

Exploring the Dominance of the English Language on the Websites of EU Countries

Created by:  Minas Pergantis,  Andreas Giannakouloupoulos
Version received: 12 May 2020



The English language is the most dominant language in the western world and its influence can be noticed in every aspect of human communication. This article presents data about the prevalence of the English language in the different countries and regions of the European Union. By analyzing and processing the findings from over 100000 websites from every country in the EU, a solid foundation is set that is used to explore the dominance of the English language in the European World Wide Web in general. The results of this process demonstrate that the English language is not only available on more than one quarter of all websites of non-English speaking EU countries, but is also available in the vast majority of multilingual and bilingual websites of the EU, while at the same time being the only available language in a number of monolingual websites and shown preference over the national language in a significant number of cases.

This topic review is primarily based on the article "Exploring the Dominance of the English Language on the Websites of EU Countries" ^[1] which has been published in the "Future Internet" Open Access Journal.

The English Language in the EU World Wide Web

In an effort to increase international communication, people and organizations within the EU area and Europe in general have been adopting the use of English as a common language for a long time. Although the use of the English language is considered an important factor for the internationalization of institutional and organizational websites ^{[2][3]}, there is still, according to our knowledge, no study that focuses on the EU countries and spans multiple disciplines from education to government and commerce. The facts that the EU member states are archetypal nation-states and language is the common denominator of national identity strengthen the need for a study such as the one at hand, which will try to provide data regarding the influence of a foreign language in member states of a multinational formation such as the EU. The purpose of this research is to study the diffusion of the English language, and measure its spread in countries of the EU, and more specifically in the ones that don't have English as one of their official languages.

In summary, the research goals of this work are: (i) to statistically examine and measure the existence of English content availability in a wide range of websites from each EU member state and (ii) to examine the proportion of websites in which English appears to have greater prevalence than each EU member state's official language or languages. In order to meet these goals, the study provides accurate and fully quantifiable data extracted by a very large and representative sample of EU websites. This information can become a valuable instrument in understanding not only the use of language to achieve better international reach, but also the changes underway in the national identity of EU member states.

This research makes use of the Internet and specifically the World Wide Web to develop a method to analyze the diffusion of English. The World Wide Web is a major driving force in creating a global community but is also a technological entity that can to some extent be measured. By studying what languages are available or hold prominence in a large number of websites from every country of the EU we can get an accurate metric of the diffusion of English.

For the purposes of this study we consider the use of a National Top-Level Domain (NTLD) as an intentional action from the website's owner to associate their website with a specific country. There are no technical requirements in procuring a Global Top-Level Domain (GTLD). The website hosting server does not need to be in a specific location and the language of the content presented in the website does not need to meet specific criteria. Hence, it is safe to assume that anyone who desired not to associate their website with a specific country or region would opt for the

selection of a GTLD. On the other hand, purposefully selecting a NTLD can be seen as a clear indication of a relation between the website and the equivalent country. This notion is further reinforced as in the past use of the country's NTLD has been linked to cultural characteristics and national pride for example in Sweden^[4] or India^[5].

Following that reasoning, from a grand total of more than 5.9 million recorded websites that belong to European national top-level domains, a sampling pool of over 100 thousand websites was created. These websites were then automatically traversed, and their content analyzed in order to infer the language they use, as well as any additional languages they might be offering, through a Language Inference Algorithm (LIA) developed exclusively for the purposes of the research at hand.

Quantifying the Language Landscape of the EU Web

By taking the average for the percentages of monolingual, bilingual and multilingual websites for the main NTLD of each member state as detected by the LIA we can get a unified picture of what's going on in the EU. Figure 1 displays a pie chart of that average for Websites for non-English speaking member states and its equivalent for English speaking member states.

