

Accesses to ETDs in languages other than Portuguese @PUC–Rio: a different view

Ana M. B. Pavani

IEEE Member

Pontifícia Universidade Católica do Rio de Janeiro, Brazil

apavani@puc-rio.br

Abstract

Portuguese is among the 10 world languages with the most native speakers; it is the third Western language in the ranking. At the same time, it is neither the *lingua franca* of science and technology, nor of the Internet. In Brazil, in the last decade, many institutions started allowing ETDs to be published in languages other than Portuguese; English is the most used. This work addresses characteristics of the ETD collection of a university in Brazil in terms of the languages of the works. An overview of the numbers/percentages of ETDs in Portuguese and other languages is presented in the time frame 2008–2018. Some aspects of the accesses are discussed too, but they are from 2016 on since before this year the numbers of ETDs in English were very low. The numbers of ETDs in foreign languages – numbers of institutions and of works – in other Brazilian institutions that contribute to the Brazilian national consortium are presented and discussed too.

Keywords: ETDs in Portuguese, ETDs in English, accesses, Portuguese speaking countries

Introduction

Electronic Theses and Dissertations (ETD) are very important products of graduate education. There are ETD Programs in institutions all over the world and ETDs are written in many different languages. A quick look at ND LTD – Global ETD Search (<http://search.ndltd.org/>) shows that there are over 5.8 million works whose metadata records come from almost 200 providers. A provider is an institution that has an ETD Program and adopts OAI-PMH – Open Archives Protocol for Metadata Harvesting

(<https://www.openarchives.org/pmh/>) to share the metadata records of its collection. It is important to remark that the protocol applies to any digital collection not only to ETDs.

The Global ETD Search database hosts ETDs in 14 languages. The three languages with the highest number of works are English (almost 1.4 million), Taiwanese Chinese (almost 0.99 million) and Portuguese (almost 0.77 million). Richardson et al. (2008) addressed knowledge discovery in digital libraries of ETDs and used NDLTD as a case study. Their concern was on finding works of interest and one of the points was language. Among the over 70 contributors of ETD records, the Brazilian ETD consortium, BDTD – Biblioteca Digital de Teses e Dissertações (<http://bdtd.ibict.br/>), ranked third in terms of numbers of records (a little over 10%). This indicates a strong presence of records of ETDs in Portuguese. An important fact about BDTD is that its database shows that the number of Brazilian ETDs and their languages. This topic will be addressed in the next section.

Some years after Richardson's work, Cayabyab (2015) addressed challenges of ETDs and mentioned global visibility. Global visibility is impacted by the possibility of discovery which can be related to language.

Many universities in countries whose one of the official languages is not English allow and even stimulate students to publish their ETDs in English – it is the language of scholarly communication. This is the situation at Pontifícia Universidade Católica do Rio de Janeiro (PUC–Rio) in Rio de Janeiro, Brazil. In 2008, PUC–Rio started allowing ETDs to be published in languages other than Portuguese without a special authorization. Currently, there are ETDs in English, French and Spanish, besides Portuguese, naturally.

The Maxwell System (<https://www.maxwell.vrac.puc-rio.br/>) is the Institutional Repository (Lynch, 2003) of Pontifícia Universidade Católica do Rio de Janeiro (PUC–Rio). The system is the publishing platform for scholarly communication, online educational resources, research data, journals etc. The system yields many data that can be used to monitor the publishing of ETDs, as well as other digital resources, by language and year. Accesses are recorded too, allowing the tracking of where they come from and when they happen.

ETDs @PUC–Rio and at Other Brazilian Institutions

The ETD program of PUC–Rio started in 2000 – the first ETD was published in May. ETDs became mandatory in August 2002. The ETD program has been under constant enhancement and two points will be examined in this work:

- (1) The evolution of the adoption of ETDs in languages other than Portuguese. The focus will be all ETDs and ETDs in the three areas that offer graduate programs. The objective is to track evolution and to compare how the areas have adopted other languages for ETDs.
- (2) The impact of the publishing in foreign languages on the accesses of ETDs on the system. In this case, all ETDs will be examined and ETDs of six graduate programs will separately be

analyzed. Details on the selection of the graduate programs will be presented in a following section.

