
Quantifying Americanization: Coverage of American topics in different Wikipedias

This is a preprint version of the paper:
Konieczny, K. & Lewoniewski, W. (2024)
*Quantifying Americanization: Coverage of
American Topics in Different Wikipedias.*
Social Science Computer Review.
DOI: [10.1177/08944393231220165](https://doi.org/10.1177/08944393231220165)

SAGE

Abstract

We quantify the concept of Americanization on a global scale through comparative analysis of the coverage of American topics in different language versions of Wikipedia. We analyzed over 90 million Wikidata items and 40 million Wikipedia articles in 58 languages. We discuss whether Americanization is more or less dominant in different languages, regions, and cultures. We show that the interest in American topics is not universal. Western, developed countries are more Americanized (more interested in topics related to America) than the rest of the world. This is the first global, quantitative confirmation of issues often hypothesized, or assumed, in the literature on Americanization and related phenomena. This study shows that Wikipedia and Wikidata can allow quantification of social science concepts that previously were considered not realistically measurable. Finally, the presented research is also relevant to the discourses on the biases of Wikipedia.

Keywords

Wikipedia; Americanization; Wikidata; Big Data; Data Mining

Introduction

It is commonly accepted that American culture has a significant impact on the globalizing world, a phenomenon usually referred to as Americanization. Numerous works have been published on this topic, usually taking a moral stance on the pros and cons of Americanization, or describing its worldwide history. Such works have generally approached this topic from a qualitative and theoretical perspective, as quantifying and objectively comparing cultures is rather challenging. Some authors have argued that Americanization cannot be quantified and measured, while others attempted to measure singular dimensions, often economic (such as box office sales for American movies, or possession of jeans). Despite the growing literature on the subject, no attempt to present

a quantitative overview of Americanization, through an index or similar metric, or even a simple ranking of countries has been attempted. As we will demonstrate, a grander approach, however, is becoming more realistic in the era of Big Data.

In this paper, we attempt to quantify the prevalence of Americanization as visible in the coverage of American topics in different language versions of Wikipedia. There are over a hundred different language editions of Wikipedia, each independently written from the others by local volunteers. Comparing the proportional amount of content about the United States present in various editions of Wikipedia, and its popularity (page views) can allow us to quantify the concept of Americanization on a global scale, and answer such questions of whether Americanization is more or less dominant in different regions and cultures. Such a comparative quantification of American content in various Wikipedias should contribute to the discourse on the ramifications of American culture in globalization (King & O'Boyle, 2002; Barjot, 2003; Beck, Sznaider, & Winter, 2003), what Schröter (2005, p. 220) described as the still under-researched topic of "differences in economic and social culture" that can act as facilitators or barriers to Americanization. However, due to limitations of space, this study can only briefly outline some related issues and topics, as any more in-depth discussion of the Americanization of almost 60 language groups (roughly corresponding to that many countries), the topic tackled here, would be challenging for a book, and certainly impossible for a paper-sized treatment. As such, while we attempt to direct readers to relevant literature on topics such as, for example, the Americanization of Italy, we ask the readers to remember that the main focus of this paper is to demonstrate that quantification of the global impact of American culture, i.e. Americanization, is possible, not an attempt to explain the observed differences, which instead is a fertile ground for further research.

This study is also relevant to the discourses on the biases of Wikipedia (Jemielniak, 2014, p. 77). While a number of authors have criticized the English Wikipedia for being focused primarily on English-speaking (Western) topics, little if any studies have been done to measure such bias in non-English Wikipedias. In this paper, we focus on Americanization (coverage of US topics) (Nehring, 2004; Berghahn, 2010), although a similar analysis for Westernization (coverage of Western topics) would also be feasible.

The choice of Wikipedia is motivated by theoretical relevance and practical implications. The importance of Wikipedia comes from the simple fact that Wikipedia is widely popular and treated by many as a reliable information source (Keegan, 2013; Avieson, 2019). For more than a decade, it has been consistently ranked as one of the 10 most popular websites according to Similarweb and former Alexa (Okoli, Mehdi, Mesgari, Nielsen, & Lanamäki, 2014). It has been called the 'most influential source of information in the world' and 'our first destination when we want to understand something' (Kleeman, 2015). Bilić (2015) noted that Wikipedia represents one of the modern centers "for knowledge production, dissemination, and consumption in the network society". As such, it stands to reason that by measuring the coverage of American topics in different language editions of Wikipedia, we can measure the global strength of this phenomenon, a Big Data extension of the approach used in studies of Americanization and similar concepts, previously limited to conceptually smaller studies about particular brands, items or concepts (Crothers, 2021, p. 184).

As for the practical implications, the database structure of Wikipedia and its sister site Wikidata makes it feasible to employ Big Data techniques to analyze it as both a convenient and reliable database of much of human knowledge (see related studies such as Konieczny and Klein (2018), Miquel-Ribe and Laniardo (2021), Laouenan et al. (2022)).

Americanization in the Digital Age

Americanization is an ambiguous concept with a number of definitions (Barjot, 2003; Schröter, 2005; van Elteren, 2006; Dębska, 2010). In its international aspect, on which this paper is focused, it can be understood as the influence of American culture on other countries outside the United States of America, visible through various dimensions, from pop culture and cuisine to business practices and political techniques, following van Elteren's (2006, p. 103) definition ("a process in which economic, technological, political, social, cultural and/or sociopsychological influences emanating from America or Americans impinge on values, norms, belief systems, mentalities, habits, rules, technologies, practices, institutions and behaviors of non-Americans"). It was perhaps first coined in 1902 by the British journalist William Stead, who used this term in the title of his book, *The Americanization of the World*, in which he discussed the growing popularity of the "American ideas". It was not originally a pejorative term but these days is often used by critics of America, who are against the spread of its influences (a process sometimes described as cultural imperialism). As such, the concept of Americanization has been also described as a term related to anti-American sentiments, sometimes also known as McDonaldisation (Fabbrini, 2004; Schröter, 2005; van Elteren, 2006, p. 106-107; Dębska, 2010; Berghahn, 2010). It is also related to wider concepts of Westernization (Nehring, 2004) and of course globalization (Dębska, 2010) and in a wider theoretical perspective, it is an example of cultural assimilation (Ramsey, 2015).

According to some, Americanization, understood in this way, is losing its drive, as the "American century" has ended, or will soon, impeded among others by the resistance to the American economic model (Schröter, 2005, p. 220), and the growing power of India and even more so, China (van Elteren, 2006, p. 196, 210). Yet according to others, it has become more prevalent since the end of the Soviet Union in 1991 and especially since the advent of widespread high-speed Internet use in the mid-2000s. Indeed, a new dimension of anti-Americanism is the fear of the pervasiveness of American Internet technology, and the near-monopoly of new media companies such as Google or Facebook (Berghahn, 2010; Kroes et al., 2003, p. 235-256). And there is something to be said here about Wikipedia being ostensibly an American (non-profit) enterprise (operated by the Wikimedia Foundation, an entity registered in the USA), representing American-supported values such as free speech which led to it being banned or censored in a number of places, such as China (Clark, Faris, & Heacock Jones, 2017).

Some authors have argued that Americanization cannot be quantified and measured (ex. Mueller, 2004; Vučetić, 2018, p. 10). Others have attempted to identify and quantify the increase in other cultures' exposure to the American way of life as evidenced by travel, marketing, and appropriation of elements of American culture; much of

such work has been focused on the economic aspects of globalization, as trade flows are usually better documented and less subjective (Barjot, 2003; Schröter, 2005). Economics, however, is only one of the three major dimensions of globalization, the other being political and cultural (Crothers, 2021, p. 18), although arguably some metrics such as movie ticket sales attempt to encompass more than one dimension. Political Americanization measures can be found, but generally in the context of Americanization of immigrants to the United States (see for example the metric employed in Rodolfo & Yang, 2019, p. 69). Cultural Americanization measures found in the literature include surveys of the population's awareness of or attitude to certain topics, such as American classic brands (Coca-Cola, McDonald's, NFL, Facebook), or their possession of certain items associated with America, such as jeans (Crothers, 2021, p. 184). Such piecemeal studies are valuable, but certainly hard to reliably generalize from; as observed by van Elteren (2006, p. 122), "consumption of American imports is not automatically proof of an attendant psychological change", and further, narrow definitions can hinder a deeper understanding of the studied phenomena (van Elteren, 2006, p. 103).

