# Mapping the Americanization of English in space and time

Bruno Gonçalves[1], Lucía Loureiro-Porto[2], José J. Ramasco[3], David Sánchez[3]*

1 Center for Data Science, New York University, 60 5th Avenue, New York, NY 10012, United States of America, 2 Departament de Filologia Espanyola, Moderna i Clàssica, Universitat de les Illes Balears (UIB), 07122 Palma de Mallorca, Spain, 3 Institute for Cross-Disciplinary Physics and Complex Systems IFISC (CSIC-UIB), 07122 Palma de Mallorca, Spain

* david.sanchez@uib.es
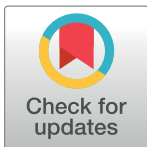
Abstract

As global political preeminence gradually shifted from the United Kingdom to the United States, so did the capacity to culturally influence the rest of the world. In this work, we analyze how the world-wide varieties of written English are evolving. We study both the spatial and temporal variations of vocabulary and spelling of English using a large corpus of geolocated tweets and the Google Books datasets corresponding to books published in the US and the UK. The advantage of our approach is that we can address both standard written language (Google Books) and the more colloquial forms of microblogging messages (Twitter). We find that American English is the dominant form of English outside the UK and that its influence is felt even within the UK borders. Finally, we analyze how this trend has evolved over time and the impact that some cultural events have had in shaping it.

Introduction

With roots dating as far back as Cabot's explorations in the 15th century and the 1584 establishment of the ill-fated Roanoke colony in the New World, the British empire was one of the largest empires in Human History. At its zenith, it extended from North America to Asia, Africa and Australia deserving the moniker "the empire on which the sun never sets". However, as history has shown countless times, empires rise and fall due to a complex set of internal and external forces. In the case of the British empire, its preeminence faded as the United States of America –one of its first colonies– took over the dominant role in the global arena.

As an empire spreads so does the language of its ruling class because the presence of a prestigious linguistic variety plays an important role in language change [1, 2]. Thanks to both its global extension, late demise, and the rise of the US as a global actor, the English language enjoys an undisputed role as the global *lingua franca* serving as the default language of science, commerce and diplomacy [3, 4] (see Fig 1). Given such an extended presence, it is only natural that English would absorb words, expressions and other features of local indigenous languages resulting in dozens of dialects and topolects (language forms typical of a specific area) such as "Singlish" (Singapore), "Hinglish" (India), Kenyan English [5], and, most importantly, American English [6] a variety that includes within itself several other dialects [7, 8].

word appears in the Google Books corpus each year.

The transfer of political, economical and cultural power from Great Britain to the United States has progressed gradually over the course of more than half a century, with World War II being the final stepping stone in the establishment of American supremacy. The cultural rise of the United States also implied the exportation of their specific form of English resulting in a change of how English is written and spoken around the world. In fact, the "Americanization" of (global) English is one of the main processes of language change in contemporary English [9]. Although it is found to work along with other processes such as colloquialization and informalization [10], the spread of American features all over the globe is generally assumed to be result of the American 'leadership' in change [9].

As an example, if we focus on spelling, some the original differences between British and American English orthography (most of which are the result of Webster's reform [11]) are somehow blurred and, for instance, the tendency for verbs and nouns to end in *-ize* and *-ization* in America is now common on both sides of the Atlantic [12]. Likewise, a tendency for Postcolonial varieties of English in South-East Asia to prefer American spelling over the British one has been observed, at least, for Nigerian English [13], Singapore and Trinidad and Tobago [14], regarding spelling and lexis, for Indian English [15] and the Bahamas [16], regarding syntax, and for Hong Kong [17], regarding phonology. In addition, a growing tendency for Americanization has been observed for Philippine English, which, despite being rooted in American English, has experienced a rise in the frequency of American forms [18]. Although this Americanization is found in different registers, web genres have been highlighted as a text-type where American forms are preferred [19]. Electronic communication has indeed been considered to play a role in linguistic uniformity [20]. It is in this sense that this paper will make a contribution to the study of the Americanization of English, since a corpus of 213, 086, 831 geolocated tweets will be used to study the spread of American English spelling and vocabulary around the globe, including regions where English is used as a first, second and foreign language.
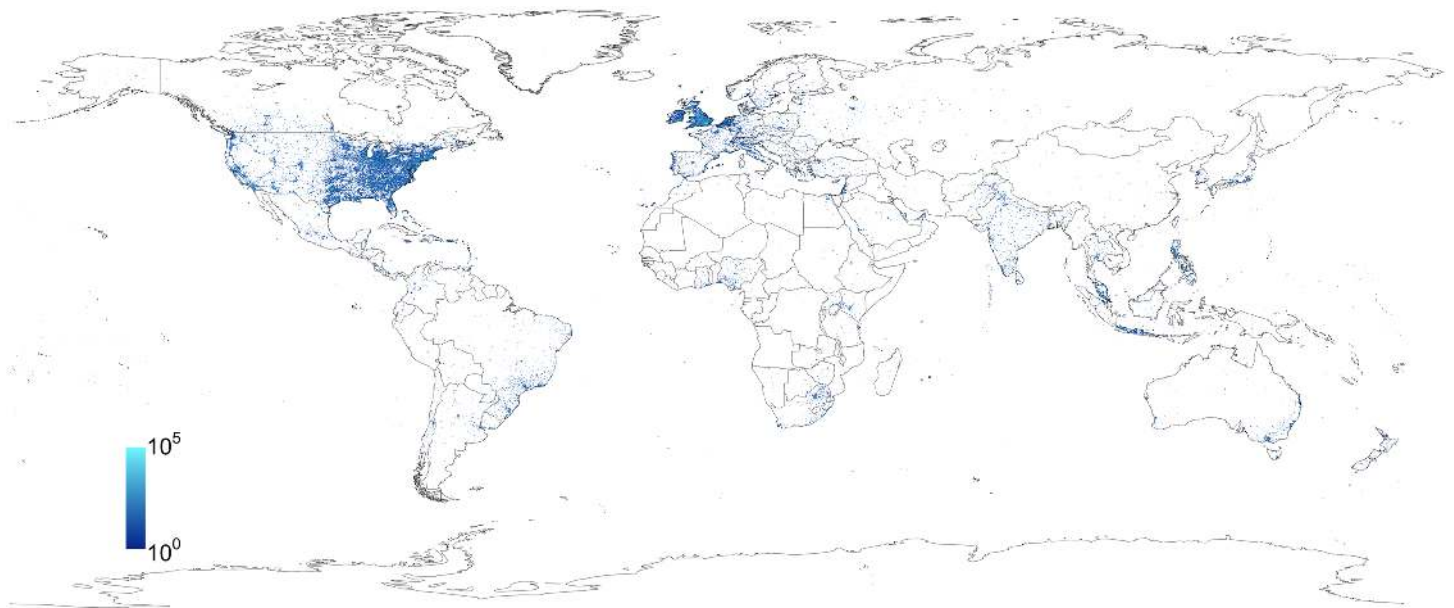