To examine these points, it is necessary to have data. Data are gathered from the Maxwell System. The Maxwell System has many programs that organize data from the database to yield information on both the production (publishing) of digital contents and on the accesses they have. Access data are obtained by mining data from the Apache server and storing them on the database. This process was described in a work of a previous ETD Symposium (Pavani, 2017). The same work addressed the first point in a very limited way. The second was superficially mentioned in the extended abstract of another ETD Symposium (Pavani, 2018).

Other institutions that contribute to BDTD allow ETDs to be published in foreign languages too; there are works in English, French, Italian, Portuguese and Spanish. The profiles of the collections compared to PUC-Rio's will be addressed later in this work.

The subsections that follow address the publishing of ETDs in foreign languages both at PUC-Rio and at other institutions. The next section presents accesses to PUC-Rio's ETDs analyzing the patterns by the languages. Access data are not available on the BDTD system and institutions record accesses in different ways on their repositories.

ETDs in Languages Other Than Portuguese @PUC-Rio From 2008 On

PUC-Rio offers graduate programs in three areas that are under the responsibility of three university centers: (1) CCS - Center for Social Sciences; (2) CTC - Center for Science and Technology; and (3) CTCH - Center for Theology and Humanities.

After a thesis or dissertation is defended, the student is granted a deadline that ranges from one to six months after this event to submit the final version. The committee decides the deadline based on the number and depth of changes the work requires. In some special cases a longer time may be authorized.

Since 2008 was the year when foreign languages started being allowed with no special authorization, this work focuses on ETDs presented from 2008 to 2018 - an interval of 11 years. ETDs presented in 2019 have already been published but, since the number is still less than 10% of the average of the last 5 years, they are not considered in this work. Since data are available for all ETDs, table 1 presents the numbers per language in the interval and in the complete collection. The complete collection contains theses & dissertations that were defended before 2000 and were digitized as well as ETDs presented in 2019. The Graduate Program in Electrical Engineering digitized all theses & dissertations and Mechanical Engineering is currently doing it.

	Complete collection	2008–2018
English	542 (4.838%)	527 (7.741%)
French	2 (0.018%)	2 (0.029%)
Portuguese	10,646 (95.037%)	6,268 (92.068%)
Spanish	12 (0.107%)	11 (0.162%)
Total	11,202 (100.00%)	6,808 (100%)

Table 1

Numbers and percentages of ETDs in different languages as of August 31, 2019.

The Maxwell System makes available time series of published materials both by language and in all languages. This information can be grouped or separated by course/graduate program. Data may be combined to generate time series by Center too. The information on the system comes in tables and in most cases with additional graphics.

The use of the same time series, limited to the interval observed in this work, allows the generation of the graphics of figures 1–4. The graphics show the percentages of ETDs in Portuguese and in English from 2008 to 2018 in four different sets of ETDs: (1) PUC-Rio – complete collection; (2) CCS – Social Sciences; (3) CTC– Science and Technology; and (4) CTCH – Theology and Humanities. The numbers for French and Spanish were not plotted because they are insignificant (less than 1%).

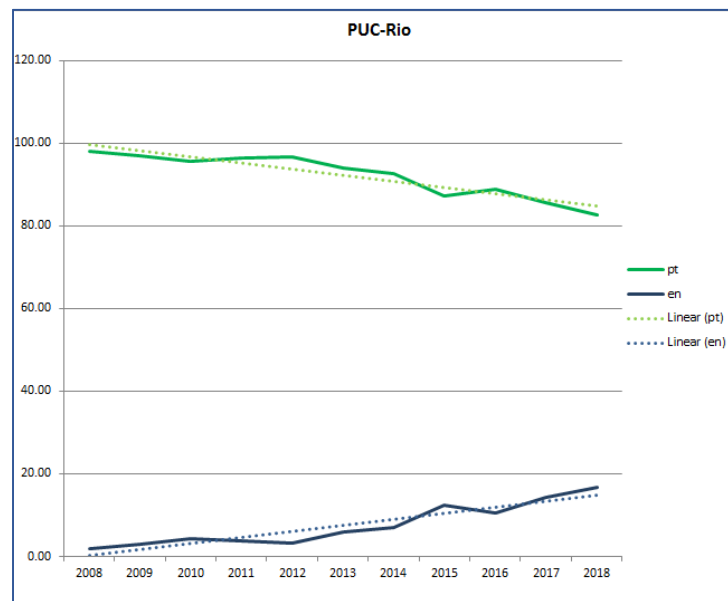


Figure 1

Percentages of ETDs in Portuguese and English – all ETDs as of August 31, 2019.