Due to the limitations of the quantification approach, no universally agreed-upon measure exists which would allow us to confidently say which countries, or regions, are more Americanized. Much of the existing research is theoretical and qualitative in nature. Schröter's study, limited to Europe, did not provide any metrics that would allow comparing or ranking of countries, despite being focused on the economic dimension. Similarly, Stephan et al. (2005), while discussing the Americanization of Europe, likewise provide no such metrics. Similar issues are found in research on the Americanization of Latin America (ex. Tota, 2010; Rodolfo & Yang, 2019). Bajrot (2003) suggested that Americanization is most common in Europe, Latin America, and the Far East. van Elteren (2006, p. 133) suggested this parallels, from the state-central perspective, the historical projection of America's power and foreign outreach, which "began modestly" in the mid-nineteenth century in these regions, although "by the end of the twentieth century virtually no place of the globe was left untouched by American influence in one way or the other".

In the mid-2000s Schröter (2005, p. 7, 218) noted that "scholarly literature on Americanization is not yet extensive" (and in the geographical dimension, there were next to no studies of its impact outside Europe, Latin America, and Japan). Likewise, Pasquino (2005) noted that while the issue of Americanization of Italy has been of interest to some scholars, it has yet to be systematically studied. Stephan et al.'s (2005) discussion of Americanization of Europe is in practice only that of Americanization of the UK, France, Germany, Sweden, Denmark, Austria, Poland, Russia, Italy Greece, and Spain, as the work contains chapters only about these particular countries, and is emblematic of the lack of comprehensive understanding of the subject. The varying dimensions of Americanization, as discussed by Volker (2010), arguably require a book-length treatment. Increasingly, they receive it, ex. Beck et al., 2003, van Elteren, 2006, Conrad, 2014, Rodolfo & Yang, 2019, Crothers, 2021, but we are still in a relatively early period of understanding this phenomenon. In particular, none of such treatments have attempted to present a quantitative analysis of Americanization around the world.

We will now turn to the discussion of the second issue to which this study is relevant, which is our growing understanding of the biases of Wikipedia.

Systemic bias on Wikipedia

Most of the existing research on bias in Wikipedia focuses on unequal contributions, in particular, gender bias, followed by age bias, related to the findings that the majority of Wikipedia contributors are young males (Donlan, 2010; Shaw & Hargittai, 2018; Worku, Bipat, McDonald, & Zachry, 2020; Young, Wigdor, & Kane, 2020). Additional findings also suggest that Wikipedia is less likely to be edited by individuals who are less educated (Shaw & Hargittai, 2018).

With regards to the global digital divide and the resulting systemic bias, fewer studies have tackled this issue, although there is an agreement that editing Wikipedia is an activity more commonly associated with individuals living in developed countries, as well as a consensus in the literature that the resulting bias is a significant problem (Apic, Betts, & Russell, 2011; Rogers, Sendjarevic, et al., 2012; Greenstein & Zhu, 2012; Jemielniak, 2014, p. 77; Martin, 2018; Laouenan et al., 2022).

Overall, existing literature suggests that systemic bias on Wikipedia may mirror that of global culture, for example through focusing on topics of interest to the Western audience or male readers. As such, it has been suggested that Wikipedia is likely biased toward Internet pop culture, with an overemphasis on topics such as pop culture, technology, and current events (Konieczny & Klein, 2018).

The reasons for bias vary. Miquel-Ribé and Laniado (2018) distinguish discourse and structural reasons. Discourse effects are based on the idea that since each language edition of Wikipedia constitutes a community, its editors tend to develop and hold a shared cultural background that influences the contents of said Wikipedia. This is also why Young et al. (2020) suggested that most of the non-neutral bias originates from less active, occasional editors and tends to be corrected in time by the more active, core editors. Structural effects are based on the idea that context and culture are relevant factors that affect editor interests and consequently content coverage. The commonly identified reason for this is that most Wikipedia contributors are young, male, English-speaking, educated, technologically aware, and wealthy enough to spare time for editing.

This model can be further refined, and as such, we propose that within the structural effects, we can distinguish socioeconomic/demographic and cultural effects.

The former is much better understood than the latter (Rask, 2008; Graham, Straumann, & Hogan, 2015; Miquel-Ribé & Laniado, 2018; Konieczny, 2020) and is related to the structure and stratification of society. It can be understood from the perspective of the age, gender, and education of Wikipedia editors. In other words, the socioeconomic and demographic factors are the same factors responsible for the global digital divide - Wikipedia is mostly written by people who can afford to both access the Internet and develop digital hobbies, something that is still not a norm in the developing world (Rask, 2008; Graham et al., 2015; Konieczny, 2020). And articles are more complete and of higher quality in the language editions associated with their relevant cultures (Callahan & Herring, 2011). It is also such factors, with the stress of demography (Wikipedia is

written mostly by young males) that are responsible for commonly observed project-wide biases such as the dominance of topics of interest to young, Western culture-influenced males (Konieczny & Klein, 2018). In addition to researchers, the Wikimedia community and the Wikimedia Foundation concur that this is a significant issue that both have been trying to address for years; first with regards to gender (see ex. Konieczny & Klein, 2018), and more recently, culture as well. Asia and Africa are in particular underrepresented, with Asia accounting for 27% of the Wikipedia editors (and 60% of the world's population), and Africa, respectively 1.6% and 16% (Meta-Wiki, 2021).

Explanations for the cultural factors are much sparser, although they are uneniably important; Jemielniak & Wilamowski (2017) also found that the very "nature of understanding what makes a good description of a given phenomenon is culture-dependent". On the surface, they also appear pretty simple: people will write about topics related to their culture (Callahan & Herring, 2011). Kolbitsch and Maurer (2006) argued that local Wikipedias emphasize "local heroes" and thus "distort reality and create an imbalance", although Callahan and Herring (2011) who observed differences between Polish and English Wikipedia, concurred that local Wikipedias are more likely to "promote their local heroes and local values", but saw in it a potential for representing multi-cultural diversity, particularly once automated translation tools allow interested readers to compare multiple versions of the article, a sentiment also echoed by Miquel-Ribé (2019). They also noted that "there is no evidence that the resulting biases are intentional attempts to deceive or distort, as the word bias may connote", and that differences were mostly an unintentional result of differing interests and experiences of editors, and the simple fact that there are many more editors interested in English topics than Polish ones. More recently, a big data analysis by Miquel-Ribé and Laniado (2018) of about 40 different languages Wikipedias found that about a quarter of each Wikipedia language edition is dedicated to representing the corresponding cultural context, although this number varies from 7 to 49% depending on the particular Wikipedia analyzed. They also noted that such content is less likely to be included in other Wikipedias, compared to more general or international topics. At the same time, follow-up research (Miquel, Laniado, & Kaltenbrunner, 2021) indicated that it is this very content that is among the most highly demanded, regionally (in other words, readers are highly interested in their local topics and much less in foreign local topics). Likewise, Wiggers (2018) observed that Wikipedia favors mainstream languages and content, and the visibility and impact of content from more niche languages and cultures is very limited. Already in 2009 Hecht and Gerge described the cultural clustering of topics and difficulty of diffusion of such articles to other cultures as "self-focus bias".

Miquel-Ribé and Laniado (Miquel-Ribé & Laniado, 2018) applied Hecht's cultural contextualization theory and concluded that cultural and geographical context influences communities' common interests and in turn, gaps and unbalance in coverage of individual Wikipedia projects; however, they did not attempt to answer whether there are any patterns in the observed data - in other words, answer the question of what are the differences and effects in cultural coverage. The size and quality of different Wikipedias cannot be explained just by the socioeconomic factors such as a country's size or development level (Konieczny, 2020), who found that contributing to Wikipedia is more

common for countries that are located near the self-expression and rational-secular ends of the Inglehart-Welzel cultural clusters model, and the uncertainty avoidance, masculinity, and long-term orientation dimensions of the Hofstede cultural dimensions model. Further, there is the question of what affects the widely uneven proportion of articles (7 to 49%) about local cultural topics, as observed by Miquel-Ribé and Laniado (2018).