Fig 1. English tweets. A heatmap showing the location of geolocated English tweets in our dataset that match our keywords.

https://doi.org/10.1371/journal.pone.0197741.g001

The study of diatopic variation using Twitter datasets is a relatively new subject [21]. The use of geotagged microblogging data [22] allows the quantitative examination of linguistic patterns on a worldwide scale, in automatic fashion and within conversational situations. The global extension and the real time availability of the data constitute major methodological advantages over more traditional approaches like surveys and interviews [23]. Importantly, the resulting corpora are publicly available [24], although due to their nature most of the literature has been concerned with lexical variation (for an exception that addresses semantic and syntactic variation, see Ref. [25]). Thus, different variables can be mapped after carefully removing lexical ambiguities [26]. A Bayesian approach shows good agreement between baseline queries and survey responses [27]. Machine learning techniques applied to Twitter corpora reveal the existence of superdialects [28, 29], which can be further analyzed with dialectometric techniques [30]. Linguistic evolution in social media appears to be strongly connected to demographics [31]. Age and gender issues can be additionally introduced in the analysis [32]. Moreover, an investigation of lexical alternations unveils hierarchical dialect regions in the United States [33]. Twitter can be also employed in the study of specific varieties departing from the standard form [34]. However, online social media are more suitable for a synchronic approximation to language variation. If one aims at understanding the diachronic evolution of language, we need a corpus well established over time. This is available with the Google Books database [35], which has already been used for the analysis of relative frequencies that characterize word fluxes [36, 37] or the applicability of Zip's and Heaps's law with different scaling regimes [38]. Here, we will complement our Twitter study of the Americanization of English with an analysis of the dynamic process that is taking place since 1800.

In this paper we analyze how English is used around the world, in informal contexts, using a large scale Twitter dataset. Due to the written nature of our corpus we consider in detail both how vocabulary and spelling of common words varies from place to place in order to understand how American cultural influence is spreading around the world. We complement this synchronic analysis with a diachronic view of how the prevalence of British and American vocabulary and spelling have evolved over time in British and American publications using the Google Books dataset.

Methods

Datasets

The goal of this manuscript is to analyze how English is used across both time and space. We study the geographical variation of English by using the Twitter Decahose from which we collect [39] all tweets written in English between May 10, 2010 and Feb 28, 2016 that contain geolocation information (S1 File). The language is detected using Chromium Compact Language Detection library as in Ref. [39]. One might ask whether those tweets arising from outside the English-speaking world are from native-English speakers residing in or visiting those countries. In fact, it has been shown that the vast majority of the tweets in a given location arises from speakers residing in that country [40, 41]. Therefore, our approach is reliable to a very good extent but has the limitations common to geographical studies based on Twitter datasets, as illustrated, e.g., in Ref. [42].

The temporal evolution of English is analyzed using the Google Books dataset [35] of books published by both British and American publishers (S2 File). The dataset contains the number of times individual words were used in books scanned by Google and dating back to the 15th century. However, due to the poor statistics in earlier periods, we restrict our analysis to the period between 1800 and 2010. Importantly, for a given year in our dataset each of the two corpora (British and American) include at least several million word instances, ranging from a

minimum 18.4 million for US in 1800 (98.5 million in the UK) to a maximum of 8.2 billion for the US in 2000 (2.2 billion in the UK). To the best of our knowledge, these are the largest corpora of this kind ever gathered. Admittedly, the Google Books dataset has its own shortcomings (prolific authors, overrepresentation of scientific texts, etc. [43]) but these are expected to affect both corpora equally.

Our two main data sources are different in nature: Twitter contains more colloquial expressions, while the language recorded in the books is more formal. As a result, these two sources, in combination, can provide a useful perspective on the spatio-temporal patterns developed or developing in English. Yet, in our study we do not distinguish between the two types. Rather, our objective is to show that, despite the fact that both corpora have different register features, the English Americanization is evident in the two of them.

In our analysis, we consider two factors of differentiation between American and British English: spelling and vocabulary with different word lists used for each case. A given concept is expressed with two lexical alternations (either British or American) or two different spellings. The complete list of words and expressions employed in each case can be found, respectively, in Table 1 (vocabulary) and Table 2 (spelling). It is the result of compiling information in reference books [12] and online sources such as the Oxford Dictionaries (https://en.oxforddictionaries.com/usage/british-and-american-terms). In order to make sure that the items in the sources really represented British or American English, all the words in the list were subsequently checked in two widely used representative corpora of both varieties, namely, the British National Corpus (BYU-BNC) and the Corpus of Contemporary American English (COCA) [44, 45]. Only pairs of words in which one of the members exhibits a significantly higher frequency in either of the two varieties were considered for inclusion in the list. Thus, for example, *railway* is significantly more frequent in BYU-BNC whereas *railroad* is significantly more frequent in COCA, which makes the pair valid for our purposes. Verbs ending in *-ize/-ise* (and the corresponding nouns in *-isation* and *-ization*) are, on the contrary, not considered because, as stated in Sec. 1, the spelling *-ize* is common on both sides of the Atlantic, with the exception of *analyse/analyze* and *paralyse/paralyze*, which do form part of our list insofar as they have been found to be reliable British/American spellings. Inflectional forms (e.g., *solicitor, solicitors, solicitor's, solicitors'*) as well as derived (e.g., *amphitheater*) and compound forms were also included in the search (e.g., *sportscenter*). We are aware of the fact that departing from a list of Britishisms and Americanisms may appear to be a simplification of reality, because some Postcolonial Englishes may opt for vernacular forms, rather than for the British or the American one. However, our purpose is not to describe all varieties of English but to measure which of the two main inner-circle varieties is predominant in territories where English is used as a first, second and foreign language.

A word of caution is here needed. There exists certain semantic ambiguity in the selected lexical alternations. Nevertheless, this is an unavoidable effect that is inherent to computational studies on language variation (e.g., Refs. [28] and [33]). Our compromise is to keep the overall polysemy to a small degree while at the same time providing a selection of words sufficiently large to allow for a quantitative analysis. We have checked that the variants of each pair in our list can be exchanged, quite generally, in many contexts and are thus valid for the aim of this work.