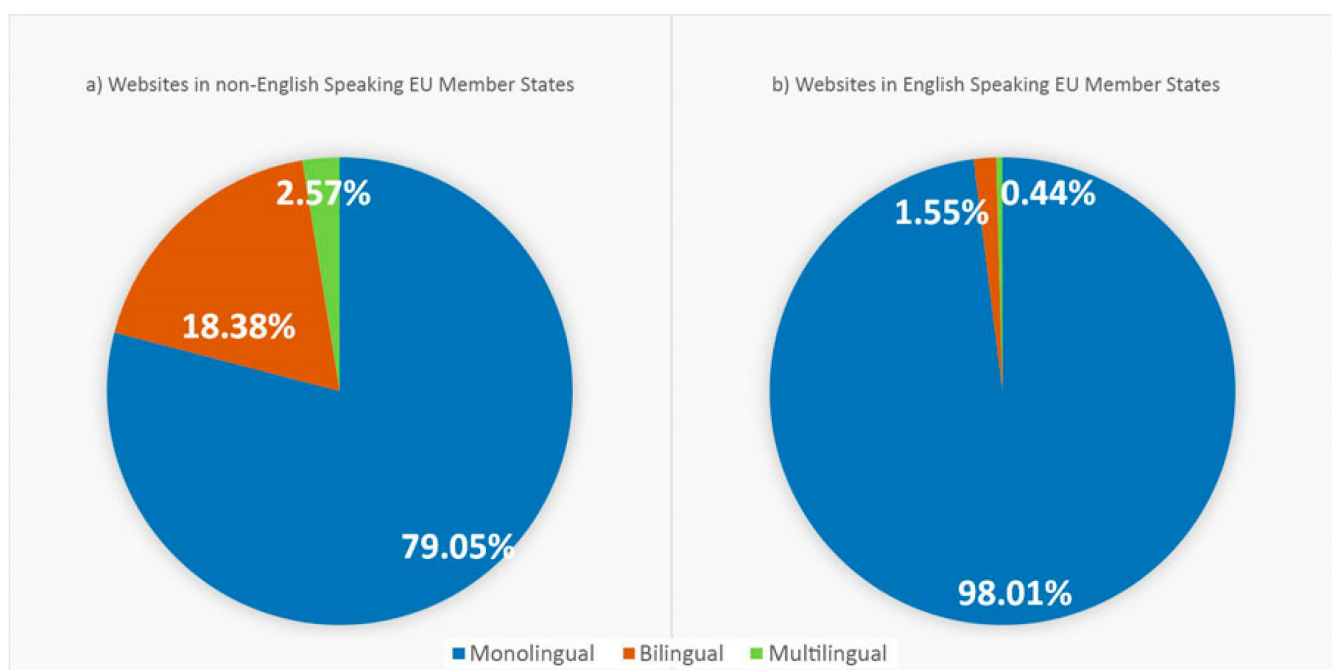


Figure 1

Having established an overview of the number of languages available, the next step is focusing on which languages are preferred. Using the results inferred by the LIA it is calculated that an estimate of 25.64% of websites in Non-English speaking EU countries are available in English as seen in Figure 2. This includes monolingual websites that only offer the English language and bilingual or multilingual websites that offer English among other choices. In English speaking countries this percentage reaches 99.1%.