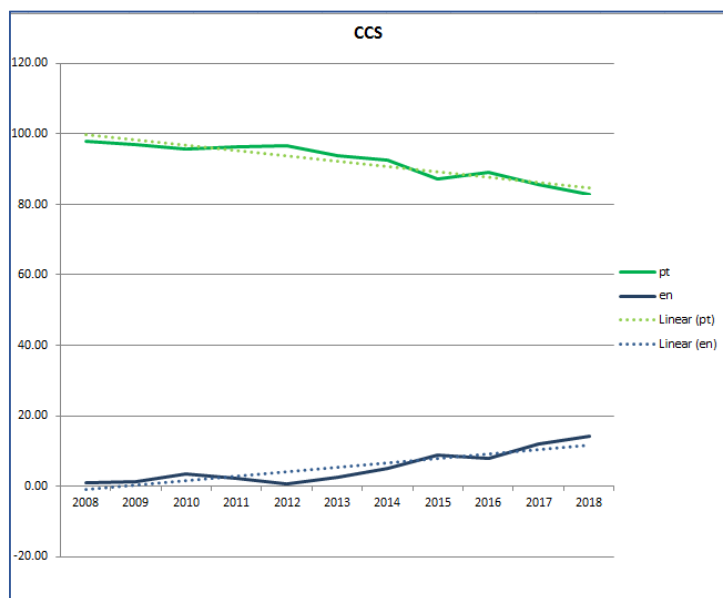


Figure 2

Percentages of ETDs in Portuguese and English – Social Sciences as of August 31, 2019.

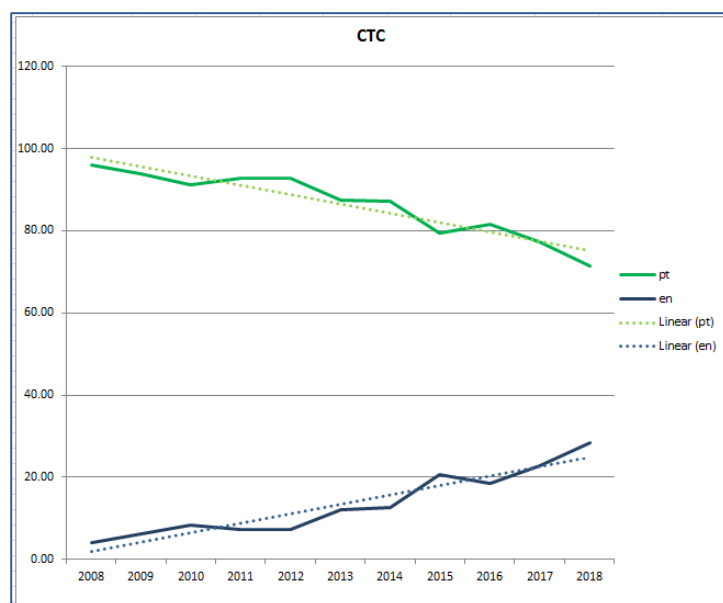


Figure 3

Percentages of ETDs in Portuguese and English – Science and Technology as of August 31, 2019.

