for a total of 114 citing text, after removing inconsistencies, as commented in the methodology below.

3. Data Set and Methodology

As mentioned, data collection was carried out in September 2021. We selected three highly-cited articles from current indexes, and then selected the texts that cited these three articles. The three articles are in the field of Linguistic Studies that come from a Brazilian journal, a Chilean journal, and a South African journal. All in all, that adds up to 114 citing texts, after removing distortions, as we will explain below. In short, the criterion for the selection of journals and articles was the h5-index.

Still about the collection, we clarify that our option was to work with journals outside the context of the United States and the United Kingdom, which, as can be seen in Table 3, have the highest h5-index in Google Scholar. Therefore, this selection prioritized, among the journals with the highest h5-index, those that were outside these two geographic spaces. This is a criterion that privileges disciplinary field and geographic space that are not dominant.

In Table 4, below, we present a view of this data:

Identification of the article/journal	**h5-index and h5-median of the journal**	**Number of authors of the article**	**Year of article's publication**	**Number of citations of the article (Google Scholar data)**	**Total citing texts analyzed; inconsistencies removed**
Brazilian Journal Article (BJA)	h5-index: 13 h5-median: 19	01	2016	42	33
Chilean Journal Article (CJA)	h5-index: 14 h5-median: 20	07	2016	70	53
South African Journal Article (SAJA)	h5-index: 08 h5-median: 10	03	2017	30	28

Table 04 – Information about the research data.
Source: prepared by the authors.

The h5-index and h5-median numbers registered by Google Scholar at the time of collection (September 2021), for journals in the field of Linguistic studies, were the parameters for the constitution of the research data. Once we identified the three (03) journals with the highest rates in the indicated field, provided that they did not originate from the United States or the United Kingdom, as we noted above, we selected the article with the highest number of citations from each of these 03 (three) journals.

The second step was, also through the information provided by Google Scholar, to access all citing texts of each of the selected articles, indicated by the tool. In this step, we identified some inconsistencies, which led to the numbers shown in the last column of Table 4. In short, we detected some differences between what was recorded in our consultation to Google Scholar and what the data obtained through this tool revealed.

The main inconsistencies identified were: (i) duplicate version of the text that cites the article (this is a copy of the same article that was previously counted by Google Scholar); (ii) version of the same text quoting the article, in another language; (iii) the text selected by Google Scholar did not mention the cited article; (iv) the year of publication of the text cited in the citing text is different from that indicated by Google Scholar.

Based on this inventory, we have the following results regarding data collection: (i) 42 citations are attributed to the BJA, when it is actually 33 (78.6% of the total detected; loss of 21.4% of the texts originally indicated as citing); (ii) 70 citations are attributed to the CJA, when it is actually 53 (75.7% of the total detected; loss of 24.3% of the texts originally indicated as citing); (iii) 30 citations are attributed to the SAJA, when it is actually 28 (93.3% of the total detected, loss of 6.7% of the texts originally indicated as citing). In total, therefore, there were 114 citing texts analyzed, out of a total of 142 citing texts indicated.

Finally, we note that, as metrics are used for promoting rankings of researchers/universities and for the provision of funding, the identification of inconsistencies seems to be an important aspect to be considered in this discussion.

4. Data Analysis

The data analysis is organized[7] to accomplish the four specific objectives, namely, (i) the year of publication of the citing texts; (ii) the language of publication of the citing texts; (iii) the genres of the citing texts; (iv) the location of the institutions to which the authors of the citing texts are affiliated. The examination of such aspects, not directly visible by an article's h5-index, may favor a clearer understanding of the bibliometric indexes and the functioning of the disciplinary culture of citation in scientific articles in this field of knowledge.

[7] The organization of the content of the 114 documents was done with the MAXQDA 2020 software (VERBI Software 2020). We are grateful to Amanda de Carvalho Valadão, João Vitor Moreira, Taíne Soares de Jesus, Thaís Cristina de Assis, Tiago Ruas Dieguez, undergraduate students at PUC Minas, and Tamiris Vianna da Silva, doctoral student at Unesp, for their help in data collection.

4.1. Year of Publication of the Citing Texts

The analysis of the 33 citing texts of the Brazilian journal article (BJA) showed that they were mainly published in the years of 2019-2020 (19 of the 33, corresponding to 57.5% of the total amount), it means at least three or four years after the publication of the initial source text, in 2016. Seven (07) of the 33 citing texts were published in 2018 (21.2%); six of them (18.1%), in 2021, and only one in 2016. There is no register of a quoting from BJA in 2017.