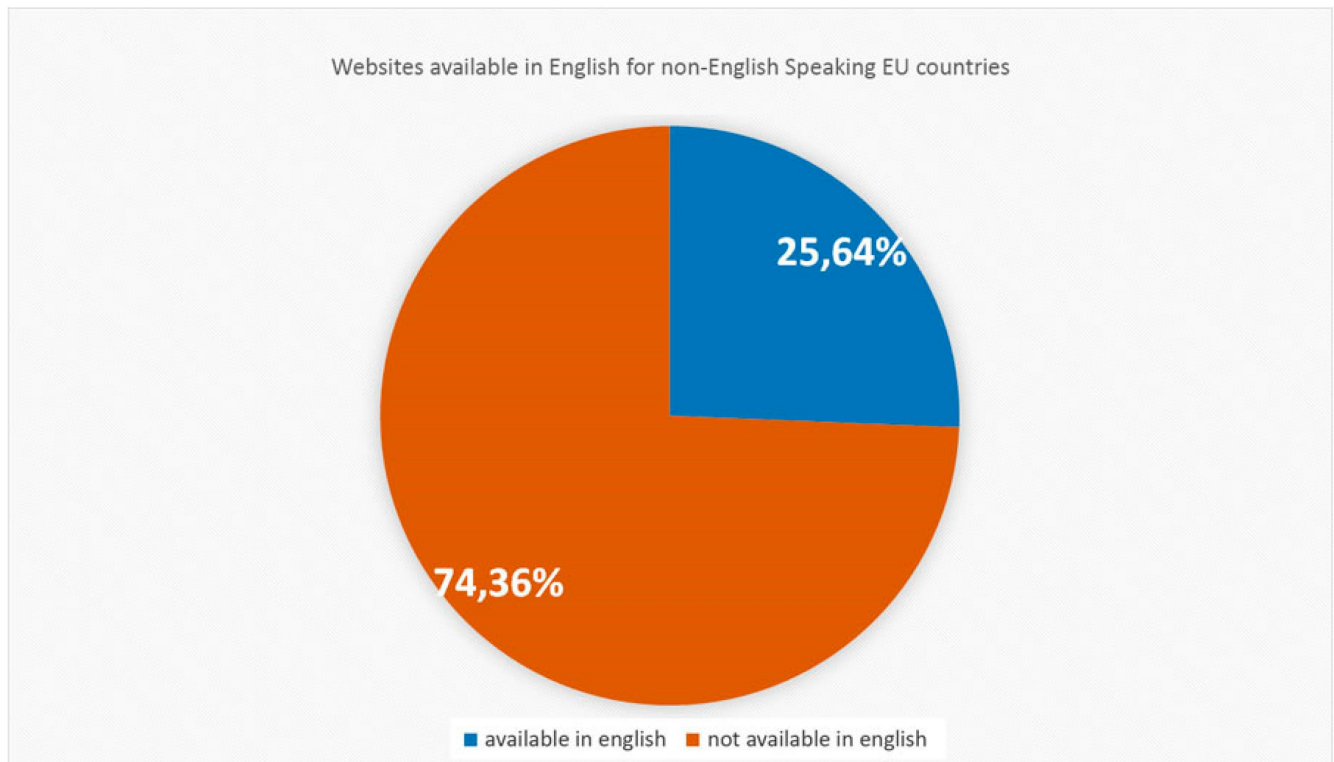


Figure 2

Having the English language available on a website can be achieved by developing a website only in English, by adding English on top of each country's established official language as an option or by providing it alongside a group of other languages. In Figure 3 we can compare the availability of the English Language in monolingual, bilingual or multilingual sites.