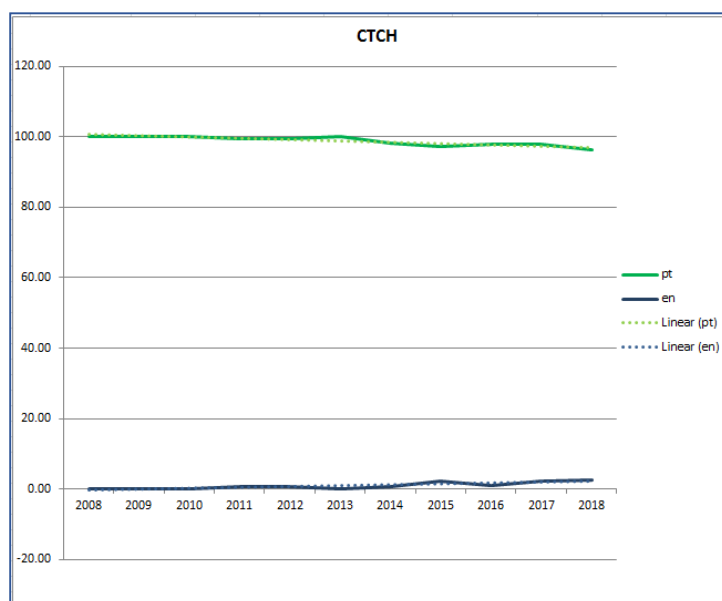


Figure 4

Percentages of ETDs in Portuguese and English – Theology and Humanities as of August 31, 2019.

The curves in figures 2–4 are quite different. It is clear that Science and Technology presents a much steeper descent for Portuguese than the other two. Theology and Humanities has a very slow change. Social Sciences has a pattern that is similar to the complete collection. The examination of figures indicates that the behaviors of the three centers have been different since 2008. Table 2 shows the percentages of ETDs in English in 2008 and in 2018.

The percentages of ETDs in English in 2008 and 2018 are shown in table 2.

	2008	2018
PUC-Rio	2.01%	17.36%
CCS	1.12%	14.08%
CTC	4.01%	28.32%
CTCH	0.00%	2.62%

Table 2

Percentages of ETDs in English by area in 2008 and 2018 – 2018 percentages as of August 31, 2019.

The numbers in table 2 and the graphics in figures 1–4 show that Science and Technology is the area that has adopted ETDs in English with the most enthusiasm. At the same time, they indicate that the percentages of ETDs in English are increasing.

ETDs in Languages Other Than Portuguese at Other BDTD Member Institutions

In order to gather data about BDTD member institutions, it was necessary to access the union catalog and perform some searches. Total numbers of metadata records from each institution are quite easy to find. On the other hand, finding information on the languages of

the works requires searches and there are different codes for the same language language.

There are many Brazilian institutions that allow ETDs in languages other than Portuguese. Make graduate work results available worldwide is a concern among them and works in English are one of the tools to achieve this. The union catalog hosts metadata records for ETDs in English, French, Italian, Portuguese, Russian and Spanish. As in the case of PUC-Rio, the foreign language with a significant number of works is English.

The compilation of the search results performed on August 31, 2019 yielded the numbers in table 3 and in figures 5-8.

	Number	Percentage
All Programs	115	
Programs with ETDs in English	61	53.043 %
Programs with ETDs in Spanish	35	30.435 %
Programs with ETDs in French	16	13.913 %
Programs with ETDs in Italian	1	0.870 %
Programs with ETDs in Russian	1	0.870 %

Table 3

Institutions with ETD Programs that Accept ETDs in Languages Other Than Portuguese.

Table 3 leaves no doubt about English being the most popular foreign language. For this reason, it is the language that will be focused in this work along with Portuguese. At the same time, the number of contributing institutions with ETDs in English is a little over 50% of all institutions. The numbers and percentages of ETDs in English in the 61 collections are quite different. Figures 5 and 6 show, respectively the numbers and percentages of ETDs in English (horizontal axis) and the numbers of ETD programs.