Metrics

Language variation in space is analyzed by means of a grid of cells of $0.25° \times 0.25°$ spanning the globe. The polarization, V_w^c, for a concept w in cell c during the data collection period is

Table 1. Word list comprising British and American vocabulary variants.

British	American
railway	railroad
MA dissertation	MA thesis
doctoral thesis	doctoral dissertation
draughts	checkers
abseil	rappel
antenatal	prenatal
anticlockwise	counterclockwise
aubergine	eggplant
barrister, solicitor	attorney
biscuit	cookie
car park	parking lot
caster sugar, icing sugar	confectioner's sugar, powdered sugar
corn flour	corn starch
cupboard	closet
demister	defroster
drawing pin	thumbtack
Father Christmas	Santa Claus
handbrake, hand brake	emergency brake
hire purchase	installment plan
inside leg	inseam
mobile phone	cell phone
motorway	expressway, freeway
nappy	diaper
notice board	bulletin board
number plate	license plate
plasterboard	wallboard
polystyrene	styrofoam
porridge	oatmeal
perspex	plexiglass
pushchair	stroller
rubbish	garbage
skirting board	baseboard
spring onion	green onion
sticky tape	scotch tape
sweets	candy
torch	flashlight
tracksuit	sweatsuit
trousers	pants
valuer	appraiser
wellington boots, wellingtons	rubbers, rubber boots, rain boots
windscreen	windshield
lorry	truck
chemist's	drug store
elastic band	rubber band
estate agent	realtor
cot	crib
off-licence	liquor store
crayfish	crawfish

Table 2. Word list comprising British and American spelling variants.

British	American
skilful	skillful
wilful	willful
fulfil, fulfils	fulfill, fulfills
instil, instils	instill, instills
appal, appals	appall, appalls
flavour	flavor
mould	mold
moult	molt
smoulder	smolder
moustache	mustache
centre	center
metre	meter
theatre	theater
analyse	analyze
paralyse	paralyze
defence	defense
offence	offense
pretence	pretense
revelling, revelled	reveling, reveled
travelled, travelling	traveled, traveling
travelle	traveler
marvellous	marvelous
plough	plow
aluminium	aluminum
jewellery	jewelry
pyjamas	pajamas
whisky	whiskey
neighbour	neighbor
honour	honor
colour	color
behaviour	behavior
labour	labor
humour	humor
favour	favor
harbour	harbor
tumour	tumor
vigour	vigor
rumour	rumor
rigour	rigor
demeanour	demeanor
clamour	clamor
odour	odor
armour	armor
endeavour	endeavor
parlour	parlor
vapour	vapor
saviour	savior

(*Continued*)

Table 2. (Continued)

British	American
splendour	splendor
fervour	fervor
savour	savor
valour	valor
candour	candor
ardour	ardor
rancour	rancor
succour	succor
arbour	arbor
catalogue	catalog
analog	analog
acknowledgement	acknowledgment
goitre	goiter
foetus	fetus
paediatrician	pediatrician
oesophagus	esophagus
manoeuvre	maneuver
oestrogen	estrogen
anaemia	anemia

defined as the ratio:

$$V_w^c = \frac{A_w^c - B_w^c}{A_w^c + B_w^c},$$

(1)

where A_w^c (B_w^c) is the number of American (British) forms of the concept w observed in cell c. The polarization is then constrained to be in the $[-1, 1]$ domain, with -1 corresponding to purely British and 1 being purely American forms.

The polarization of each cell, V^c, is then determined by taking the average polarization over all words observed in cell c:

$$V^c = \frac{\sum_w V_w^c}{W^c},$$

(2)

where W^c is the number of different words observed in cell c. Similarly, the polarization score of a country is defined as the average polarization taken over all the cells within that country. By considering the average polarization we are able to compare countries of varying sizes.

In the case of Twitter, the polarization signal is measured over the complete time period of the database since, as it is not long enough to allow for large variations in the language use patterns. On the other hand, when the time evolution of written language is considered with Google Books, space is not relevant, beyond the country of origin of the published book, and we add an index referring to the year y considered. The polarization V^y is then defined as:

$$V^y = \frac{\sum_w V_w^y}{W^y},$$

(3)

where V_w^y is the concept polarization for year y and W^y refers to all the books published in the country considered, the US or the UK, during year y.
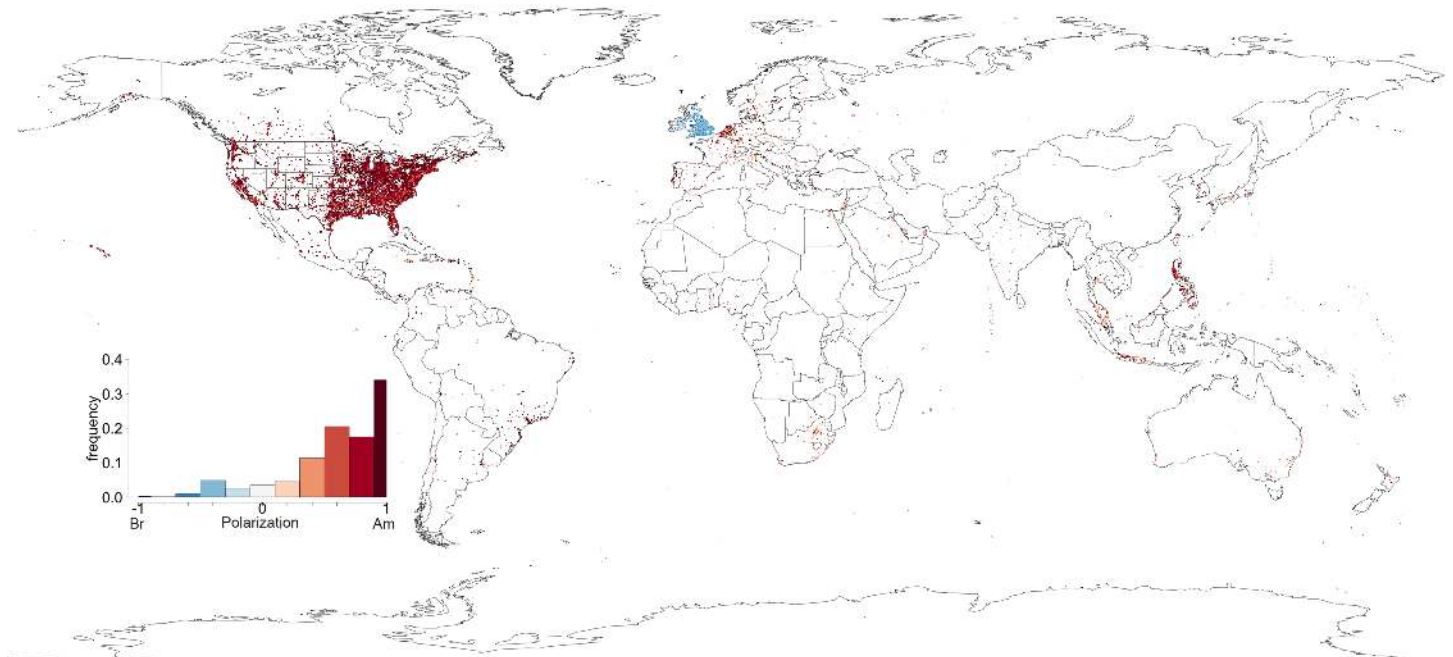
Fig 2. Vocabulary. The polarization ratio of each cell around the world according to the vocabulary used within each cell. The inset barplot is an histogram of the number of cells as a function of the ratio.