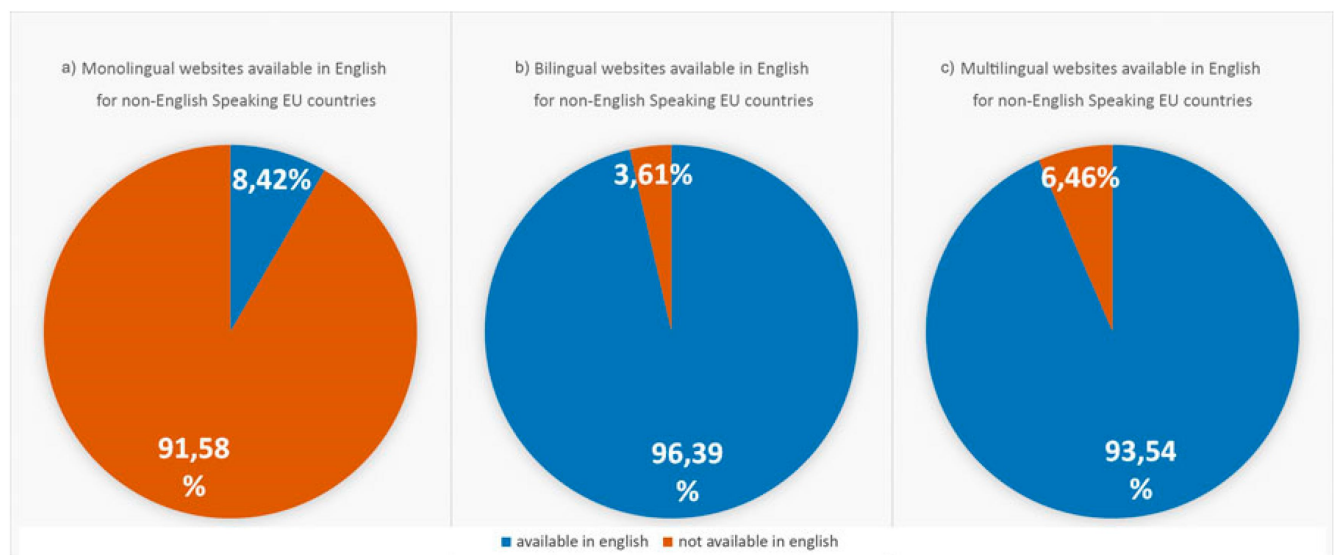


Figure 3

While looking at averages can provide a good summary of the language landscape in the European part of the World Wide Web, it is also important to notice the trends in each individual country. In Figure 4 the availability of the English language can be seen for non-English speaking countries in a comprehensive map.