The discussion of the global digital divide and resulting systemic bias on Wikipedia has been, however, outside the exploratory study by Miquel-Ribé and Laniado (2018), mostly anecdotal and untraced in a longitudinal fashion. This study will provide an analysis of the systemic bias in the context of Americanization (focus on American topics) visible in the coverage of several Wikipedias, in an attempt to quantify how significant this bias really is, whether it shows any changes over time, and whether it affects the “international” English Wikipedia more - or less - than the local ones. In this way, it attempts to provide some of the answers requested by Graham et al. (2015) when they asked for “further quantitative, qualitative, and time-series data and research. . . to better understand key mechanisms and practices that either reinforce or circumvent digital divisions of labor and informational magnetisms” in line with initiatives such as the recently launched Wikipedia Diversity Observatory (Miquel-Ribé & Laniado, 2021).

Data and methods

Our research question is as follows: “Is the coverage of American topics the same in all Wikipedias (of different languages)? If not, what can explain the differences?”

American topics are defined as topics associated with the United States, such as geographical locations, people with American citizenship, entities created by American citizens and entities (works of art, media, businesses, products) related to US.

There are different ways of locating Wikipedia articles related to a given topic. One is based on information from Wikipedia categories which group together pages on similar subjects. However, each language version can define its own structure of categories and connections between them (Lewoniewski, Węcel, & Abramowicz, 2019). In some Wikipedia languages such structure is often too fine-grained to be directly analyzed (Boldi & Monti, 2016). Moreover, there are over 10 million categories in various language versions, which can be used to describe Wikipedia articles at different levels of abstraction, and on average, on English Wikipedia, each article belongs to 22 categories (Lewoniewski et al., 2019).

Another way to identify the topic of the Wikipedia article is to use information from related items from other knowledge bases, such as Wikidata or DBpedia. Those semantic databases have shown their effectiveness in recognizing the topics of articles at different levels of abstraction (Lewoniewski et al., 2019; Lewoniewski, 2022; Lewoniewski, Węcel, & Abramowicz, 2022).

Wikidata is an open semantic knowledge base operating as a wiki service. It provides structured data to Wikimedia projects such as Wikipedia, ensuring that information is standardized, up-to-date, and consistent across different language versions. Unlike

traditional encyclopedic content, Wikidata's structured format is machine-readable, facilitating its use in data analytics, artificial intelligence, and other computational tasks.

Because Wikidata provides a vast amount of linked data, it plays an essential role in the Semantic Web. Developers and organizations utilize it to link their datasets with a common, standardized dataset. Several automated tools and bots that operate on Wikipedia and other platforms use Wikidata to fetch and validate information, helping maintain the consistency and accuracy of the data (Pfundner et al., 2015; Laouenan et al., 2022). Some digital assistants and search engines (such as Google Search) utilize Wikidata's structured information to enhance their results and provide direct answers to users' queries (Mora-Cantalops, Sánchez-Alonso, & García-Barriocanal, 2019; Ait-Mlouk & Jiang, 2020). Additionally, Wikidata helps to improve different library repositories (Tharani, 2021). There are also approaches related to computational biology and COVID-19 combating (Waagmeester et al., 2021; Turki et al., 2022).

Each Wikidata item has a collection of different statements structured in the form: "Subject-Predicate-Object". Figure 1 shows Wikidata item Q83873577 ("COVID-19 pandemic in the United States") with some statements.

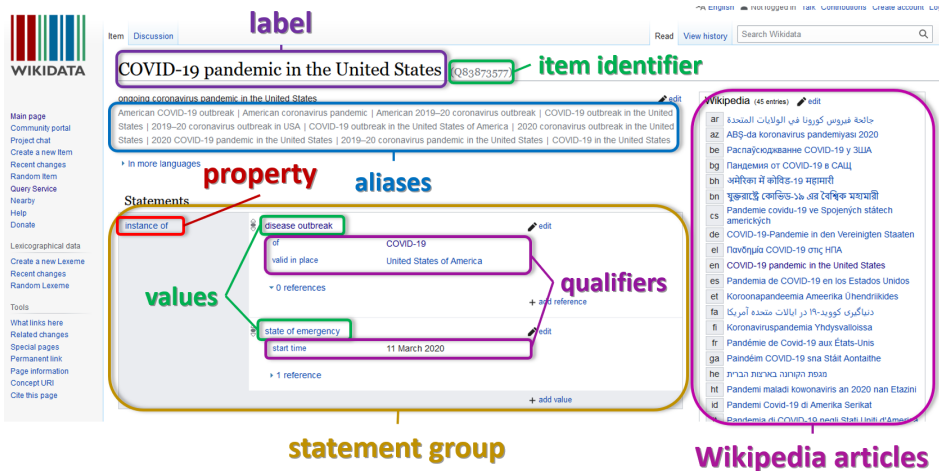


Figure 1. Scheme of the Wikidata item related to COVID-19 pandemic in the United States (Q83873577).

Some of the statements of the Wikidata can help to identify the relationship of an object with a specific country, culture, or language. For example, such a set of Wikidata properties might include: P17 (country), P19 (place of birth), P27 (country of citizenship), P276 (location), P495 (country of origin), and others (Miquel-Ribé, 2019).

Combining the Miquel-Ribé & Laniado (2018; 2019) model based on Hecht's cultural contextualization theory and Konieczny's (2020) use of the Inglehart-Welzel cultural clusters model, we have analyzed over 90 millions of Wikidata items and using its

semantic connections to Wikipedia investigated which of over 40 million articles in 58 considered language versions are related to American topics.

To be able to obtain more complete data about views of the Wikipedia articles on different subjects we analyzed alternative titles (redirects) of the Wikipedia articles. Figure 2 shows how the title of the article on COVID-19 pandemic in English Wikipedia was changed at the beginning of the pandemic and when some alternative titles (redirects) were created. The supplementary materials are at ([Supplementary materials, 2022](#))

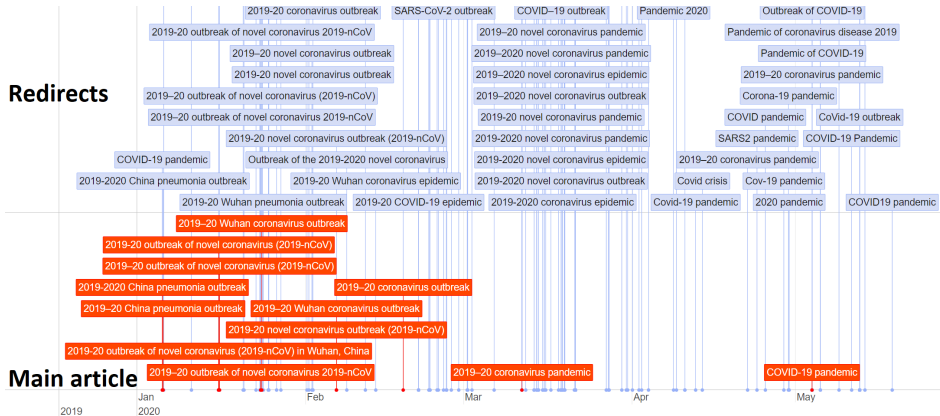


Figure 2. Article names and redirects on COVID-19 in English Wikipedia in the period from January 2020 till May 2020. The extended interactive version of the figure is available on: <https://data.lewoniewski.info/americanization/timeline>

The results are presented for 58 Wikipedias. 53 can be tied to one primary language and a corresponding country (ex. Polish Wikipedia represents the Polish language and the country of Poland). 4 represent language aggregates of more than one country (English, Spanish, Portuguese, and Arabic). We also present results for groups corresponding to Inglehart-Welzel cultural clusters (based on [Haerpfer et al., 2020](#), the partially overlapping clusters are English-speaking, Latin America, Catholic Europe, Protestant Europe, Islamic, South & West Asian, Orthodox, and Ex-communist).

In addition, the content was also coded as to whether it concerns topics that are primarily related to developed or developing countries. That classification was based on World Bank's ([World Bank Data Team, 2019](#)) 4-trier division, with high and upper-middle income countries classified as developed and low and lower-middle income countries, as developing.