Results

The analysis yields 30, 898, 072 tweets matching the list of words. A heatmap illustrating the geographical distribution of matching tweets is shown in Fig 1. The relation between bias of the data and population may lead to undesired fluctuations in our maps. This is particularly true in the US and the UK. On the other hand, in countries where English is not the mother tongue the real problem is the lack of data in certain cells. In the latter case, few English tweets may have a strong influence in the final value of the polarization for a given cell. We fix this issue in a twofold way. First, we impose for each cell a minimum threshold of ten matches from our list of concepts. Second, we consider a sufficient number of cells. A balance between these quantities causes fluctuations to average out and they do not have a strong influence in the overall results.

Let us start by considering how the vocabulary used for common terms such as *lorry/truck* or *motorway/freeway* changes around the world by defining the ratio of each cell as given by Eq (2). The results are plotted in Fig 2. Unsurprisingly, we find that the British Islands tend to be blue while the United States is predominantly red as befits the representatives of each trend. Interestingly, Western Europe where English teaching has traditionally followed British norms the American influence is undeniable. Most areas are depicted in various shades of red while some of the largest international metropolises such as Madrid, Paris, Amsterdam, Berlin, Milan or Rome are visible in light shades, indicating intermediate values, in no doubt due to their role as touristic and transportation hubs, see Fig 3 (left). A more marked British influence is easily seen in former colonies such as South Africa, Australia, New Zealand ("the only large areas in the Southern hemisphere where English is spoken as a native language" [12]), and which have reached a very advanced phase of development, according to Schneider's 2007 Dynamic Model [46]) or India (where English is spoken as a non-native language, but which has followed an exonormative model, i.e., strongly based on British rules [47]) displaying large
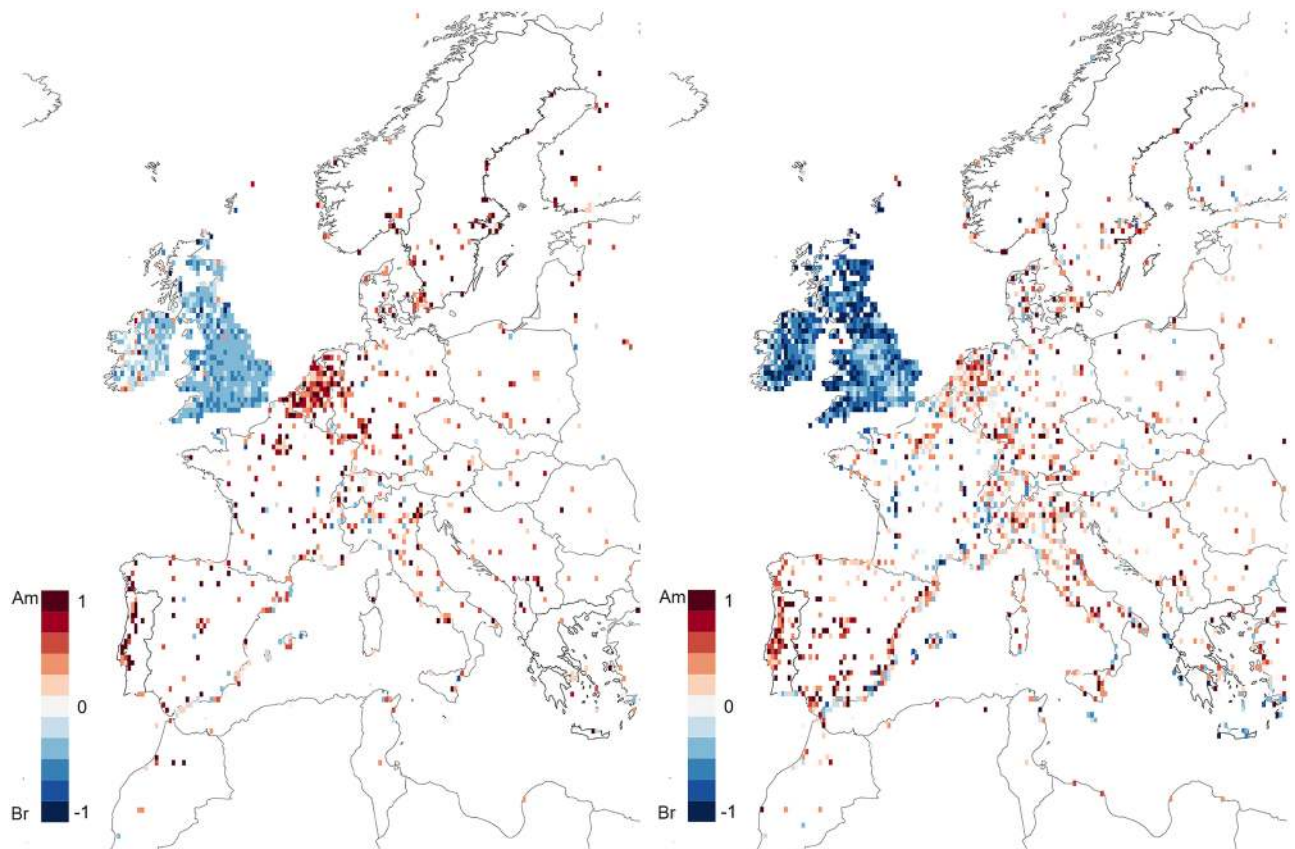
Fig 3. Europe. Side by side comparison of the vocabulary (left) and spelling (right) results for countries in continental Europe. The tension between British spelling and American vocabulary is clearly visible by the shift towards lighter shades of blue and darker shades of red between the left and the right plots.

areas of blue side by side with tell-tale patches of white in the most international areas such as Pretoria, Melbourne, Sidney, Auckland, New Delhi or Mumbai. Furthermore, countries such as the Philippines (one of the few postcolonial varieties of English with an American superstratum [46]), as well as Taiwan, South Korea and Japan (where English is spoken as a second language) attest their strong American influence with full displays of red.