The analysis of the 53 citing texts of Chilean journal article (CJA) showed that they were mainly published in the years of 2019-2020 (31 of 53, corresponding to 58.5% of the total amount), that means at least three or four years after the publication of the initial source text, in 2016, as we also saw in the Brazilian case. Fifteen (15) of the 53 citing texts were published in 2017-2018 (28.3% of the total amount), that means one or two years after the publication of the initial source text. In the year of publication of that text, there is a register of five citing texts (9.4% of the total amount). There are also two citing texts recorded in 2021.

The analysis of the 28 citing texts of South African journal article (SAJA) showed that they were mainly published in the years 2019-2020 (18 of 28, corresponding to 64.3% of the total amount), that means two or three years after the publication of the quoted text. Eight (08) of the 28 citing texts (25%) were published in 2018, one year after the publication of the quoted text. There are also three citing recorded in 2021.

Two of the most cited articles selected from h5-index of the journals from the Linguistics field were published in 2016 and one in 2017. The concentration of citing texts is in the years 2019-2020 (68 out of 114, corresponding to 59.6% of the total amount). The data analysis shows that it takes at least two years for articles from the field of Linguistics to be cited.

This period of at least two years highlights the disciplinary culture in the field. There is a demand of time for the text to be read, discussed, quoted in the field, also taking into account the publication time of the citing text. In other fields the issue of time appears (more agile) as an imperative in the functioning of citation and the production of knowledge (Ioannidis and Boyack and Wouters 2016). It is well known, for example, that

in the period of the COVID-19 pandemic, the scientific information and communication cycle was accelerated, with preprints, rapid systematic reviews, and adoption of a fast track process in article submission (Sepúlveda-Vildósola et al. 2020). The linguistics field, in the Humanities, demands a time (of reflection) for citation.

4.2. Language of Publication of the Citing Texts

The most cited article from the Brazilian journal (BJA) was written in Portuguese; the Chilean one (CJA), in Spanish, and the South African in English (SAJA). Concerning the language of publication of the citing texts, we observe that the citing texts of BJA were mainly published in Portuguese (28 of 33, corresponding to 84.8% of the total amount). Three citing texts were published in English (9% of the total amount), one in German and one in Spanish.

The citing texts of CJA were mainly published in Spanish (43 of 53, corresponding to 81.1% of the total amount), but also in English (9 of 53, almost 17% of the total amount) and in Portuguese (1 of 53). In turn, the citing texts of SAJA were mainly published in English (22 of 28, corresponding to 78.5% of the total amount). Four of the 28 citing texts were published in Spanish (14.2% of the total amount), one in Norwegian and one in Danish.

Lillis and Curry (2010) have critically discussed the issue of English as a lingua franca of journals or even as a “language of Science” for years. The authors problematize how English plays a central role in globalized systems and practices, given the pressure for publication metrics, citation and the “internationalization” of institutions. In the top-ranked journals, English appears as a “presumed requirement” in different fields of knowledge and now sharply in Humanities.[8]

In this context of politics of academic knowledge production, we highlight that in the general set of this study, the language of the citing texts in Linguistics is predominantly Spanish (48 out of 114, corresponding to 42.1% of the total amount). English appears as the

[8] In this paper in which we discuss the disciplinary culture of citation in scientific articles in the field of Linguistics, we have chosen to write the text in English. This is not, as we seek to discuss, a "choice", but rather the injunctions that determine literate practices in the scientific field.

second most used language (34 of 114, corresponding to 29.8%), followed by Portuguese (29 of 114, corresponding to 25.4%).

Some critics might say that this trend is conditioned by the fact that article CJA, the one that received the most citations (n=53), was written in Spanish. The citing texts were predominantly written in Spanish (43 of 53 citing texts). The tendency for the citing texts to be published in the same language as the most cited article is also verified in BJA, written in Portuguese, with 28 of the 33 citing texts published in Portuguese, and in SAJA, written in English, with 22 of the 28 citing texts published in English. This trend may reflect the dialogue that researchers in the language sciences establish with their peers, considering the specificities of the linguistic expertise involved in the production of knowledge.

One of the major challenges of academic discourse, as pointed out by Hyland (2017), is how this discourse represents disciplinary realities. In the case of the Humanities, says the author, there is the employment of "abstraction rather than technicality, moving from instances to generalizations by gradually shifting away from particular contexts" (Hyland *ESP* 7-8). In this scenario, scientific communication in a language that is not the researcher's native one is not at all obvious.

We observe that the demand for English in scientific production is still recent in the field of linguistic studies. Gradually, this practice is being assumed by the force of institutional coercions of internationalization made up by funding agencies, universities, and post-graduate programs. Local and regional interests of scientific production are left aside, as noticed by Curry and Lillis (2018).