To organize our data, and address the limitations discussed in the subsequent chapter, we have developed the following metrics: * PPCRW - percentage of people in a primary country reading that language Wikipedia (primary country refers to the largest country associated with a given language, ex. Germany for German language); * VPC - percentage of views from the primary country directed to that Wikipedia; * US RAS

- percentage of United States related articles share on that Wikipedia; * US RAVS - percentage of United States related articles views share on that Wikipedia.

Limitations

This study is limited by several issues. First, to what degree Wikipedia volunteers and readers are representative of the general populace? Second, the limitations of the Wikipedia model based around languages (instead of countries) make it impossible or hard to study certain groups. Lastly, while the Wikidata content is generally considered reliable, it is still affected by certain biases.

As discussed previously, we know that Wikipedia is widely popular (Keegan, 2013; Bilić, 2015; Miquel-Ribé, 2019). It is expected that the digital divide is a factor and that Wikipedia's readers, like its contributors, are more likely to be younger, more educated, and wealthy. The topic of Wikipedia readership is rather under-researched (see Okoli et al., 2014), but we do know more about contributors to Wikipedia, who from the demographic perspective are mostly younger, better-educated males. According to Graham et al. (Graham et al., 2015), one's choice to join and remain in the Wikipedia volunteer community can be impacted by cultural and organizational factors that differ between various Wikipedias, such as the local community's attitude towards new editors, existence or lack of local Wikimedia chapters or groups that support the community, and efficiency of the dispute resolution procedures. Additionally, factors that do not generally impact one's attitude to and participation in the wider society may affect attitudes to and participation in Wikipedia, for example in countries where freedom of speech is an issue. Overall, we believe that Wikipedia is a rather reliable proxy for how Internet users see the world, and in turn, Internet users are increasingly representative of the general populace; nonetheless, readers should be aware that the digital divide still limits the generalization of such conclusions (for more, see Rask, 2008; Okoli et al., 2014; Graham et al., 2015; Konieczny, 2020).

The second issue we run into is the fact that limitations due to countries and languages having, sometimes literally, blurry boundaries. As such, a number of countries and cultures could not be included in our analysis.

Each Wikipedia project is organized around a language. Since Wikipedias are based around languages, not countries, this means that the macro-level unit of analysis, while simplified in discussions as country/culture, is de facto a cultural language cluster, so references to, for example, Poland, are in fact references to Polish language speakers.

This is a major issue as some languages span multiple countries (from the international lingua franca of English to regional languages such as Spanish or Russian), and other countries are multilingual (like Switzerland and India). Some interferences can be drawn from projects such as Catalan Wikipedia or Arabic Wikipedia. Nonetheless, this means that some populations had to be excluded from our study: using our method we cannot reliably measure, for example, the Americanization of India, as there simply is no Indian Wikipedia (there is a Hindi Wikipedia, but also a Marathi Wikipedia, Tamil Wikipedia, and others, and with over 80% of visits from India are directed to English Wikipedia, any attempt to generalize about India would be rather unreliable). Additionally, we need

to remember that many people are also multilingual (Lemmerich, Sáez-Trumper, West, & Zia, 2019 found that 20% of Wikipedia sessions involve users switching from one language version to another).

Attempting to control for multi-language contexts and large diasporas illustrates the scope of the problem. We have identified 53 Wikipedias that correspond to one dominant country. Three major Wikipedias (English, Arabic, and Spanish) can be seen as significantly international, while the Portuguese Wikipedia is heavily dominated by Brazilians. Only 3 Wikipedias (Italian, Japanese, and Polish) meet the condition of “at least 75% of the population uses a single Wikipedia corresponding to the main language of the country and at least 75% of the views of that Wikipedia originate from that country”. Respectively 18 and 37 meet weaker conditions (for half and a quarter of a population and views, rather than three quarters). Individually, 19 Wikipedias meet the 50% population, and 51, meet the 50% of views criteria. On average, the Wikipedias used in our study have values of 40% and 72%, respectively. See the table 1 for details.

Overall this means that for some, relatively homogeneous countries with small diasporas, we can very reliably correlate the use of Wikipedia in that country to a single specific Wikipedia. For example, about 90% of Wikipedia page views from countries like Poland, Japan, or Italy go to their respective Wikipedias, and conversely, a similar proportion of views to these Wikipedia originate from their countries. In other words, this means that the Japanese Wikipedia is primarily read and written by the Japanese people who live in Japan, and the same is true for the Italian and Polish cases, as so the percentage of American topics in Polish, Japanese, and Italian Wikipedias is a rather reliable representation of the popularity of American culture in these countries. However, a country like Belgium shows the distribution of 34% views to Dutch Wikipedia, 32% to French, and 28% to English, with no Belgian Wikipedia in existence, while Dutch Wikipedia receives 69% of its views from the Netherlands and 20% from Belgium. Therefore we cannot measure, reliably, the Americanization of Belgium, and the Americanization of the Netherlands is captured less reliably than that of Poland. The least reliable results can be expected for Ukraine, Belarus, Malaysia, Kirghizstan, China, Nepal, Sri Lanka, Uzbekistan, the Philippines, Cambodia, Bangladesh, Laos, Myanmar, Pakistan, Haiti, Hindu/India, and Somalia, where fewer than 25% of views are directed to their expected, national-language Wikipedias. Any discussion of the four significantly international Wikipedias is inevitably a discussion of multi-country entities. English Wikipedia can be seen as representing primarily the United States, United Kingdom, India, Canada, and Australia (accounting for 70% of the views). 60% of the views of Arabic Wikipedia come from Saudi Arabia, Egypt, Morocco, Algeria, Iraq, and Jordan. 70% of views for the Spanish Wikipedia come from Mexico, Spain, Argentina, Colombia, Peru, and Chile, with Spain accounting only for 20% of the total views. Portuguese Wikipedia is primarily read by Brazilians (80% of views), with views from Portugal accounting only for 7%. In the last two cases, due to the expected influences of the digital divide, the number of editors from Spain and Portugal is likely higher than readership proportions would suggest.

As suggested by Konieczny (2020), development level and country size are likely a factor here, as less developed, smaller countries will have less developed Wikipedias

(due to fewer active volunteers). Compare, for example, German Wikipedia (over 2 million articles, 4,500 active editors, representing about 100 million native speakers) to Arabic Wikipedia (which while representing 350 million native speakers sports only a million articles, supported by less than 1,000 active volunteers), or Bengali Wikipedia (representing over 240,000 million native speakers in Bangladesh and India, with about 125,000 articles and less than 200 active volunteers); the latter is at the level of - for example - Estonian Wikipedia (a bit over a million native speakers, yet with about 250,000 articles and over a 100 active editors). Inevitably, readers will choose to use much larger Wikipedias in languages they are more familiar with (or, these days, can simply machine translate directly from their browser). Indeed, more than half, and sometimes 90% of visits from these countries are directed to larger Wikipedias (usually English, sometimes others, like Russian or French). English Wikipedia accounts for approximately half of all Wikipedia views worldwide.

With regards to groups of countries corresponding to Inglehart-Welzel cultural clusters, our study does not contain an African cluster, as we did not identify any reliable African-language Wikipedia outside the borderline Somali language one (Egyptian Arabic Wikipedia exists and is reasonably large, but seems very limited use within Egypt, with only 2% of views compared to 60% that Arabic Wikipedia receives from that country). The English-speaking cluster is effectively the English-language Wikipedia, Latin America is represented by Spanish and Portuguese language Wikipedias. Catholic Europe is represented by 10 Wikipedias, Protestant Europe by 11, Orthodox by 10, Islamic by 11, South & West Asia by 13, and Ex-communist by 22 projects. Hebrew Wikipedia, representing (Jewish) Israel, was not included in any cluster.