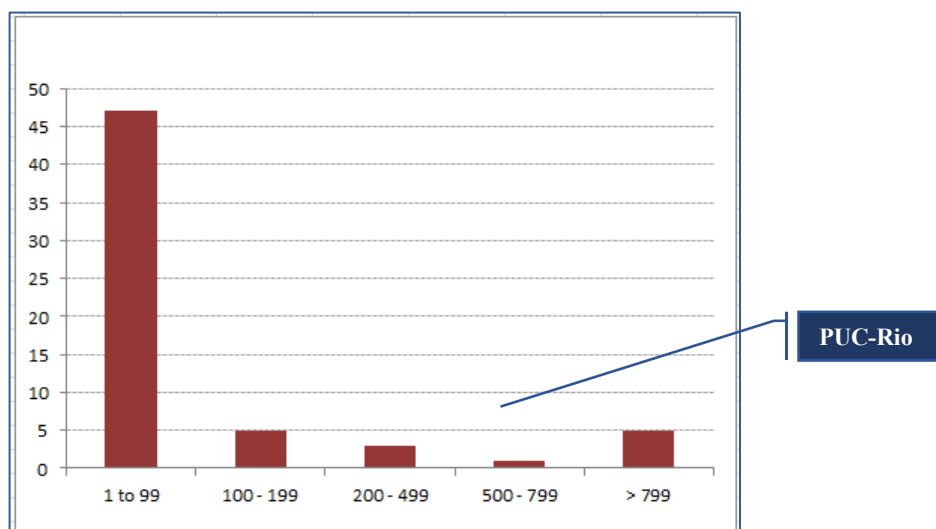


Figure 5

Numbers of ETD Programs (vertical) by Numbers of ETDs in English (horizontal).

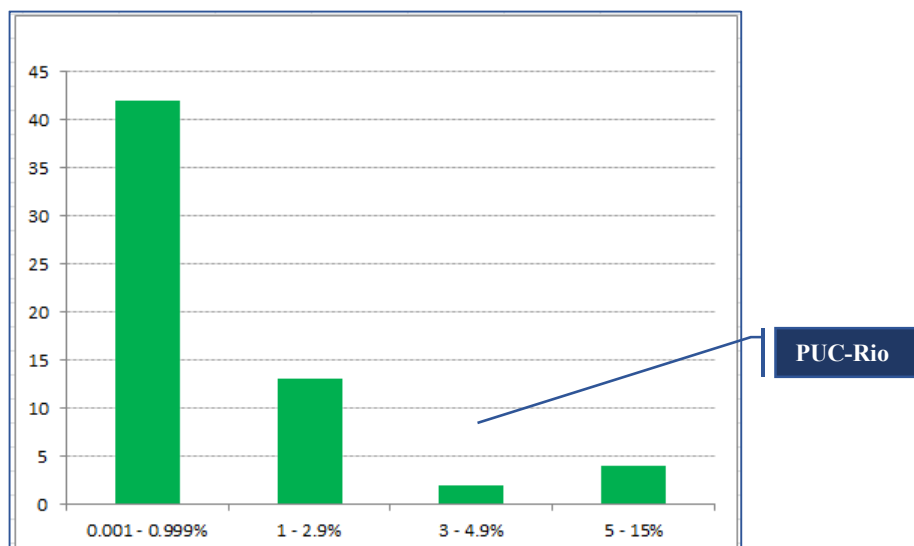


Figure 6

Numbers of ETD Programs (vertical) by Percentages of ETDs in English (horizontal).

Figure 6 shows that the percentages of ETDs in English in the collections is still very low – most ETD programs have less than 1% of ETDs in English. Many do not have ETDs in other languages. Figures 5 and 6 can be better understood if some facts about BDTD are mentioned.

It was created in 2001 with three institutions that had ETD programs at that time. The institutions were PUC–Rio, UFSC – Universidade Federal de Santa Catarina and USP – Universidade de São Paulo. PUC–Rio and USP are located in the Southeastern Region of Brazil and UFSC in the Southern Region. The three institutions have solid graduate programs. One of the objectives of BDTD was to disseminate ETDs and digital publishing to all regions. Many institutions were new in graduate education but were welcome to join. As time went by, 115 institutions joined BDTD. They are from all regions of the country and their graduate programs have very different profiles. Figure 7 shows the numbers of institutions in the different regions.