4.3. The Genres of the Citing Texts

The selected articles were cited in different genres of discourse, with a predominance of journal articles. The most cited article from the Brazilian journal (BJA) was quoted in other journal articles (21 out of 33 citing texts, corresponding to 63.3% of the total amount). There are also citations in Ph.D. thesis (4 of 33 citing texts), Master's dissertation (3 of 33), book chapters (3 of 33) and completed work in annals (2 of 33).

The most cited article from the Chilean journal (CJA) was quoted in other journal articles (35 of 53 citing texts, corresponding to 66% of the total amount). There are also

citations in book chapters (5 of 53 citing texts), and term papers (3 of 53), reviews (2 of 53) and preface, research project and completed work in annals, each one with one occurrence. The most cited article from South African journal (SAJA) was also mainly cited in journal articles (24 of 28 citing texts, corresponding to 85.7% of the total amount). There are also citations in papers (2 of 28 citing texts), and in a Ph.D. thesis and abstract, each one with one occurrence.

We observed a tendency for the most cited articles from the Linguistics field to be quoted in other journal articles. In the case of the South African journal article, this tendency is more accentuated, with 85.7% of the citing texts concentrated in this genre, against 63.3% in the case of the Brazilian journal article and 66% in the case of the Chilean journal article.

There are different reasons why the journal article is the most recurrent citing text genre in the data set. From a linguistic perspective, it should be noted that the discourse genre has a communicative purpose circumscribed to a certain sphere of discursive communication. As Hyland (*Disciplinary* 20-21) reminds us, there are epistemic conventions of the disciplines in the discursive construction of knowledge, a "way of formulating and negotiating knowledge" that "defines what it takes knowledge to be". In academic writing, the "journal article" genre expresses a form of socialization through which the scientific production presented seeks to contribute to the advancement of other studies and research. The citation in the journal article provides, "an intertextual framework for new work, allowing the writer to construct an effective justification for an argument and demonstrate the novelty of his or her position." (Hyland, *Disciplinary* 25).

The adoption of the journal article thus satisfies an institutional expectation of scientific communication not only in the domain of Humanities and Linguistics, but also in different fields of knowledge. The quality of what is presented through this genre is assessed by a peer review system in journals that have experts from different institutions on their editorial board. This is a mode of circulation that seeks to respond to the coercions of scientific production, also in dialogue with other fields of knowledge. In the field of linguistic studies, as seen, the average time between the publication of the cited article and the publication of citing texts is at least two years.

Besides the journal article genre, we point out that the field of Linguistics also uses other genres, such as book chapters (8 of 114 citing texts), Ph.D. theses (7 of 114) and Master's dissertations (6 of 114). There are also records, in the general set, of complete work in annals (3 of 114), term papers (3 of 114), papers (2 of 114), reviews (2 of 114), an abstract (1 of 114), a research project (1 of 114) and a preface (1 of 114). This diversity shows that scientific communication in the field is not restricted to journal articles, but also includes other genres.

4.4. Location of the Institution connected to the Authors of the Citing Texts

The most cited article from the Brazilian journal (BJA) was published in Portuguese by one single author from a public higher institution in Brazil. The Chilean one (CJA) was published in Spanish by seven authors affiliated with Chilean, Brazilian, Colombian and U.S. universities. The most cited article from the South African journal (SAJA) was published in English by three authors affiliated with South African and Danish universities.

In Figure 1, the location of the institution connected to the authors of the citing texts of BJA is represented by quotations marks:

Source: Created by the authors with Google My Maps (2022).