There are certain issues with the Wikidata dataset, mirroring issues with Wikipedia. Some are related to previously discussed biases due to the nature of its editor base as well as readership that in turn may reflect wider issues such as those related to gender bias or the digital divide. Others are the outcome of the grassroots nature of the project. Given the relatively small number of volunteers active within some language editions of the Wikimedia movement, some of the bias may be due to the preferences of a small number of gatekeepers (Li & Farzan, 2020). For example, the Swedish Wikipedia is unproportionally large due to the efforts of several volunteers, who coded automated scripts (bots) that machine-translated or otherwise created numerous articles about plants, animals, and certain geographical entities. That does not mean, however, that Swedish readers are above-average interested in such topics, but it does skew the metrics based on article count. To control this, we have moved our metric beyond the total or proportional article count and focused on page views. This is because regardless of the actions of some Wikipedia content creators, the public interest in the content will be the same. In other words, while the Swedish Wikipedia may have a proportionally smaller number of articles about American topics than the Norwegian Wikipedia, due to having many more articles about plants and animals, this should not affect the interest of the Swedish or Norwegian public interest in American topics. This is why, while American topics represent only 3% of Swedish Wikipedia's content, and 7.5% of Norwegian, both have comparable levels of page views (respectively, 10.5% and 8%). Fortunately, this

limitation is easily addressed by focusing not on the number of articles (either total or American-focused) in a given Wikipedia but on their percentage.

This is also related to the exclusion of certain large Wikipedias from our research. Such language versions as Cebuano, Egyptian Arabic, or Waray formally have over 1 million articles, but have a very low collaborative quality (or editing depth - https://meta.wikimedia.org/wiki/Wikipedia_article_depth) and a very low average number of visits per article. This usually happens when such language versions were massively enriched by bots and without checking the content by humans (this means that any content may appear there and that may not fully reflect the preferences of the editors or readers of that language version).

Finally, we also need to consider the fact that low viewership may represent a lack of content rather than interest. Mulina (2018) noted that "the use of Wikipedia and the perception of its usefulness with non-English speaking users depends, by and large, on the completeness of the Wikipedia content in their native language as well as their knowledge of both second and foreign languages". For example, the two smallest Wikipedias in our study are Laotian (about 4,000 articles) and Somali (about 8,000). It is likely that such Wikipedias may lack numerous articles that would be highly read if they existed. Low Americanization values for such Wikipedias may indicate not the lack of interest in such topics, but the lack of relevant articles to be accessed. Fortunately, 52 out of 58 Wikipedias used in our study are over 40,000 articles long (the size of the 2014 Global version of Encyclopedia Britannica); with an average of 715,000 articles. The six Wikipedias below 40,000 articles in our study are: Khmer, Sinhalese, Somali, Mongolian, Nepali, and Laotian.

The lack of content, however, can be caused not just by the small size of a project, but also by more restrictive inclusion criteria, known within the Wikipedia movement as "notability". While all Wikipedia projects are in broad consensus over the need to include articles on encyclopedic topics, the specific guidelines for what merits inclusion and what does not are determined by local, not global, consensus. For example, some Wikipedia projects are less likely to allow the creation of articles on topics such as current events or topics related to popular culture than others. Unfortunately, no comprehensive comparison of differing notability standards between different languages Wikipedias exists as of now. For more on the concept of notability within the Wikipedia movement, see Taraborelli and Ciampaglia (2010) and Jemielniak (2014, p. 13, 24, 157–158), as well as Jemielniak & Wilamowski (2017) on how similar processes can be affected by cultural factors.

Future research may address the issue of cultural bias by comparing articles from general fields of knowledge, such as mathematics or medicine. This could be done by using lists of articles maintained by, for example, English Wikipedia WikiProjects such as WikiProject Mathematics or WikiProject Vital Articles. This would allow measuring discrepancies that could be attributed to the project's maturity rather than cultural domination.

Therefore the findings of this study are most reliable when applied to the countries which are not multilingual, and which do not have large diasporas, as the logical relation between country and language is most relevant for that group. The total

population of speakers in a given language, and the development levels of their countries, which translate into having larger, more comprehensive Wikipedias, is also a factor. For developing countries, the results are also less likely representative of the general populace, given the still persisting digital divide (although as of 2022, Internet penetration rates in Africa are over 40%, Asia, 65%, and Latin America and the Middle East, approaching 80%). As non-English Wikipedias grow and their coverage improves, this should reduce such potential errors in data, and the reliability of the entire dataset warrants reporting revised results in perhaps another decade or so.

The findings may be least reliable for 11 Wikipedias which fail two or more of the criteria discussed above (have less than 40,000 articles, “at least 50% of the population uses a single Wikipedia corresponding to the main language of the country”, and “at least 50% of the views of that Wikipedia originate from that country”: Khmer (Cambodian), Urdu (Pakistani), Sinhalese (Sri Lanka), Bengali (Bangladesh), Uzbek, Somali, Mongolian, Nepali, Belarusian, Laotian and Haitian. Conversely, however, 47 remaining Wikipedias from our sample (over 80% of the total) meet two or more of these conditions and should not be significantly affected by the biases discussed here. The reliability criteria of each Wikipedia are included in Table 1.

For the list of all Wikipedias used in our study (57 mentioned above plus the Hindu Wikipedia, an imperfect but feasible way to include India in such a study), and their values with regards to the parameter PPCRW, VPC and Wikipedia size, see Table 1.

Findings

We will now use the obtained data to compare Wikipedia language versions. The figure 3 shows bubble chart where the placement of each language depends on US RAS and US RAVS parameters, the size of the symbol indicates the number of articles in the Wikipedia chapter, colors indicate the range of PPCRW values: if 0-49%, such languages were marked as red, otherwise (for 50-100%) blue color was selected. For better visibility in the paper for this chart a limited logarithmic scale has been chosen.

We can observe that the "blue" languages often have a higher values of the US RAS and US RAVS parameters than the "red" ones. Also often "blue" Wikipedia editions have more articles.

Additionally, we also assessed the US RAVS values for each month of 2021. Figure 4 shows the relevant heatmap. As we can see, each language version has some differences in relative popularity of US-related topics depending on the month. In some languages, there are months when US-related Wikipedia articles were especially popular. For example, Haitian (hr) Wikipedia has relatively high values of US RAVS parameter in February and March of 2021. However, usually, those differences are not significant within the selected language version of Wikipedia.

Extended and interactive versions of charts presented in this paper can be found at [Supplementary materials, 2022](#).