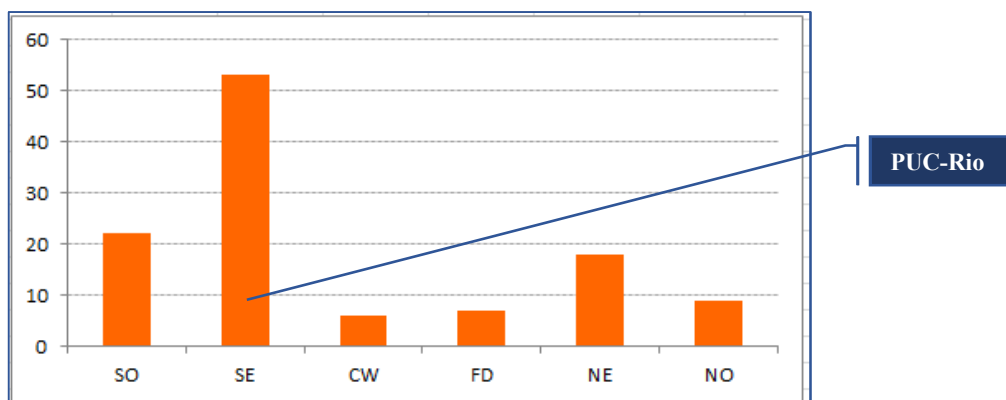


Figure 7

Numbers of ETD Programs (vertical) by Region of Brazil.

The ETD collections have different sizes. Figure 8 shows the numbers of ETD programs by the sizes of the collections.

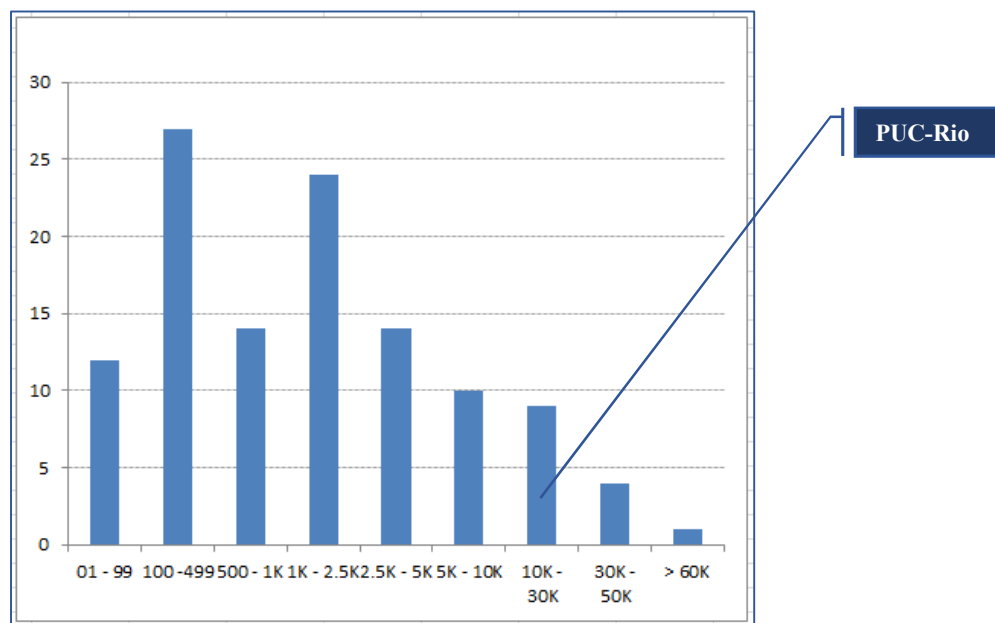


Figure 8

Numbers of ETD Programs (vertical) by Collection Size.

Accesses to ETDs in Different Languages @PUC-Rio

The main objective of writing ETDs in English is making them more internationally accessible. For this reason, understanding the relations among accesses and languages is part of this paper. In the work presented at ETD 2017 (Pavani, 2017), accesses were addressed in a very simple manner and the numbers of ETDs in English were very low at that time.

In the two years since this first work, the subject was addressed with more accuracy in ETD 2018 (Pavani, 2018); an extended abstract was published and a dataset made available (<https://doi.org/10.17771/PUCRio.ResearchData.34785>). In this extended abstract, accesses were considered for all ETDs. It was a simple analysis too. Now a more accurate analysis is made.

Decisions on Data to be Examined

In order to get a better understanding of the impact of using English on the patterns of accesses (numbers and countries), new programs were developed to extract data to easily yield more meaningful information. The criterion for the selection of the Graduate Programs to be examined was based on the following premises:

- (1) Accesses in 2016, 2017, 2018 and 2019 (partial results) for works published in a window three years ending in the year of measurement.
- (2) All graduate programs as a complete set (ALL).

(3) Individual graduate programs with at least 20 ETDs in English representing a minimum of 10% of the total number in the graduate program.