Regarding spelling, the case for American influence becomes even stronger as displayed in Fig 4. The British Isles attain significantly lighter shades of blue as do the former British colonies with South Africa, Australia and New Zealand becoming predominately red. This dichotomy between spelling and vocabulary, illustrated in Fig 3 for Europe, is perhaps a testament to the conflicting forces of traditional formal education and media influence. Individuals who studied in school systems that subscribe to the British form of English are more prone to continue writing words in the way they originally learned them. However, through the influence of American dominated television and film industries they have acquired new (American) vocabulary. This can be clearly seen in Fig 5 where we plot the average polarization for both vocabulary and spelling for 30 countries around the world, including countries belonging to Kachru's [48] Inner Circle, i.e., where English is spoken as a native language (e.g., UK, Ireland), Outer circle, i.e., where English is spoken as a second language (e.g., India, South Africa) and the expanding circle, i.e., where English is spoken as a foreign language (e.g., Portugal, Finland, Russia). Interestingly enough, in all expanding circle territories, American
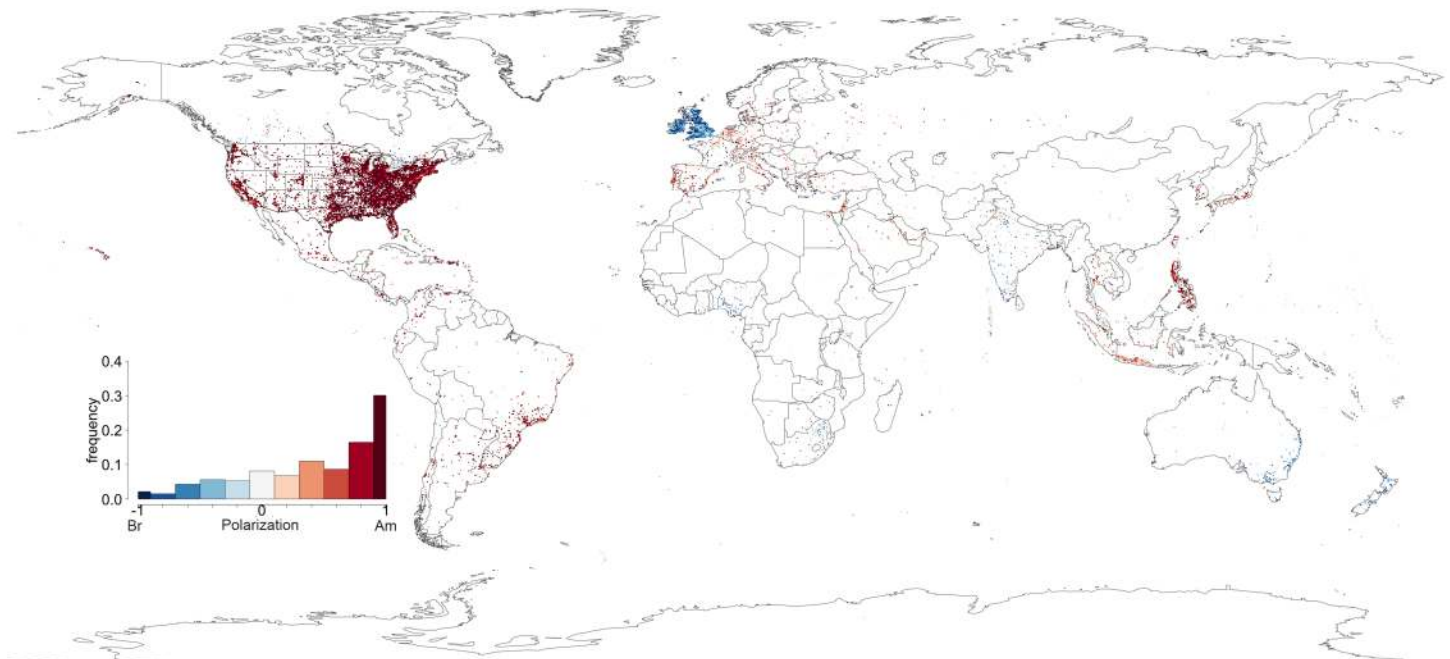
Fig 4. Spelling. The polarization of each cell around the world according to the spelling used within each cell. The inset barplot is an histogram of the number of cells as a function of the ratio observed.

orthography and vocabulary dominate, and the same happens, obviously, in the United States and in the Philippines, a former American colony. The bottom part of the figure includes Inner and Outer circle varieties, where American vocabulary is also chosen over British forms, with the notable exception of India, UK and Ireland, whose green bars are always towards the left hand (British) side of the ratio spectrum. India's alignment with the UK is clearly the result of an exonormative model and postcolonial prescriptivism in this former colony of the United Kingdom [47, 49]. Surprisingly, we find that in some ex-colonies which still hold strong ties with the British empire, such as South Africa, Australia and New Zealand, the drift towards American vocabulary is unmistakable.