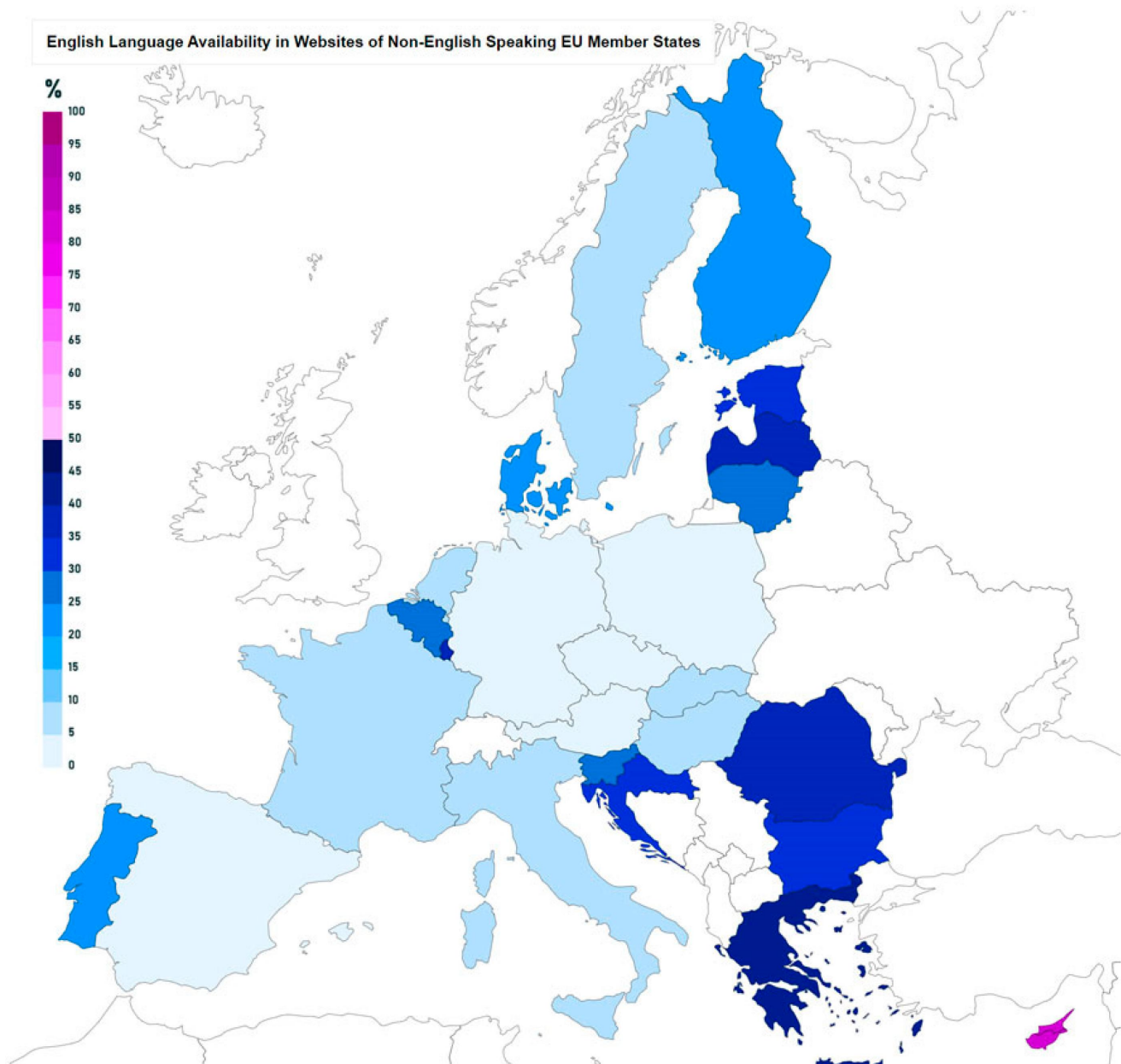


Figure 4

The first language a user encounters on a website's landing page can be safely considered the primary language of that website. As seen in Figure 5 the English language appears as the primary language of 9.92% of websites on average in non-English speaking EU member states.