Condition (1) was chosen so that research topics were current to the timeframe of observation. Some graduate programs have digitized all a part of their theses & dissertations, so some works can be considered “old”.

The graduate programs that fulfilled condition (3) were: Economics (ECO), Electrical Engineering (ELE), Informatics (INF), International Relations (IR), Mathematics (MAT) and Mechanical Engineering (MEC).

Data were collected from the database using the programs and were organized in a dataset (<https://doi.org/10.17771/PUCRio.ResearchData.38014>). Data in this dataset can be viewed in many different ways and two were chosen to be addressed in this work. Others will be subject of future analysis and publication.

Examination of the Data on Accesses

The first “view” of data is related to the Portuguese-speaking countries that are in the top 10 in numbers of accesses. Table 4 shows the ranks of the countries for each ETD set, each language and each timeframe.

	2016			2017			2018			2019(*)		
	all	pt	en	all	pt	en	all	pt	en	all	pt	en
ALL												
Brazil	1	1	1	1	1	1	1	1	1	1	1	2
Angola	9	9		7	7		7	7		6	6	
Portuga	7	7	10	6	6		6	6		4	4	
Mozam	8	8		5	5		5	5		5	5	
INF												
Brazil	1	1	1	1	1	2	1	1	1	2	1	2
Angola		10			9		10	5			9	
Portuga	9	8		7	5		9	6		9		
Mozam		9		10	7			10			9	
ELE												
Brazil	1	1	2	1	1	1	1	1	1	1	1	2
Angola	8	6		7	7		5	4		5	4	
Portuga	9	9		5	5		6	6		8	8	
Mozam	6	5		6	6		7	7		7	7	
ECO												
Brazil	1	1	1	1	1	2	1	1	2	3	2	3
Angola					9		9	6		7	5	
Portuga										10	7	
Mozam		7		8	5		8	5				
MEC												
Brazil	1	1	2	1	1	1	1	1	1	1	1	2
Angola	5	5		5	6		6	5		7	5	
Portuga	9	9	10	6	6		9	7		8	6	
Mozam	10	10		8	8		7	6		6	4	
MAT												
Brazil	1	1	1	1	1	1	1	1	1	1	1	3

Angola	9	9		7	7		8	7		9	9	
Portuga	3	3		5	5		5	5		6	5	
Mozam	5	4		3	3		3	3				
IR												
Brazil	1	1	2	1	1	1	1	1	1	1	1	2
Angola	6	6		6	6		6	6		4	4	
Portuga	9	9		7	7		7	7	8	8	7	
Mozam	5	3		2	2		5	5		5	5	

Table 4

Ranking of Portuguese-speaking countries in the terms of accesses – different sets of ETDs and different time frames.

(*) Accesses computed until August 31, 2019.

Examination of table 4 indicates that:

(1) In almost all cases Portugal, Angola and Mozambique are among the top 10 when ETDs in all languages and in Portuguese are considered – this is an expected result since Portuguese is the language of more than 95% of the ETDs.

(2) Accesses from Brazil are at the top – 1st in 68 out of 84 cases, otherwise 2nd or 3rd.

(3) ETDs in English had no accesses from Portuguese speaking countries, except Brazil, in 2019 (as of August 31, 2019).

(4) Accesses from Brazil are not in in the first place for ETDs in English in 2019 (as of August 31, 2019).

(5) The Graduate Program in Economics shows a pattern that is quite different from the others – Portuguese-speaking countries do not play significant roles in the accesses – this is to be futher explored to be understood.

If the data set is examined, it is clear that France, Germany and the United States are always in the top 10.

The second “view” is related to the average numbers of accesses to ETDs – ALL and the six graduate programs of table 4. Numbers were normalized as specified in the dataset (<https://doi.org/10.17771/PUCRio.ResearchData.38014>) and they are organized in table 5.