We now consider a temporal view of how English as a language is evolving. Using the word counts provided by the Google Books digitalization efforts, we measure the vocabulary and spelling average ratio per year [Eq (3)] for books published by American and British publishing houses. Considering the averages suffices for our purposes since averaging over the huge number of word instances in our two corpora ensures negligible error bars. An analysis of the resulting timelines as shown in Fig 6 provides several interesting insights. First, we can see that the divergence in spelling between the American and British forms has significantly increased in the last 200 years. Indeed, from this time series we can pinpoint the beginning of the trend to around 1828 when Noah Webster published *An American Dictionary of the English Language* [50] with the explicit goal of systematizing the way in which English was written in America. As [51] puts it: "He is certainly responsible for establishing (though not inventing) the common differences between traditional British and American spellings" the final *-or* versus *-our* in *color*, *labor*, *savor*, and the like; *-er* versus French *-re* in *theater*, *center*, *meter*; and the simplification of final *-ck* as in *physic*, *music*, *logic*. This is now considered to have been the first American English dictionary and it started the Merriam-Webster series of Dictionaries that is still dominant today. The US vocabulary curve follows a similar but less pronounced
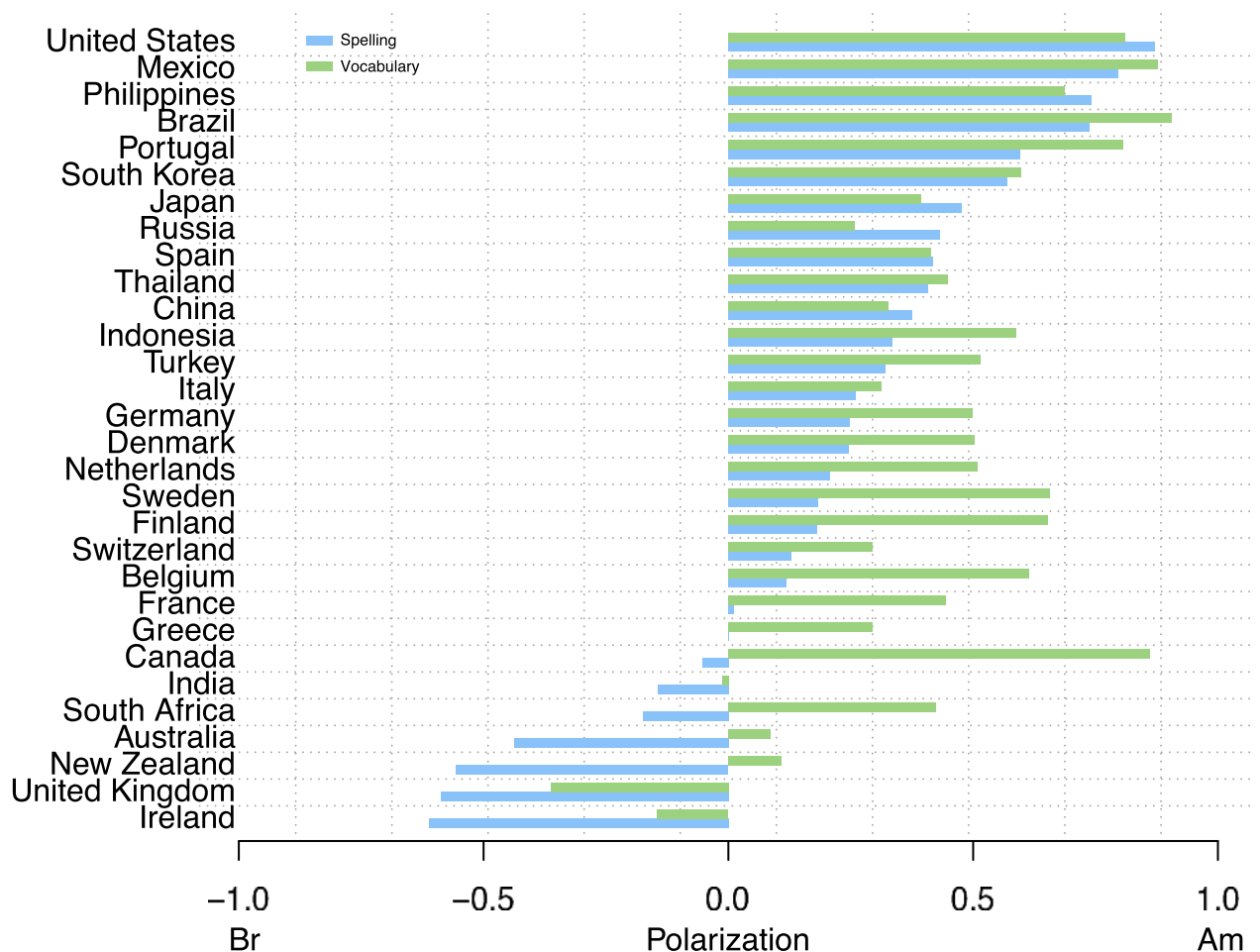
Fig 5. Countries. Vocabulary and spelling polarization ratio by country.

trend as it takes longer for new words to be created than for people to agree on a common spelling form.

Another interesting feature of these timelines is the pronounced "Britishization" of American English in the years following World War II as seen by the declining slope that extends until after 1960. This can likely be explained by the large influx of European migrants that moved to America in search of a better life away from a destroyed or warring Europe. In the immediate aftermath of WWII, Congress passed the War Brides Act in 1946 and the Displaced Persons Act in 1948 to facilitate the immigration to the US by the people affected by the war. It is estimated that between 1941 and 1950 over 1 Million people [52], mostly of European descent, immigrated to the United States that at the time had a population of 150 million. In the following decade, this number doubled to over 2 Million [53].

Interestingly, while the ratio timelines within the United Kingdom had been towards becoming ever more British, we find a significant change of trend in the last 20 years of our dataset, corresponding to the period after the fall of the Berlin Wall and the end of the Cold War that left America as the world's only superpower. A position that was only reinforced with the advent and popularization of the Internet just a decade later. It is the status quo resulting from the aftermath of this trend that we are able to observe in the Twitter analysis above.
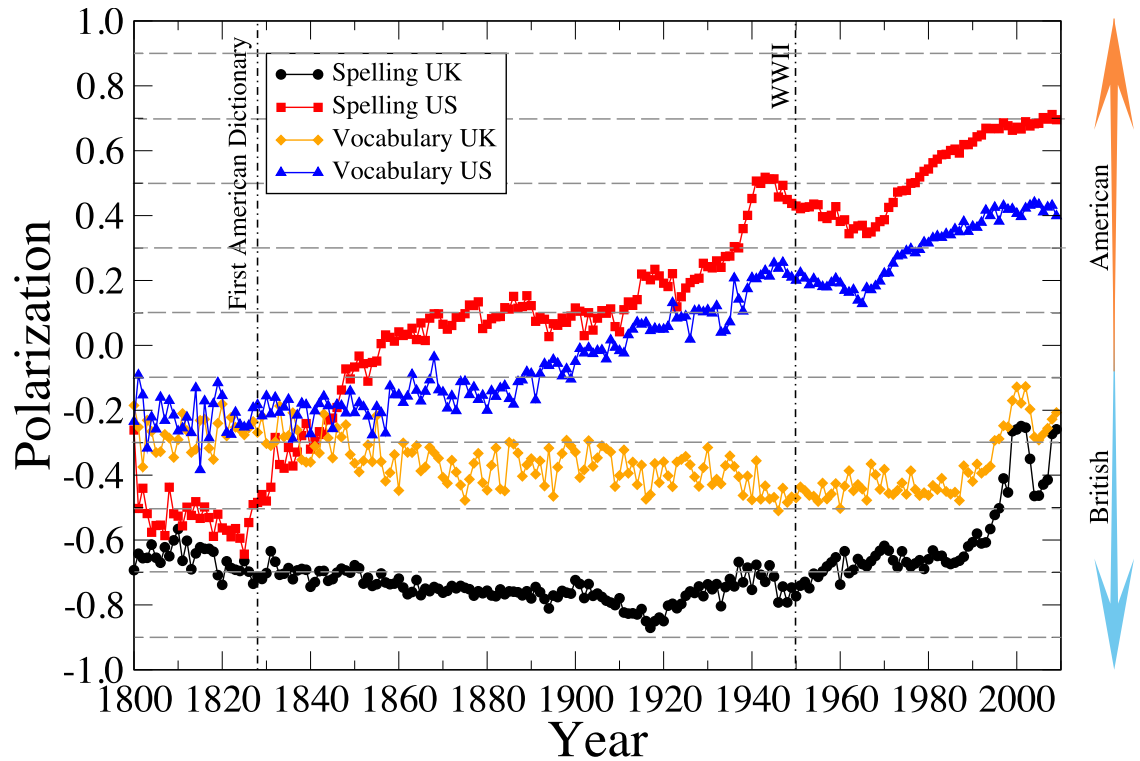
Fig 6. Americanization of English over time. Averaged polarization ratio of vocabulary and spelling for books published by US and UK publishing companies in the 1800 – 2010 period.