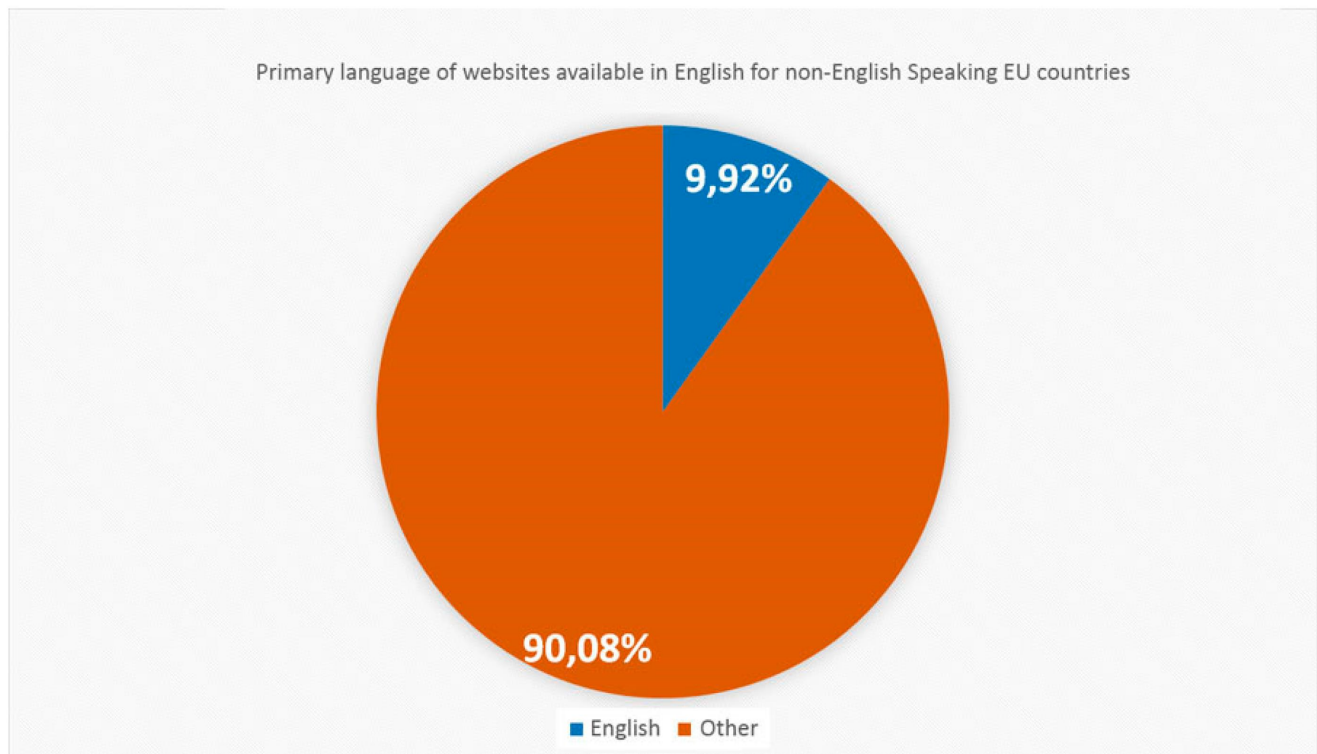


Figure 5

Moving away from the average and trying to establish an overview of the situation in each individual country, Figure 6 presents a comprehensive map depicting the websites that have English as their primary language in the non-English speaking countries of the EU.