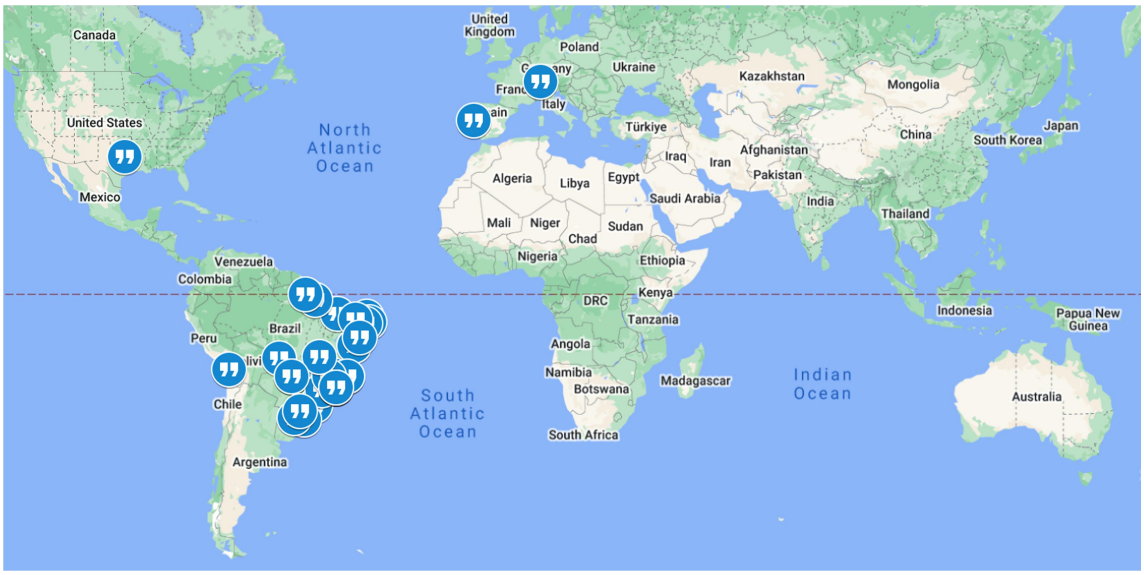

Figure 1 – BJA (n=33): Location of the institution connected to the authors of the citing texts

The authors of the citing texts of BJA are affiliated with 32 universities in Brazil, Chile, Germany, Portugal and the United States. In Brazil, the authors of the citing texts come from 24 public and private universities. As we have said, this article published in 2016 had received 33 citations by September 2021, the collection date.

In Figure 2, the location of the institution connected to the authors of the citing texts of CJA is represented by quotations marks:

Source: Created by the authors with Google My Maps (2022).

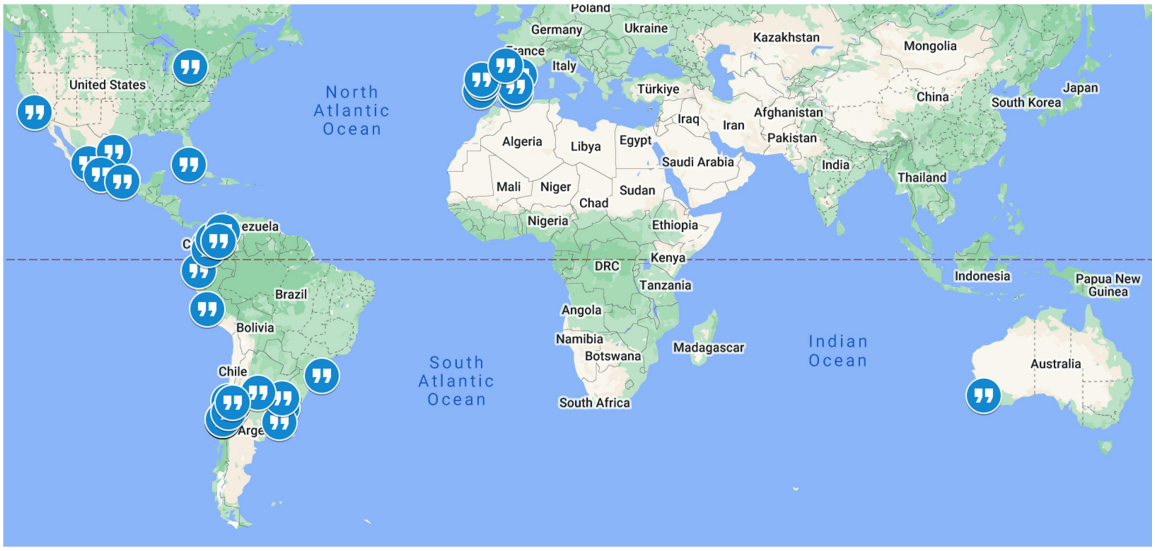

Figure 2 – CJA (n=53): Location of the institution connected to the authors of the citing texts.

The authors of the citing texts of CJA are affiliated with 48 universities in Argentina, Australia, Brazil, Canada, Chile, Colombia, Cuba, Ecuador, Spain, the United States, Mexico and Peru. This article published in 2016 had received 53 citations by the collection date, the highest number of citations in the set.

In Figure 3, the location of the institution connected to the authors of the citing texts of SAJA is represented by quotations marks:

Source: Created by the authors with Google My Maps (2022).