Language	Abbr.	Articles	PPCRW	VPC	US RAS	US RAVS
Albanian	sq	85,575	30%	69%	7.36%	2.19%
Arabic	ar	1,170,101	60%	70%	13.69%	6.11%
Armenian	hy	292,230	40%	75%	6.34%	3.62%
Azerbaijani	az	183,728	38%	65%	3.85%	2.28%
Belarusian	be	218,884	1%	33%	2.59%	4.41%
Bengali	bn	123,097	17%	38%	5.93%	2.09%
Bulgarian	bg	281,858	48%	82%	8.03%	8.47%
Burmese	my	103,346	9%	60%	1.03%	2.42%
Chinese	zh	1,279,573	7%	68%	6.73%	9.61%
Croatian	hr	211,957	34%	58%	7.66%	6.68%
Czech	cs	504,439	69%	86%	7.17%	8.12%
Danish	da	279,739	42%	87%	8.47%	10.21%
Dutch	nl	2,091,679	34%	69%	4.82%	11.92%
English	en	6,506,889	60%	80%	16.31%	31.58%
Estonian	et	227,891	41%	78%	5.33%	3.94%
Finnish	fi	531,864	61%	92%	9.60%	11.54%
French	fr	2,427,282	76%	68%	9.13%	16.26%
Georgian	ka	160,873	38%	72%	5.11%	3.95%
German	de	2,694,463	73%	77%	7.03%	16.01%
Greek	el	210,657	42%	86%	5.68%	7.03%
Haitian	ht	64,619	1%	12%	37.82%	13.29%
Hebrew	he	316,462	57%	88%	11.84%	13.21%
Hindi	hi	151,479	8%	93%	3.10%	1.72%
Hungarian	hu	504,989	65%	82%	7.13%	13.52%
Icelandic	is	54,280	8%	58%	7.01%	8.83%
Indonesian	id	622,051	51%	85%	6.87%	4.33%
Italian	it	1,756,899	86%	92%	11.37%	18.89%
Japanese	ja	1,328,108	94%	97%	6.23%	6.10%
Kazakh	kk	231,609	31%	93%	4.95%	0.72%
Khmer	km	9,199	13%	74%	1.86%	1.59%
Korean	ko	590,723	67%	87%	9.33%	6.74%
Kyrgyz	ky	81,377	20%	80%	12.06%	0.90%
Lao	lo	4,007	5%	43%	2.59%	5.76%
Latvian	lv	113,341	27%	75%	7.71%	5.67%
Lithuanian	lt	202,915	45%	79%	3.33%	5.12%
Malay	ms	357,983	10%	78%	3.37%	2.07%
Mongolian	mn	21,187	36%	85%	7.22%	2.87%
Nepali	ne	32,100	2%	54%	5.06%	2.96%
Norwegian (Bokmål)	no	590,618	46%	87%	7.66%	8.16%
Persian	fa	913,195	69%	72%	13.73%	11.43%
Polish	pl	1,524,085	81%	81%	7.28%	10.31%
Portuguese	pt	1,092,595	54%	80%	11.31%	15.32%
Romanian	ro	430,674	40%	75%	5.49%	6.89%
Russian	ru	1,824,504	85%	59%	5.92%	14.10%
Serbian	sr	659,144	40%	67%	7.13%	6.87%
Sinhalese	si	17,849	5%	81%	2.57%	2.07%
Slovak	sk	240,861	44%	83%	3.91%	6.61%
Slovenian	sl	176,444	35%	78%	4.45%	4.29%
Somali	so	8,234	11%	38%	4.48%	2.63%
Spanish	es	1,779,399	70%	60%	10.85%	16.79%
Swedish	sv	2,556,199	57%	89%	2.92%	10.68%
Tagalog	tl	43,109	7%	71%	6.85%	3.86%
Thai	th	148,010	68%	95%	6.84%	4.75%
Turkish	tr	487,817	54%	42%	11.10%	7.55%
Ukrainian	uk	1,157,506	24%	85%	7.20%	7.35%
Urdu	ur	171,536	1%	35%	20.60%	2.06%
Uzbek	uz	142,454	14%	44%	14.17%	2.43%
Vietnamese	vi	1,272,845	71%	92%	3.78%	5.06%

Table 1. Wikipedia language versions and their estimated parameters used in the study. PPCRW - people in a primary country reading that language Wikipedia, VPC - Views from the primary country, US RAS - United States related articles share, US RAVS - United States related articles views share

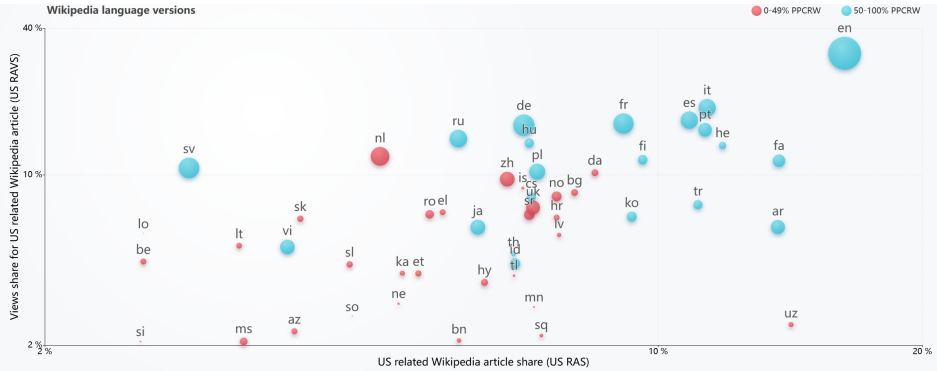


Figure 3. Bubble chart with selected Wikipedia languages which uses US RAS and US RAVS values for location. Size of the item indicates the number of articles in Wikipedia language, color indicate range of PPCRW values. More extended and interactive version on: <https://data.lewoniewski.info/americanization/bubble>

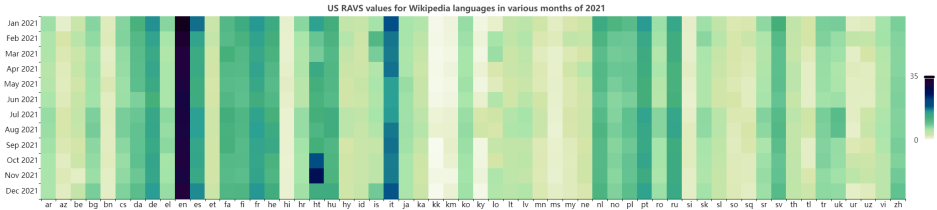


Figure 4. US RAVS values for each considered language Wikipedia version and each month of 2021. Interactive version on: <https://data.lewoniewski.info/americanization/heatmap>

Discussion

We know from previous studies (Miquel-Ribé & Laniado, 2021) that local Wikipedias have a bias toward content associated with their own cultures. English Wikipedia has the most content about the United States (here, we consider values for Haitian and Urdu Wikipedias skewed by small size and/or automated edits), and its views of American topics are twice that of any other Wikipedia. This is a reassuring “common sense” confirmation that our data is measuring something that has relevance to the real world. As English Wikipedia is significantly edited and ready by Americans, it stands to reason that it will be heavily focused on American topics. It would be a red flag if our data suggested that any other Wikipedia, or cultural region represented by them, is more Americanized. Topics related to the United States will, as expected, always be of most interest to American citizens first. On the other hand, English Wikipedia is also said to be an “international project”, and the reported value might be underrepresenting the Americanization of the “English-speaking world”, being moderated by the interests of volunteers and readers from other, less Americanized regions.

Next, our data suggest that Americanization is strongest in Europe and Latin America. We can consider it a confirmation of the outcome hypothesized in prior studies (Barjot, 2003; van Elteren, 2006, p. 133), although no prior research was able to prove it with quantifiable data. Our data further suggest that the Americanization of Latin America is very significant, as at 16% (popularity) and 11% (article count) it is higher than the values for Europe. The latter have been aggregated, following the Inglehart-Welzel model, into several clusters: Catholic Europe (10% popularity, 6% article count), Protestant Europe (without English-speaking countries; 9% and 6%, respectively), and Orthodox Europe (7% and 6%, respectively).

Unfortunately, the limitations of our data make it impossible to comparatively analyze the Americanization of Latin American countries (in the case of Venezuela or Cuba different from Brazil or Argentina? This cannot be answered with the Wikidata approach). Further, we cannot completely untangle the influence of the European countries of Spain and Portugal here. Nonetheless, the majority (80-90%) of views of Spanish and Portuguese Wikipedias originate from Latin America, so we expect our findings to allow for relatively reliable generalization for that region. Readers interested in exploring this topic might want to turn to Rodolfo and Yang (2019).

Within Europe, we can see some significant differences. Baltic, Balkan, and Eastern (usually, Orthodox) countries appear to be less Americanized than the Western ones. Italy emerges as the most Americanized country in the world (with almost 19% popularity and 11% article count). What could explain this? Some insights are found in Pasquino (2005), however the special issue it is a part of is focused on political globalization, tackling issues such as American influence on Italian foreign policy or the judicial system. Arguably, more relevant might be the dimension of cultural globalization, yet that still awaits a comprehensive study, to be pieced from parts of the puzzle, such as the extensive penetration of Italian culture by Disney comics (Stajano, 1999). Perhaps the best overview is the chapter on Italy, “Containing Modernity, Domesticating America in Italy”, by David W. Ellwood in Stephan et al. (2005).

Italy is followed by France and Germany (with about 16% popularity and 8% article count). Relatively high Americanization of Russia (14% popularity, although only 6% article count) might suggest the counterintuitive effect of Kremlin propaganda, which often reports on American topics (although usually in a pejorative way). For all three countries, the corresponding chapters in Stephan et al. (2005) make worthwhile reading (here, “A Special German Case of Cultural Americanization” by Alexander Stephan, “From French Anti-Americanism and Americanization to the “American Enemy”?” by Richard J. Golsan, and “From Cold War to Wary Peace: American Culture in the USSR and Russia” by Marsha Siefert).