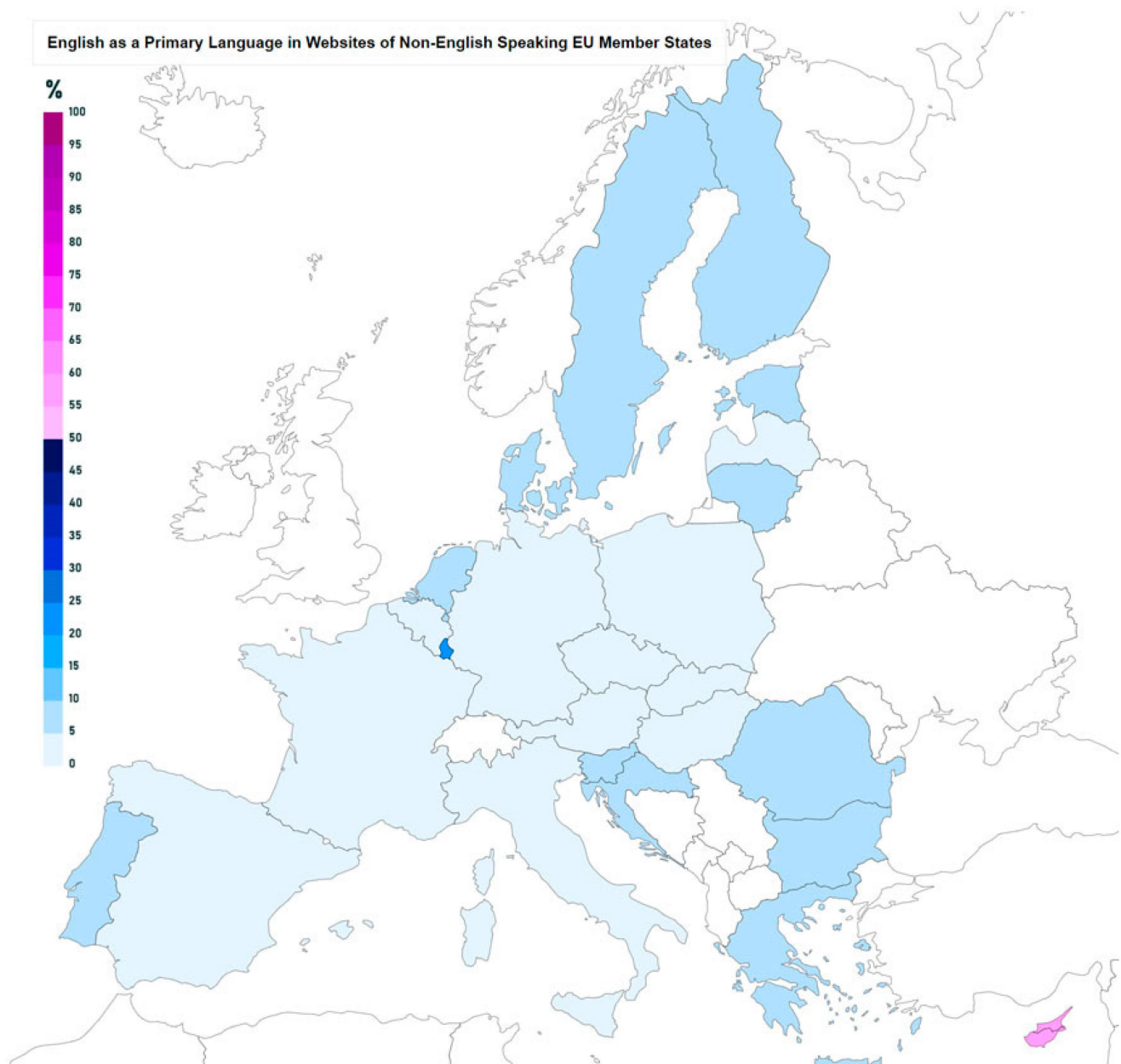


Figure 6

A Deeper Look behind the Numbers

As indicated by the results, more than one quarter of websites belonging to NTLDs of non-English speaking countries, on average, offer their content in the English language. On the other hand, websites of English speaking countries very rarely offer the option for any other language besides English. This clearly indicates that for both English speaking countries and non-English speaking ones the use of the English language is viewed as the most efficient way to attain international reach. On top of that, a significant number (almost 10%) of websites from non-English speaking countries prioritize English over their own official language or languages. This further reinforces the status of English as the most prevalent language for international communication in the wider European Union.

When studying the reasons that lead to greater or lesser availability of the English language in websites of non-English speaking EU member states, a statistically significant moderate negative correlation was discovered between the population and GDP of a member state and the availability of English in its websites. This signifies that in general, larger countries population-wise and more affluent countries put less effort in providing their website content in English than smaller countries do. This correlation reinforces the notion that language is a hard barrier in achieving greater reach. Countries with a smaller number of people speaking their official language need to compensate by putting more effort into providing their content in English, in order to attain the reach that an equivalent website in a larger non-English speaking country or in an English speaking country would have.

Besides population and GDP, some characteristics that were identified to influence the prevalence of the English language were: i) a country's position in the EU periphery (in the Balkans, the Baltic states, Scandinavia and Portugal) ii) an already existing multilingual culture due to more than one official language (in Belgium and Luxemburg) and iii) a close historical connection with British culture (in the case of Cyprus).

The main contribution of this article lies in the provision of quantified information regarding the usage of the English language in the websites of EU, obtained from a large-scale data research all over the web of EU member states. In addition, this research paves the way for the exploration of the current status of European integration in quite a different way than the traditional approaches by the use of unintentionally and freely provided data on the Web.

References

1. Andreas Giannakouloupoulos; Minas Pergantis; Nikos Konstantinou; Aristeidis Lamprogeorgos; Laida Limniati; Iraklis Varlamis; Exploring the Dominance of the English Language on the Websites of EU Countries. *Future Internet* **2020**, *12*, 76, 10.3390/fi12040076.
2. Costales, A.F. The internationalization of institutional websites: The case of universities in the European Union. *Transl. Res. Proj.* **2012**, *4*, 51–60.
3. Nunes-da-Cunha, I.; Martinez, F.M.; Fernandez-Llimos, F. A Global Comparison of Internationalization Support Characteristics Available on College of Pharmacy Websites. *Am. J. Pharm. Educ.* **2019**, *83*, 6592.
4. Schlesinger Wass, E. Addressing the World: National Identity and Internet Country Code Domains; Rowman & Littlefield Publishers: Lanham, MD, USA, 2003; p. 16.
5. Sharon Strover; A Review of: "Addressing the World: National Identity and Internet Country Code Domains". *The Information Society* **2006**, *22*, 59-60, 10.1080/01972240500388313.

Keywords

web presence; geographical domains; Internet statistics; European Union; English language; multilingual websites; globalization; World Wide Web



© 2020 by the author(s). Distribute under a Creative Commons CC BY license