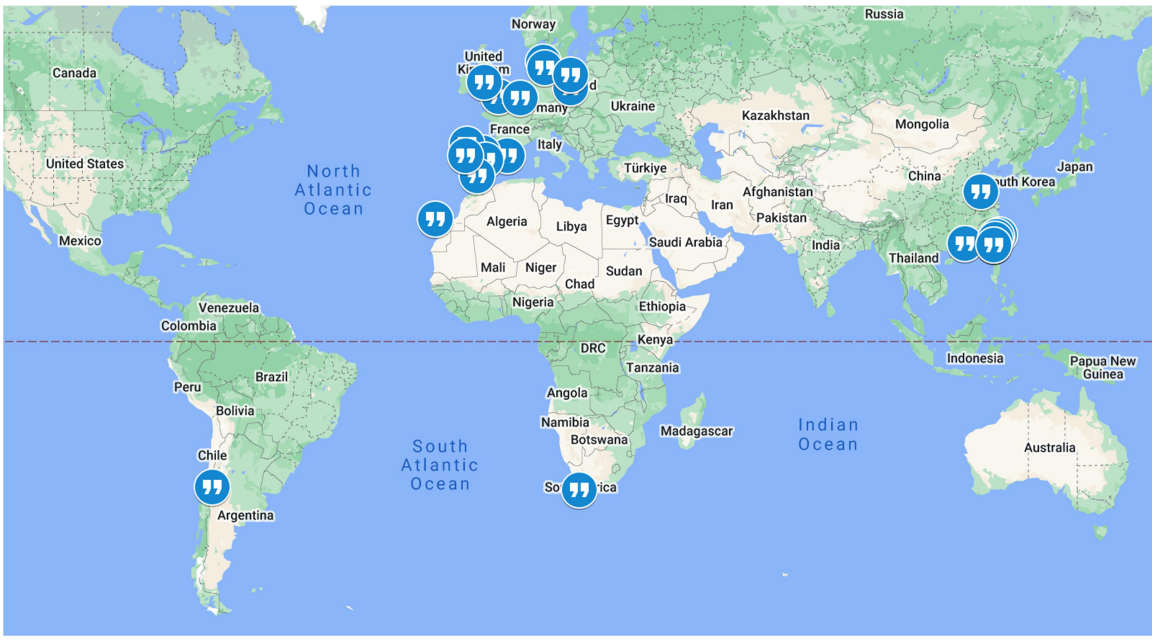

Figure 3 – SAJA (n=28): Location of the institution connected to the authors of the citing texts.

The authors of the citing texts of SAJA are affiliated with 25 universities in Belgium, Chile, China, Denmark, South Africa, Spain, Poland, Portugal, United Kingdom and Taiwan. This article published in 2017 had received 28 citations by September 2021.

Some critics might say that the more restricted circulation of the BJA would be related to the language in which the text was published (Brazilian Portuguese). However, the language issue was not a barrier to this journal article being cited in other English, Spanish and German texts, although with much lower frequency. Besides the linguistic expertise of the researchers of the citing texts, we remark that improved automatic web translators and other tools have been supporting scientific production in reading and writing practices and also an internationalization agenda for research (Pinto et al. 2021). However, the discussion about how the most cited articles circulate and their consequent visibility cannot be conditioned only to the language issue in which this text is published, as we will discuss in the following findings and final considerations (Lillis and Curry 2010; Angermuller and Hamann 2019; Hyland and Jiang 2019).

The Chilean journal article (CJA) was written in Spanish, in a multi-authored work involving seven researchers affiliated with Chilean, Brazilian, Colombian, and North American universities. The Spanish language is among the most spoken languages in the

world in terms of native speakers. The article was quoted in Spanish, but also in English and in Portuguese. The number of authors involved in this scientific production, their insertion in the disciplinary field, the diversity of the location of the institutions to which these authors are affiliated, and the theme of the article (historical and contrastive review of studies in the field) are also aspects that corroborate the high number of citations of CJA and its circulation beyond the countries of the Americas.

The South African journal article (SAJA) was written in English by three authors connected to universities in South Africa and Denmark. The citing texts were published mainly in English, but also in Spanish, Norwegian and Danish. This most cited article in South Africa was published one year after the other most cited articles selected in this study (BJA and CJA). Although it has the lowest number of citations (28 against 33 of BJA and 53 of CJA), it appears geographically distributed in more locations, among researchers from Africa, America, Europe, and Asia. Again, the English language issue would be the most obvious justification for this transnational circulation that reaches researchers from Asian countries (a feat not found in the other most cited articles of the set). However, when looking at the curriculum of one of the authors of SAJA, collaboration with a university in China is verified, which can be understood as a factor of his insertion in Asian countries, beyond the academic interest of the topic discussed in the article.

5. Findings and Final Considerations

Bibliometric indexes are rational and objective; they serve to evaluate the scientific, economic, and cultural performance of researchers. Their use in this evaluation, however, generates problems of distortion. The differences between fields of knowledge are not considered (with several prejudices to the Humanities), as well as the differences between researchers of different ages, countries, native speakers of certain languages and genres. A second distortion is in the very conception of what knowledge production is, which can be confused with the number of citations. In this sense, we can assume, with Angermuller and Hamann (89), that academia “is subject to a regime of ‘discursive capitalism’ that allows few members of a community to occupy subject positions which are made visible in the discursive practices of the many members of the Community”. Finally, it should be noted

that, as these indexes are used for career evaluation and for granting research funding, there is a direct impact on scientific development.