	2016			2017			2018			2019 (*)		
	all	pt	en	all	pt	en	all	pt	en	all	pt	en
ALL	49.64	50.88	27.64	63.54	69.49	19.50	68.19	75.58	20.27	56.19	61.53	28.69
INF	24.85	24.96	23.55	18.72	24.62	13.77	18.64	25.01	13.47	21.57	30.05	16.45
ELE	59.84	58.77	72.20	60.08	63.27	28.83	55.46	65.85	27.98	38.57	38.73	45.52
ECO	14.83	12.25	19.05	15.68	9.97	20.40	12.24	13.68	11.97	28.08	29.39	26.72
MEC	54.69	58.83	21.79	73.14	78.83	28.26	58.92	72.22	22.57	51.28	59.53	31.10
MAT	68.28	71.81	6.50	134.74	146.03	20.14	51.16	66.00	12.67	41.71	58.14	21.84
IR	31.82	39.93	8.97	50.90	54.58	33.93	71.29	78.96	40.70	59.79	59.02	56.24

Table 5

Normalized accesses to ETDs – all collection and six graduate programs – in a three-year time frame of publication ending in the years of accesses. (*) Accesses computed until August 31, 2019.

Data in table 5 are difficult to interpret. Some characteristics can be inferred though:

(1) The numbers for 2019 are expected to be smaller since data were computed for 8 months (up to August 31, 2019).

(2) The numbers for all languages and Portuguese are close to each other due to the percentage of ETDs in Portuguese. This happens for ALL and each graduate program in all the years of observation.

(3) In 2016, 2017 and 2018 numbers for English were much lower than the corresponding numbers for Portuguese in most cases.

(4) The graduate program in Economics does not follow the pattern – accesses for ETDs in English are either higher or almost equal to the ones in Portuguese;

(5) In 2019, numbers for Portuguese and English became closer in all situations, but these are partial results.

If the two tables are considered at the same time, it is reasonable to conclude that the higher numbers of accesses are due to the three Portuguese-speaking countries. This is confirmed by the fact that ETDs in Economics have higher or similar numbers of accesses in English and they do not have Angola, Mozambique and Portugal in the top 10 as many times as the others.

Final Comments

The numbers indicate that ETDs in English do not have higher numbers of accesses than the ones in Portuguese. There are countries that yield accesses regardless of the language but Portuguese-speaking countries are the largest contributors to accesses for ETDs in Portuguese and in all languages. At the same time, ETDs in English seem to be having higher accesses averages as time goes by. The compilation of data at the end of 2019 will allow a confirmation (or not) of this trend. Time series of average accesses will be examined.

Since the percentage of ETDs in English is increasing, it will be interesting to analyze

when it becomes higher.

If accesses to ETDs in English grow higher, authors may become more enthusiastic in using this language for their ETDs. When this happens, it will be necessary to inform supervisors so that they can discuss with their students. Writing ETDs in English will bring an additional work to graduate students.

The fact that access data are not available with the same details in other Brazilian institutions is a limitation in the understanding the impact of writing Brazilian ETDs in English.

References

- Cayabyab, T. A. C., A review of emerging ETD initiatives, challenges and future demands, *International Journal of Information and Education Technology*, vol.5, nº. 2, pp. 772-777, 2015, retrieved from <http://www.ijiet.org/papers/609-D031.pdf>, on 28 April 2019.
- Lynch, C., Institutional Repositories: essential infrastructure for scholarship in the digital age, *ARL Bimonthly Report*, 2003, retrieved from <https://www.cni.org/wp-content/uploads/2003/02/arl-br-226-Lynch-IRs-2003.pdf>, on 21 March 2017
- Pavani, A. M. B., ETDs in languages other than Portuguese at PUC-Rio, ETD 2017 – International Symposium on Electronic Theses and Dissertations, available online in Open Access at <http://www.ocs.usetda.org/index.php/NDLTD/ETD2017/paper/viewFile/95/52>.
- Pavani, A. M. B., Revisiting ETDs in languages other than Portuguese @PUC-Rio, ETD 2018 – International Symposium on Electronic Theses and Dissertations, extended abstract available online in Open Access at https://etd2018.ncl.edu.tw/images/phocadownload/7-3_Ana_Pavani_Extended_Abstract_ETD_2018.pdf.
- Richardson, W. R., Srinivasan, V. and Fox, E. A., Knowledge discovery in digital libraries of theses and dissertations: an NDLTD case study, *International Journal of Digital Libraries*, vol.9, pp. 163-171, 2008. DOI 10.1007/s00799-008-0046-9.