Within Europe, relatively high values (10% or more in popularity) are also reported for Hungary (14% popularity, 7% article count), the Netherlands (12%, 5%), Finland (12%, 10%), Sweden (11%, 3%), Poland (10%, 7%) and Denmark (10%, 8%). Stephan et al. (2005) contain chapters on Sweden, Denmark, and Poland, but the topic of the Americanization of Hungary and Finland, like that of the Americanization of many other countries mentioned subsequently, appears to still lack any definitive treatment in the existing literature.

We find the smaller Americanization values of certain European countries quite intriguing. American topics appear least popular in Estonia (4% popularity, 5% article count), Georgia (4%, 5%), Slovenia (4%, 4%), Belarus (4%, 3%) Lithuania (5%, 3%), Latvia (5%, 8%), Slovakia (7%, 3.9%) and Croatia (7%, 8%). We can describe such countries as Balkan, Eastern European, post-communist, or Orthodox, although such groupings are not fully inclusive (it does, however, include all three Baltic states). Reasons for this might be an artifact of the Cold War, although as shown by high values for Russia (as well as Hungary), or the middling values for Poland, Bulgaria, and Ukraine, the answer is likely to be much more complex. We also note the similarity between the Orthodox cluster (7% popularity and 6%, article count) and the ex-communist cluster (which includes some Asian countries and reports values of 7% and 6%, respectively).

Norway emerges as the least Americanized Western European country (with 7% popularity and 6% article count), followed by Serbia (7%, 7%), Romania (7%, 5%), and Greece (7%, 6%).

In turn, we observe lower values of Americanisation in South & West Asia and Islamic clusters. S&WA cluster reports values of 5% popularity and 6% article count, and Islamic, 4% and 9%. The major outlier here is Hebrew Wikipedia representing Jewish Israel (not counted as part of these clusters), with 13% popularity and 12% article count, which confirms that Israel can be seen as a significantly “Western” country.

Within the S&WA cluster, Chinese Wikipedia has the highest Americanization value (10%, 7%). It should be noted that Chinese censorship and restrictions on Wikipedia use likely significantly impact the use of Wikipedia within China, and Chinese Wikipedia editors are unproportionally represented by Taiwanese and Hong Kong editors. To what degree the content of Chinese Wikipedia is biased towards non-mainland interests is an intriguing issue for future research. Korean (7%, 9%) and Japanese Wikipedias follow (6%, 6%), then Vietnamese (5%, 4%) and Thai (5%, 7%).

Within the Islamic cluster, Persian Wikipedia has surprisingly high values (11%, 14%), followed by Turkish (8%, 11%) and Arabic (6%, 14%) projects. Indonesian, Malay, and Wikipedias for Central Asian Muslim countries (Kazakhstan, Pakistan, etc.) report some of the lowest values, as do smaller S&WA ones (ex. for Cambodian and Sri Lankan projects), with the popularity of American topics at 1% to 5%. This range also includes the Hindi Wikipedia, however, as noted previously, we caution against attempts to generalize from that project to the country of India (where the majority of Wikipedia readers seem to participate in the English Wikipedia).

Finally, we find that the aggregated value for developed countries is higher than those for developing (10%, 7% versus 6%, 7%).

Conclusions

We believe that despite the outlined limitations, our findings show that Wikidata can be already used to draw valuable inferences about the entire world, allowing quantification of social science concepts that previously were considered not realistically measurable.

While our findings should still be considered exploratory, we believe they are already useful for understanding the global patterns of Americanisation (as well as the broader topic of globalization). We show that the interest in American topics is not universal. Broadly understood Western, developed countries are more Americanized (more interested in topics related to America) than the rest of the world. This is the first global, quantitative confirmation of issues often hypothesized, or assumed, in the literature on Americanization and related phenomena.

Therefore, while it is possible some of our findings are affected by unforeseen biases with Wikidata, we are convinced the broad outline presented here is correct. We are more cautious when it comes to individual countries, however, the data set is robust enough in most cases as well (47 out of 58 studied Wikipedias meet two more of our reliability criteria of having more than 40,000 articles, at least 50% of the expected population uses, and at least 50% of the expected country views. For example, with regards to the reliability of outliers such as Russian and Persian Wikipedias, whose data indicates very high popularity of American content among its readers (compared to its cultural clusters, here - Orthodox and Islamic), those Wikipedias are very large (with almost 2 million and 1 million articles, respectively, many times that of the global edition of Britannica), and they represent the main reference work used by, roughly, two-thirds of Internet users in these countries (in other words, over half of their citizens). On the other end of the spectrum (Wikipedias with low popularity of American content among its readers), the same can be said about Indonesian, Thai, Vietnamese, or Lithuania Wikipedias - while smaller, they are still large when compared to traditional reference works, and significantly used by citizens of associated countries. In other words, we believe that most of what we observed, presented, and analyzed here represents not Wikidata biases but real-world patterns in need of deeper understanding.

Further research is recommended with regard to country and culture-specific issues, as indicated by our data. For example, high values reported for Russia, and Iran, or low values reported for Baltic states or Central Asian countries are intriguing and do not appear to be significantly explored in the existing literature. Why is the popularity of American topics in Russia twice that of Ukraine? In Iran, twice that of the Arab world? In Finland and Sweden, twice that of Norway? We hope that future studies, altered to such issues by our exploratory research, will provide the answers in the foreseeable future.

References

- Ait-Mlouk, A., & Jiang, L. (2020). Kbot: a knowledge graph based chatbot for natural language understanding over linked data. *IEEE Access*, 8, 149220–149230.
- Apic, G., Betts, M. J., & Russell, R. B. (2011). Content disputes in wikipedia reflect geopolitical instability. *PloS one*, 6(6), e20902.
- Avieson, B. (2019). Breaking news on wikipedia: collaborating, collating and competing. *First Monday*.
- Barjot, D. (2003). Americanisation: cultural transfers in the economic sphere in the twentieth century. *Entreprises et Histoire*(32), 41.

- Beck, U., Sznaider, N., & Winter, R. (2003). *Global america?: The cultural consequences of globalization* (Vol. 8). Liverpool University Press.
- Berghahn, V. R. (2010). The debate on 'americanization' among economic and cultural historians. *Cold War History*, 10(1), 107–130.
- Bilić, P. (2015). 'searching for a centre that holds' in the network society: Social construction of knowledge on, and with, english wikipedia. *New media & society*, 17(8), 1258–1276.
- Boldi, P., & Monti, C. (2016). Cleansing wikipedia categories using centrality. In *Proceedings of the 25th international conference companion on world wide web* (pp. 969–974).
- Callahan, E. S., & Herring, S. C. (2011). Cultural bias in wikipedia content on famous persons. *Journal of the American society for information science and technology*, 62(10), 1899–1915.
- Clark, J., Faris, R., & Heacock Jones, R. (2017). Analyzing accessibility of wikipedia projects around the world. *Berkman Klein Center Research Publication*(2017-4).
- Conrad, P. (2014). *How the world was won: the americanization of everywhere*. Thames & Hudson.
- Crothers, L. (2021). *Globalization and american popular culture*. Rowman & Littlefield Publishers.
- Debska, M. (2010). A brief history of americanization.
- Donlan, R. (2010). *From the editor: Wikipedia and the digital divide* (Vol. 20) (No. 5). Taylor & Francis.
- Fabbrini, S. (2004). Layers of anti-americanism: Americanization, american unilateralism and anti-americanism in a european perspective. *European Journal of American Culture*, 23(2), 79–94.
- Graham, M., Straumann, R. K., & Hogan, B. (2015). Digital divisions of labor and informational magnetism: Mapping participation in wikipedia. *Annals of the Association of American Geographers*, 105(6), 1158–1178.
- Greenstein, S., & Zhu, F. (2012). Is wikipedia biased? *American Economic Review*, 102(3), 343–48.
- Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., ... others (2020). World values survey: round seven—country-pooled datafile. *Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat*.
- Jemielniak, D. (2014). *Common knowledge?: An ethnography of wikipedia*. Stanford University Press.
- Jemielniak, D., & Wilamowski, M. (2017). Cultural diversity of quality of information on wikipeidias. *Journal of the association for information science and technology*, 68(10), 2460–2470.
- Keegan, B. C. (2013). A history of newswork on wikipedia. In *Proceedings of the 9th international symposium on open collaboration* (pp. 1–10).
- King, M. J., & O'Boyle, J. (2002). Idea and value exchange worldwide: American culture in the world. *The Journal of American Culture*, 25(1/2), 1.
- Kleeman, J. (2015). The wikipedia wars: does it matter if our biggest source of knowledge is written by men? *New Statesman*, 26. Retrieved