This reality has caused discussions in the scientific community, as reported by Marques (2013) in an article for "Pesquisa Fapesp", the scientific journal of the Foundation for Research Support of the State of São Paulo, from Brazil. From this stems, according to Marques (2013), the proposal of the traditional and reputable Dutch higher education institution Utrecht University, in the Netherlands: instead of bibliometric indicators, the evaluation would consider parameters such as quality of teaching, commitment to work in teams and willingness to share research data.

In this context, we conceive as timely and relevant the discussion brought to this work, built from the objective of identifying and analyzing aspects that are not treated or cannot be visualized in the bibliometric indexes of journals and articles in the Humanities field, specifically in the Linguistics field. By analyzing the citing texts considering (i) the year of publication, (ii) the language of publication, (iii) the materialized genres of discourse, and (iv) the location of the institutions to which the authors are affiliated, we were able to demonstrate why the culture of valuing bibliometric indexes should be interrogated and why caution should guide the way we treat and use the citation index, as advocated by Angermuller and Hamann (2019).

We can highlight the contributions that our study brings to the investigated problem. The first is that it has demonstrated the inconsistencies in the counting of citations of articles, which are determined automatically by a computer program. Even though the Google Scholar database warns the user by mentioning that the citation count is approximate, clarity about this inaccuracy seems important to the discussion about bibliometric indexes. As the analysis showed, with the removal of the inconsistencies, there was a loss of up to 24.3% among the 114 citing texts: (i) 42 citations are attributed to the BJA, when it is actually 33 (78.6% of the total detected; loss of 21.4% of the texts originally indicated as citing); (ii) 70 citations are attributed to the CJA, when it is actually 53 (75.7% of the total detected; loss of 24.3% of the texts originally indicated as citing); (iii) 30 citations are attributed to the SAJA, when it is actually 28 (93.3% of the total detected, loss of 6.7% of the texts originally indicated as citing).

The second and the most important contribution is found in what it is not possible to show based on absolute numbers of bibliometric indexes, as we will recapitulate below.

The time difference between the publication date of the most cited article and the publication date of citing texts is at least two years. This is a characteristic of the disciplinary culture, which requires this minimum time for a text to be read, discussed, cited, published, and circulated in the field. This time differs from that of other disciplinary domains, in which the published text is immediately “absorbed” into the knowledge production process.

The language most used in the publication of the citing texts of the most cited articles is Spanish. Lillis and Curry (120) speak of an ideology according to which English would be used as a lingua franca and of scientific communication: “The idea that knowledge has a universal value and should be generated and shared across the world” [...] and this sharing would take place according to a “global academic utopia” in which English appears as the lingua franca. It is interesting to show the prominence of Spanish and its circulation around the world, considering the location of the institutions connected to the authors of the citing texts. It is also important to point out that, in the general set of this study, Spanish is the predominant language of the texts cited in Linguistics (48 out of 114, which corresponds to 42.1% of the total amount), followed by English (34 out of 114, which corresponds to 29.8%) and Portuguese (29 out of 114, which corresponds to 25.4%).

Although the demand for English in scientific production in the field of linguistic studies is recent, little by little this practice is being increasingly taken up, as a result of the strength of institutional internationalization coercions constituted by funding agencies, universities, and graduate programs. From a linguistic perspective, Hyland and Jiang (3) point out that today's scientific English is “an unadorned, stripped down communicative medium that bears little resemblance to the flowery prose of Boyle or Hooke in the 1600s, or even the carefully coy phrasings of Crick and Watson in the 1950s”. From a sociocultural and discursive perspective, as highlighted by Curry and Lillis (2018) and Corrêa (2020), local and regional interests in scientific production are preempted, towards a globalized pattern of scientific making.

About the discourse genre of the citing text, our research showed the prevalence of the scientific article genre (63.3% in the case of the citing texts of BJA; 66% of CJA;

85.7% of SAJA). To explain the predominance of the article genre in the data set, we recall Hyland (2006), who refers to the existence of epistemic conventions of the disciplines in the discursive construction of knowledge, regarding the ways of formulating and circulating knowledge. Thus, the journal article genre, in academic practices, embodies a form of socialization of scientific production, seeking to contribute to the advancement of new studies. In the field of linguistic studies, according to the data analyzed, the average time between the publication of the cited article and the publication of the text citation is at least two years. As can be seen, the preference for the genre of journal article is present not only in the field of Humanities and Linguistics, but also in different fields of knowledge. It is worth considering that the quality of a journal article, in journals that have experts from different institutions in their editorial board, is evaluated by a peer review system, a condition that is an effect of the coercion systems of scientific production, also in dialogue with other fields of knowledge.