https://doi.org/10.1371/journal.pone.0197741.g006

Conclusions

The way in which languages evolve in time and change from place to place has long been the focus of much interest in the linguistic community. With the advent of new and extensive corpora derived from large scale online datasets we are now able to take on a more quantitative approach to tackling this fundamental question. In this work we analyze two datasets that, when taken together, are able to provide a bird's eye view of the way English usage has been changing over time and in different countries.

The picture we are able to paint is particularly stark. The past two centuries have clearly resulted in a shift in vocabulary and spelling conventions from British to American. This trend is especially visible in the decades following WWII and the fall of the Berlin Wall. These historical events left the US as the only superpower and the influence it has exerted because of this on other cultures is evident at all levels. The presence of the American way of life, including cultural representations (literature, music, cinema and pop-culture products such as TV shows, computer games, etc.) can be felt all over the globe, as explained in Ref. [54] about the role played by the MTV as a powerful propagator of pop-culture. Naturally, the spread of the American culture is accompanied by the American linguistic variety, which ends up affecting (global) English, as we have shown, with some clearly identifiable exceptions. Indeed, when we consider the current status quo as seen through the lens of Twitter, it becomes clear that only in the countries where British influence has been strongest, such as ex-colonies with a strong exonormative influence (in Schneider's terms [46]), are British conventions still dominant to some degree.

It should be noted that both datasets we utilize in our analysis are intrinsically biased. Books are typically written by cultural elites. Also, despite their increasing democratization, GPS enabled mobile devices are, in many countries, only available to middle and higher economic strata. As a result, there are certainly factors of linguistic evolution we are missing but the fact that both datasets agree on the general picture means that we are able to capture, at the very least, the underlying trends.

Supporting information

S1 File. Twitter corpus. Python code example.
(PDF)

S2 File. Google Book corpus. Data source.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Bruno Gonçalves, Lucía Loureiro-Porto, José J. Ramasco, David Sánchez.

Data curation: Bruno Gonçalves.

Investigation: Bruno Gonçalves, Lucía Loureiro-Porto, José J. Ramasco, David Sánchez.

Methodology: Bruno Gonçalves, Lucía Loureiro-Porto, José J. Ramasco, David Sánchez.

Project administration: José J. Ramasco.

Resources: José J. Ramasco.

Supervision: Lucía Loureiro-Porto, José J. Ramasco, David Sánchez.

Validation: Bruno Gonçalves, Lucía Loureiro-Porto, José J. Ramasco, David Sánchez.

Visualization: Bruno Gonçalves.

Writing – original draft: Bruno Gonçalves, Lucía Loureiro-Porto, José J. Ramasco, David Sánchez.

Writing – review & editing: Bruno Gonçalves, Lucía Loureiro-Porto, José J. Ramasco, David Sánchez.

References

1. Labov W. The social motivation of a sound change. Word. 1963; 19:273. https://doi.org/10.1080/00437956.1963.11659799

2. Fishman JA. Bilingualism with and without diglossia; diglossia with a without bilingualism. Word. 1967; XXIII:29.

3. Crystal D. English as a Global Language. Cambridge University Press; 2003.

4. Jenkins J, Leung C. English as a Lingua Franca. The Companion to Language Assessment IV. 2013; 13(95):1605. https://doi.org/10.1002/9781118411360.wbcla047

5. Mesthrie R, Bhatt RM. The Study of New Linguistic Varieties. Cambridge University Press; 2008.

6. Grieve J. Regional Variation in Written American English. Cambridge University Press; 2016.

7. Pederson L. Dialects. In: Algeo J, editor. The Cambridge History of the English Language. Cambridge University Press; 2001. p. 253.

8. Wolfram W, Shelling N. American English: Dialects and Variation. Wiley-Blackwell; 2015.

9. Leech G, Hundt M, Mair C, Smith N. Change in Contemporary English: A Grammatical Study. Cambridge University Press; 2009.

10. Baker P. American and British English. Divided by a Common Language? Cambridge University Press; 2017.

11. Algeo J. External History. In: Algeo J, editor. The Cambridge History of the English Language. Cambridge University Press; 2001.

12. Gramley S, Pätzold KM. Survey of Modern English. Routledge; 2003.

13. Awonusi VO. The Americanization of Nigerian English. World Englishes. 1994; 13:75. https://doi.org/10.1111/j.1467-971X.1994.tb00284.x

14. Hänsel EC, Deuber D. Globalization, postcolonial Englishes, and the English language press in Kenya, Singapore, and Trinidad and Tobago. World Englishes. 2013; 32:338. https://doi.org/10.1111/weng.12035

15. Davydova J. Indian English quotatives in a real-time perspective. In: Seoane E, Suárez-Gómez C, editors. World Englishes: New theoretical and methodological considerations. Benjamins; 2015. p. 173.

