

## INTERNATIONAL VISIBILITY OF ETDs IN ENGLISH AND IN PORTUGUESE ON A BRAZILIAN REPOSITORY

### 01. Background and Introduction

Portuguese is the 9<sup>th</sup> most spoken language (4<sup>th</sup> Western language) in the world according to Berlitz (<https://www.berlitz.com/blog/most-spoken-languages-world>) or the 8<sup>th</sup> (4<sup>th</sup> Western according to Statista (<https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/>)). Both numbers mean that Portuguese has a large number of speakers. The Comunidade de Países de Língua Portuguesa (<https://www.cplp.org/>) lists nine countries that have Portuguese as one of the official languages – Angola (36.7M), Brazil (216.4M), Cape Verde (0.6M), Guinea-Bissau (2.2M), Equatorial Guinea (1.7M), Mozambique (33.9M), Portugal (10.2M), Sao Tome & Principe (0.2M) and Timor-Leste (1.4M). Their populations in 2023 were informed on WorloMeter (<https://www.worldometers.info/world-population/population-by-country/>).

It can easily be observed that Brazil holds 71.3% of the Portuguese speaking population.

This work focus on the ETD collection of a university in Brazil. The university has had graduate programs since the early 1960s, when thesis and dissertations were published in Portuguese, unless a special request was submitted. ETDs had been optional since 2000 and became mandatory in 2002. In 2008, ETDs presented in foreign languages started being accepted without special authorization. The current numbers of ETDs by language are:

Portuguese (pt) – 13,771; English (en) – 1,252; Spanish (es) – 30; French (fr) – 2; Italian (it) – 1

It is easily seen that only English and Portuguese have significant numbers.

### 02. Objectives

The objective of this work is to examine and conclude if the patterns of accesses (numbers of accesses and the numbers and languages of the countries where they originated) are different for ETDs in English and in Portuguese.

### 03. Methods

This work is based on the anlysis of data available on the Institutional Repository and organized in an Open Access dataset available on the same repository. Data on the dataset are numbers of ETDs and of public access ETDs, average numbers of partitions (for some time ETDs had to be presented in a file per chapter), numbers of accesses and the countries where they came from.

The dataset shows that the percentages of ETDs in each language are very different. The percentages of ETDs in English are contained in the interval from little less that 5.0 (2019) to a little over 8.0 (2023). The percentages of ETDs in English are growing every year, though. Examination of the statistics available on the repository shows that ETDs in English are predominately in Science & Technology and Economics.

For this reason, part of the analysis was based on averages to make up for the differences in numbers. Normalization was also used due to the fact that ETDs are partitioned and the numbers of partitions vary according to the language and to the year. Analysis is also performed on accesses from groups of countries that have either Portuguese or English as one of the official languages. The work contains tables with the numbers. Graphics are presented too.

### 04. Results

Most results are not surprising. The first is that accesses from Brazil account for almost 75% to ETDs in Portuguese. The second is that accesses from the US are over 39% to ETDs in English and accesses from countries that have neither language as official are over 35% and over 9%, respectively, for English and Portuguese. The numbers of countries are between 150 and 172 for English and between 194 and 215 for Portuguese; it is important to remark that the numbers of ETDs in Portuguese are over 10 times as many the corresponding (per year) in English.

A surprising result is the number of accesses from Portuguese speaking countries which were less than 7% for ETDs in Portuguese. A not so surprising result is that in the five years of observation the average numbers of accesses are different for both languages but ETDs in Portuguese had two to three times the average numbers of accesses when compared to the works in English.

## 05. Conclusions

There are significant differences in the patterns of accesses in the two sets. The percentages of accesses from Brazil are significant even to works in English. This is easy to explain due to the fact that the university is in Brazil. The other difference comes from the fact that ETDs in English are mostly in Science & Technology and Economics, and there are many repositories in these areas offering works in English.