As presented in these results, the disciplinary culture of the field shows, however, the appropriation of other genres in scientific communication, although in a lower occurrence, ranging from 8 to 1 of 114 citing texts, presented here in descending order of occurrence: book chapters, Ph.D. theses, Master's dissertations, work in annals, term papers, papers, reviews, abstracts, research projects and prefaces. As Hyland discussed (*Disciplinary* 21), "writing as a member of a discipline involves textualizing work in a way that colleagues can see as 'doing biology' or 'doing sociology'". In this case, the data survey shows other ways of "doing linguistics" beyond the production of scientific articles.

It is also worth noting that the location of the institutions to which the authors of the citing texts are affiliated shows a mode of circulation of the most cited articles; thus the language of the most cited article can expand or reduce this circulation. But the language issue, which would be the most obvious (in the light of the maxim that, by publishing in English, the article will have wide circulation and will be read), is not enough, as we see in our analysis about invisible aspects in the citation indexes of a scientific article.

Among them, we list the journal's h5 index. As mentioned, bibliometric indexes of CiteScore (from the Scopus database), Impact Factor (from the Web of Science database) and h5-index (from Google Scholar) are factors that promote the visibility of scientific production in search engines. In this sense, we recall the indexes of the three journals from

which the three articles were collected: BJA: h5-index 13, h5-median 19; CJA: h5-index 14, h5-median 20; SAJA: h5-index: 8, h5-median: 10.

In addition, it is worth considering the author's prominence in the disciplinary field (the “celebrity” factor addressed by Angermuller and Hamann, 2019). The author of BJA is a researcher in the field with h-index 36. The authors of CJA are also researchers in the field with h-index ranging from 9 to 56 (for the most experienced). Among the authors of SAJA, the most experienced has an h-index 36. In other words, it is the citation of an article by authors who are already recognized in the field of knowledge, according to the logic of “discursive capitalism”, as Angermuller and Hamann (2019) name it.

To all these factors, we lastly add the subject that is the focus of the article. In the case of BJA, it is emoji language. This journal article was mentioned in citing texts in Linguistics but its insertion in other fields draws attention: Administration, Communications and Semiotics, Design, Education (Physical Education, Teaching of Sciences and Mathematics Education), Computers, Business, Advertising, Public Relations and Tourism. Certainly, the novelty and timeliness of the article's theme and the interest of other disciplines in it are aspects that deserve to be considered in the discussion.

CJA presents a revision of the state of the art of the field in Latin America. The historical character of the discussion developed by the work seems to be one of the elements responsible for its citation in works published in countries in America and in other continents.

SAJA houses discussions about the use of tools for acquiring L2 and the challenges for the field. Our observations demonstrated that the country of origin of the publication does not play an important role in understanding the citation behavior of the article. However, there are three factors that matter when understanding citation behavior: the topicality, the strong interest in the issue of the use of technologies in second language acquisition, and that one of the researchers is affiliated with universities in different parts of the world.

As we have seen, article visibility, measured by citation index, is, in fact, the result of different aspects, taken together: the disciplinary culture of the field with respect to time, language, discourse genre, distribution of institutional affiliation of the authors of citing

txts; the h5-index and the h5-median of the journal; the h-index of the researcher; the thematic in the field of knowledge.

Other aspects can still be investigated, in order to expand the knowledge about this theme even further. Among them, we list, for future exploration, the way the cited article is appropriated by the citing texts, considering its role in the construction of the argumentative dimension of scientific discourse.

Acknowledgements

We thank the *National Council for Scientific and Technological Development* - CNPq, Brazil (process 309765/2019-6) and the *Coordination for the Improvement of Higher Education Personnel* – Brazil (CAPES-PrInt-Unesp Program, process 88887.310463/2018-00) for the support and help granted to the authors.

REFERENCES

Angermuller, Johannes and Julian Hamann. "The celebrity logics of the academic field. The unequal distribution of citation visibility of Applied Linguistics professors in Germany, France, and the United Kingdom". *Zeitschrift für Diskursforschung*, núm. 1, 2019, pp. 77-93.

Ampudia de Haro, Fernando. "O impacto de (não) ter impacto: Para uma sociologia crítica das publicações científicas". *Revista Crítica de Ciências Sociais*, núm. 113, 2017, pp. 83-106.

Araújo, Claudio Gil Soares de and Sardinha, Aline. "Índice-h dos artigos citantes: uma contribuição para avaliação da produção científica de pesquisadores experientes". *Revista Brasileira de Medicina do Esporte*, vol. 17, núm. 5, 2011, pp. 358-362.