16. Hackert S. Pseudotitles in Bahamian English: A Case of Americanization? Journal of English Linguistics. 2015; 43:143. https://doi.org/10.1177/0075424215577966

17. Hansen Edwards JG. Accent preferences and the use of American English features in Hong Kong: a preliminary study. Asian Englishes. 2016; 18:197. https://doi.org/10.1080/13488678.2016.1225482

18. Fuchs R. The Americanization of Philippine English: Recent diachronic change in spelling and lexis. Philippine ESL Journal. 2017; 19.

19. Mukherjee J. Response to Davies and Fuchs. English Worldwide. 2015; 36:34. https://doi.org/10.1075/eww.36.1.02muk

20. Venezky RL. Spelling. In: Algeo J, editor. The Cambridge History of the English Language. Cambridge University Press; 2001. p. 340.

21. Nguyen D, Dogrüoz AS, Rosé CP, de Jong F. Computational sociolinguistics: A survey. Computational linguistics. 2016; 42:537. https://doi.org/10.1162/COLI_a_00258

22. Melo F, Martins B. Automated geocoding of textual documents: A survey of current approaches. Transactions in GIS. 2016; 21:3. https://doi.org/10.1111/tgis.12212

23. Chambers JK, Trudgill P. Dialectology. Cambridge University Press; 1998.

24. Malmasi S, Zampieri M, Ljubesšić N, Nakov P, Ali A, Tiedemann J. Discriminating between similar languages and Arabic dialect identification: A report on the Third DSL Shared Task. Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial3). 2016; p. 1–14.

25. Kulkarni V, Perozzi B, Skiena S. Freshman or Fresher? Quantifying the Geographic Variation of Language in Online Social Media. Proceedings of the Tenth International AAAI Conference on Web and Social Media. 2016; p. 613.

26. Russ B. Examining large-scale regional variation through online geotagged corpora. ADS Annual Meeting. 2012;.

27. Doyle G. Mapping Dialectal Variation by Querying Social Media. In: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics; 2014. p. 98–106.

28. Gonçalves B, Sánchez D. Crowdsourcing Dialect Characterization Through Twitter. PLOS ONE. 2014; 9:E112074. https://doi.org/10.1371/journal.pone.0112074 PMID: 25409174

29. Gonçalves B, Sánchez D. Learning about Spanish Dialects through Twitter. RILI. 2016; 28:65–75.

30. Donoso G, Sánchez D. Dialectometric analysis of language variation in Twitter. Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial4). 2017; p. 16–25.

31. Eisenstein J, O'Connor B, Smith NA, Xing EP. Diffusion of lexical change in social media. PLOS ONE. 2014; 9:E113114. https://doi.org/10.1371/journal.pone.0113114 PMID: 25409166

32. Pavalanathan U, Eisenstein J. Confounds and consequences in geotagged Twitter data. Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2015; p. 2138–2148.

33. Huang Y, Guo D, Kasakoff A, Grieve J. Understanding U.S. regional linguistic variation with Twitter data analysis. Computers, Environment and Urban Systems. 2016; 54.

34. Blodgett SL, Green L, O'Connor B. Demographic Dialectal Variation in Social Media: A Case Study of African-American English. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016; p. 1119–1130.

35. Michel JB, Shen YK, Aiden AP, Veres A, Gray MK, Pickett JP, et al. Quantitative Analysis of Culture Using Millions of Digitized Books. Science. 2011; 331(6014):176–182. https://doi.org/10.1126/science.1199644 PMID: 21163965

36. Pedersen AM, Tenenbaum JN, Havlin S, Stanley HE, Perc M. Languages cool as they expand: Allometric scaling and the decreasing need for new words. Sci Rep. 2012; 2:943. https://doi.org/10.1038/srep00943

37. Pechenick EA, Danforth CM, Dodds PS. Is language evolution grinding to a halt? The scaling of lexical turbulence in English fiction suggests it is not. Journal of Computational Science. 2017; 21:24. https://doi.org/10.1016/j.jocs.2017.04.020

38. Gerlach M, Altmann EG. Stochastic Model for the Vocabulary Growth in Natural Languages. Phys Rev X. 2016; 3:021006.

39. Mocanu D, Baronchelli A, Perra N, Gonçalves B, Vespignani A. The Twitter of Babel: Mapping World Languages through Microblogging Platforms. PLOS One. 2013; 8:E61981. https://doi.org/10.1371/journal.pone.0061981 PMID: 23637940

40. Lamanna F, Lenormand M, Salas-Olmedo MH, Romanillos G, Gonçalves B, Ramasco JJ. Immigrant community integration in world cities. PLOS ONE. 2016; 13:e0191612. https://doi.org/10.1371/journal.pone.0191612

41. Bassolas A, Lenormand M, Tugores A, Gonçalves B, Ramasco JJ. Touristic site attractiveness seen through Twitter. EPJ Data Science. 2016; 5:12. https://doi.org/10.1140/epjds/s13688-016-0077-1

42. Leetaru KH, Wang S, Cao G, Padmanabhan A, Shook E. Mapping the global Twitter heartbeat: The geography of Twitter. First Monday. 2013; 18:5. https://doi.org/10.5210/fm.v18i5.4366

43. Pechinek EA, Danforth CM, Dodds PS. Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution. PLOS ONE. 2015; 10:e0137041. https://doi.org/10.1371/journal.pone.0137041

44. Davies M. BYU-BNC (based on the British National Corpus from Oxford Universtiy Press); 2004.

45. Davies M. The Corpus of Contemporary American English: 520 million words 1990-present; 2008.

46. Schneider EW. Postcolonial English. Varieties around the World. Cambridge University Press; 2007.

47. Schneider EW. English around the World: An Introduction. Cambridge University Press; 2011.

48. Kachru BB. Standards, codification and sociolinguistic realism: the English language in the outer circle. In: English in the world: Teaching and learning the language and literatures. Cambridge University Press; 1985. p. 11–30.

49. Collins P. Grammatical colloquialism and the English quasi-modals: a comparative study. In: Marín-Arrese JI, Carretero M, Hita JA, van der Auwera J, editors. English modality: Core, Periphery and Evidentiality. Mouton de Gruyter; 2013.

50. Webster N. An American dictionary of the English language. S. Converse; 1828.

51. Cassidy FG, Hall JH. Americanisms. In: Algeo J, editor. The Cambridge History of the English Language IV: English in North America. Cambridge University Press; 2001.

52. Census Bureau. Statistical Abstract of the U.S. 1950; 1950.

53. Census Bureau. Statistical Abstract of the U.S. 1960; 1960.

54. Jones S. MTV: The Medium was the Message. Critical Studies in Media Communication. 2005; 22:83. https://doi.org/10.1080/0739318042000333734