- from <http://www.newstatesman.com/lifestyle/2015/05/wikipedia-has-colossal-problem-women-dont-edit-it>
- Kolbitsch, J., & Maurer, H. A. (2006). The transformation of the web: How emerging communities shape the information we consume. *J. Univers. Comput. Sci.*, 12(2), 187–213.
- Konieczny, P. (2020). Macro-level differences in participation in sharing economy: Factors affecting contributions to the collective intelligence wikipedia platform across different asian countries. *Asian Journal of Social Science*, 48(1-2), 115–149.
- Konieczny, P., & Klein, M. (2018). Gender gap through time and space: A journey through wikipedia biographies via the wikidata human gender indicator. *New Media & Society*, 20(12), 4608–4633.
- Kroes, R., et al. (2003). The internet: An instrument of americanization? *Global America*, 235–253.
- Laouenan, M., Bhargava, P., Eyméoud, J.-B., Gergaud, O., Plique, G., & Wasmer, E. (2022). A cross-verified database of notable people, 3500bc-2018ad. *Scientific Data*, 9(1), 1–19.
- Lemmerich, F., Sáez-Trumper, D., West, R., & Zia, L. (2019). Why the world reads wikipedia: Beyond english speakers. In *Proceedings of the twelfth acm international conference on web search and data mining* (pp. 618–626).
- Lewoniewski, W. (2022). Identification of important web sources of information on wikipedia across various topics and languages. *Procedia Computer Science*, 207, 3290–3299.
- Lewoniewski, W., Węcel, K., & Abramowicz, W. (2019). Multilingual ranking of wikipedia articles with quality and popularity assessment in different topics. *Computers*, 8(3), 60.
- Lewoniewski, W., Węcel, K., & Abramowicz, W. (2022). Companies in multilingual wikipedia: Articles quality and important sources of information. In *Special sessions in the advances in information systems and technologies track of the conference on computer science and intelligence systems* (pp. 48–67).
- Li, A., & Farzan, R. (2020). Collaboration of open content news in wikipedia: The role and impact of gatekeepers. In *Companion proceedings of the web conference 2020* (pp. 802–805).
- Martin, B. (2018). Persistent bias on wikipedia: Methods and responses. *Social Science Computer Review*, 36(3), 379–388.
- Meta-Wiki, W. (2021). *Community insights report*. Retrieved from https://meta.wikimedia.org/wiki/Community_Insights/Community_Insights_2021_Report
- Miquel, M., Laniado, D., & Kaltenbrunner, A. (2021). The role of local content in wikipedia: A study on reader and editor engagement. *Área Abierta*, 21(2), 123–151.
- Miquel-Ribé, M. (2019). The sum of human knowledge? not in one wikipedia language edition. *Wikipedia 20*.

- Miquel-Ribé, M., & Laniado, D. (2018). Wikipedia culture gap: quantifying content imbalances across 40 language editions. *Frontiers in physics*, 6, 54.
- Miquel-Ribé, M., & Laniado, D. (2021). The wikipedia diversity observatory: helping communities to bridge content gaps through interactive interfaces. *Journal of Internet Services and Applications*, 12(1), 1–25.
- Mora-Cantalops, M., Sánchez-Alonso, S., & García-Barriocanal, E. (2019). A systematic literature review on wikidata. *Data Technologies and Applications*, 53(3), 250–268.
- Mueller, A. C. (2004). *German pop culture: How "american" is it?* University of Michigan Press.
- Mulina, N. I. (2018). What leads ukrainian university students to use wikipedia?
- Nehring, H. (2004). ‘westernization’: A new paradigm for interpreting west european history in a cold war context. *Cold War History*, 4(2), 175–191.
- Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F. Å., & Lanamäki, A. (2014). Wikipedia in the eyes of its beholders: A systematic review of scholarly research on wikipedia readers and readership. *Journal of the Association for Information Science and Technology*, 65(12), 2381–2403.
- Pasquino, G. (2005). Italy and america: politics and culture: Americanization of italian politics? *Journal of modern Italian studies*, 10(1), 3–9.
- Pfundner, A., Schönberg, T., Horn, J., Boyce, R. D., Samwald, M., et al. (2015). Utilizing the wikidata system to improve the quality of medical content in wikipedia in diverse languages: a pilot study. *Journal of medical Internet research*, 17(5), e4163.
- Ramsey, P. J. (2015). Americanization. *The Wiley Blackwell Encyclopedia of Race, Ethnicity, and Nationalism*, 1–2.
- Rask, M. (2008). The reach and richness of wikipedia: Is wikinomics only for rich countries? *First Monday*.
- Rodolfo, O., & Yang, A. S. (2019). *Americanizing latino politics, latinoizing american politics*. Routledge.
- Rogers, R., Sendjarevic, E., et al. (2012). Neutral or national point of view? a comparison of srebrenica articles across wikipedia’s language versions. *Proc. Wikipedia Academy*.
- Schröter, H. G. (2005). *Americanization of the european economy: A compact survey of american economic influence in europe since the 1800s*. Springer Science & Business Media.
- Shaw, A., & Hargittai, E. (2018). The pipeline of online participation inequalities: The case of wikipedia editing. *Journal of communication*, 68(1), 143–168.
- Stajano, F. (1999). *Disney comics from italy*. Retrieved from <https://citeseeerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.5021> (This is a revised version of an essay Francesco Stajano, “Disney-serier i Italien”, NAFS(k)uriren 30-32, 1999-2000.)
- Stephan, A. (2005). *The americanization of europe: Culture, diplomacy, and anti-americanism after 1945*. Berghahn Books.

- Supplementary materials. (2022). *Supplementary materials for this research*. <https://data.lewoniewski.info/americanization/>.
- Taraborelli, D., & Ciampaglia, G. L. (2010). Beyond notability. collective deliberation on content inclusion in wikipedia. In *2010 fourth ieee international conference on self-adaptive and self-organizing systems workshop* (pp. 122–125).
- Tharani, K. (2021). Much more than a mere technology: A systematic review of wikidata in libraries. *The Journal of Academic Librarianship*, 47(2), 102326.
- Tota, A. P. (2010). *The seduction of brazil: The americanization of brazil during world war ii*. University of Texas Press.
- Turki, H., Hadj Taieb, M. A., Shafee, T., Lubiana, T., Jemielniak, D., Aouicha, M. B., ... others (2022). Representing covid-19 information in collaborative knowledge graphs: the case of wikidata. *Semantic Web*, 13(2), 233–264.
- van Elteren, M. (2006). *Americanism and americanization: A critical history of domestic and global influence*. McFarland.
- Vučetić, R. (2018). *Coca-cola socialism. americanization of yugoslav culture in the sixties*. Central European University Press.
- Waagmeester, A., Willighagen, E. L., Su, A. I., Kutmon, M., Gayo, J. E. L., Fernández-Álvarez, D., ... Koehorst, J. J. (2021). A protocol for adding knowledge to wikidata: aligning resources on human coronaviruses. *BMC biology*, 19(1), 1–14.
- Wiggers, H. (2018). The struggle of small and non-western wikipedia editions.
- Worku, Z., Bipat, T., McDonald, D. W., & Zachry, M. (2020). Exploring systematic bias through article deletions on wikipedia from a behavioral perspective. In *Proceedings of the 16th international symposium on open collaboration* (pp. 1–22).
- World Bank Data Team. (2019). *New country classifications by income level: 2019–2020*. Retrieved from <https://blogs.worldbank.org/opendata/new-country-classifications-income-level-2019-2020>
- Young, A. G., Wigdor, A. D., & Kane, G. C. (2020). The gender bias tug-of-war in a co-creation community: Core-periphery tension on wikipedia. *Journal of Management Information Systems*, 37(4), 1047–1072.