Canagarajah, Athelstan Suresh. *A Geopolitics of Academic Writing*. University of Pittsburgh Press, 2002.

Corrêa, Manoel Luiz Gonçalves. "A inter-incomprensão polêmica e sua versão solipsista em práticas de leitura emergentes." *ComHumanitas: revista científica de comunicación*, vol. 11, 2020, pp. 68-81.

Curry, Mary Jane and Theresa Lillis. "Uma crítica ao 'inglês como língua franca' da publicação de periódicos acadêmicos". *Roseta*, vol. 1, núm. 1, 2018. Retrieved 16 October 2022, from https://www.roseta.org.br/2018/05/21/uma-critica-ao-ingles-como-lingua-franca-da-publicacao-de-periodicos-academicos/

Droescher, Fernanda Dias and Edna Lucia da Silva. "O pesquisador e a produção científica". *Perspectivas em Ciência da Informação,* vol. 19, núm. 1, 2014, pp. 170-189.

Hammarfelt, Björn. "Beyond Coverage: Toward a Bibliometrics for the Humanities". Ochsner, Michael and Hug, Sven E. and Daniel, Hans-Dieter (ed.). *Research*

Assessment in the Humanities: Towards Criteria and Procedures. Springer Open, 2016, pp. 115-131.

Hirsch, Jorge Eduardo. "An index to quantify an individual's scientific research output". *Proceedings of the National academy of Sciences,* vol. 102, núm. 46, 2005, pp. 16569-16572.

Hyland, Ken. "Disciplinary Differences: Language Variation in Academic Discourses." *Academic Discourse Across Disciplines*, edited by Ken Hyland and Marina Bondi. Peter Lang, 2006, pp. 17-45.

Hyland, Ken. "English in the Disciplines: Arguments for Specificity". *ESP Today*. vol. 5, núm. 1, 2017, pp. 5-23.

Hyland, Ken and Feng (Kevin) Jiang. *Academic Discourse and Global Publishing*. Routledge, 2019.

Ioannidis, John P. A. et al. "Bibliometrics: Is your most cited work your best?" *Nature*, vol. 514, núm. 7524, 2014, pp. 561-562.

Ioannidis, John P. A. and Boyack Kevin and Wouters, Paul F. Citation Metrics: A Primer on How (Not) to Normalize. *PLoS Biol*, vol. 14, núm. 9, 2016.

Lillis, Theresa and Mary Jane Curry. *Academic Writing in a Global Context*. Routledge, 2010.

Marcovitch, Jacques, org. *Repensar a Universidade II*. Com-Arte, Fapesp, 2019.

Marques, Fabrício. "Os limites do índice-h". *Revista Fapesp*, num. 207, 2013, pp. 35-39.

Martín-Martín, A., Thelwall, M., Orduna-Malea, E. and López-Cózar, E. D. "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations'COCI: A multidisciplinary comparison of coverage via citations". *Scientometrics*, vol. 126, núm. 1, 2021, pp. 871-906. Retrieved 19 October 2022, from https://link.springer.com/article/10.1007/s11192-020-03690-4#citeas.

Pinto, Paula Tavares, Daniela Nogueira de Moraes Garcia and Douglas Cunha dos Santos. "Ações de internacionalização para o ensino e a pesquisa na área de línguas". *Fórum Linguístico*, vol. 18, núm. 1, 2021, pp. 5618-5630.

Purvis, Andy. "The h index: playing the numbers game". *Trends Ecol. Evol*., vol. 23, núm. 8, pp. 442, 2008.

Salgado, Luciana Salazar and Letícia Moreira Clares. "Mediação editorial em artigos científicos: um estudo de injunções e apagamentos nas humanidades". *Revista do GEL*, vol. 14, núm. 3, 2017, pp. 29-58.

Santos, Raimundo Nonato Macedo dos and Nair Yumiko Kobashi. "Bibliometria, cientometria, infometria: conceitos e aplicações". *Pesquisa Brasileira em Ciência da Informação*, vol. 2, núm. 1, 2009, pp. 155-172.

Sepúlveda-Vildósola Ana Carolina et al. "Scientific Publications During the COVID-19 Pandemic". *Arch Med Res*., vol. 51, núm. 5, 2020, pp. 349-354.

Thomaz, Thomaz, Petronio Generoso, Renato Samy Assad and Luiz Felipe P. Moreira. "Uso do fator de impacto e do índice H para avaliar pesquisadores e publicações". *Arquivos Brasileiros de Cardiologia,* vol. 2, núm. 96, 2011, pp. 90-93.

VERBI Software. *MAXQDA 2021*. 2021. Software, https://www.maxqda.com.