Importance of Translators and Wikipedia in Digital Language Hierarchies within Europe

Kolya Souvorin

The Internet serves many functions that help citizens be informed and participate in democracy and the economy in the digital age—the right to access the Internet links to the right to use your native language. The EU's language policy translates all documents into its official languages, and the resulting corpus makes translation and research within these languages easier. The paper attempts to identify whether the languages of the EU have greater utility on the Internet through multiple linear regression. Economic and linguistic differences between member countries are possible explanations for the digital divide in Europe. EU official languages dovetail with high general internet utility. The study concluded that within Europe, EU language policy developed a strong Wikipedia for each EU language, which was associated with gains in content overall in those languages. Based on these results, it may be possible to increase a language's prestige by developing strong neural machine translators.

Introduction

Research into the cause of the digital divide is important for citizens. The Internet serves many functions, such as providing news, allowing communication within communities, participating in digital hiring processes, providing formal and informal education, media, shopping and selling, and a network for political organizations. Access to these resources is an immense privilege, being involuntarily excluded from them puts people at a severe disadvantage. The digital divide, or gap between those with and without access to the internet, was demonstrated within the United States during the COVID-19 pandemic (Vogels 2021). Lack of access to the internet excluded students from education. This lack of access excluded job-seekers from virtual job boards, remote interviews, and working from home. The digital divide is present throughout the globe, creating divisions between knowledge and service work in the 21st-century economy.

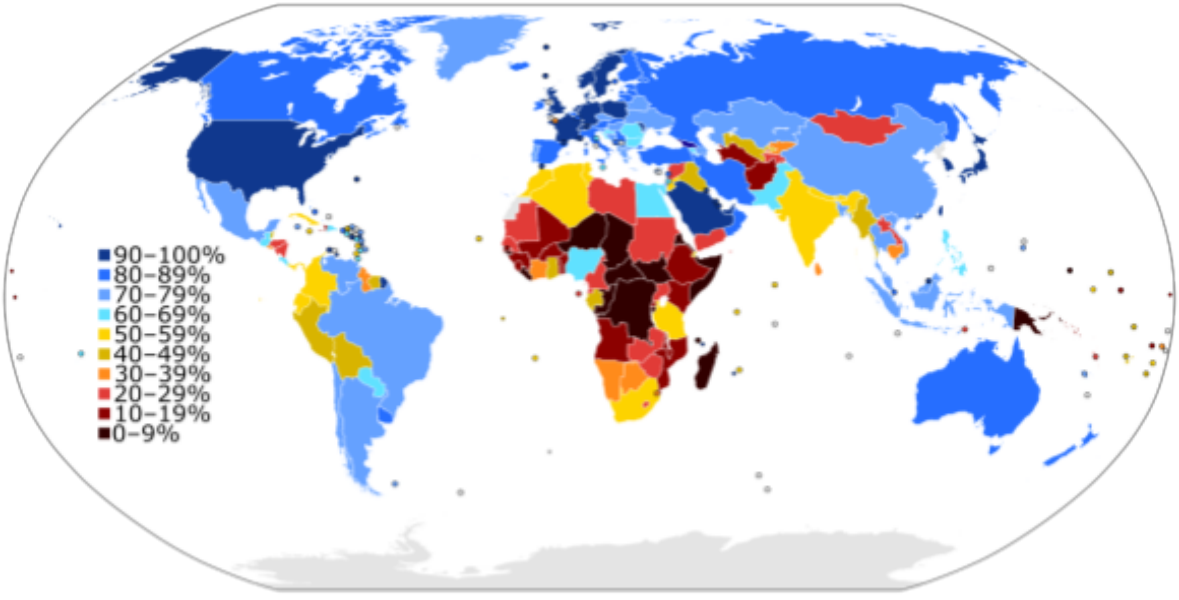

Figure 1: Global Internet Usage

Internet Rights

Human Rights Advocates have fought for your internet access. The universal right to the Internet was established by the U.N. in 2011 (Wilson 2011) and confirmed in 2016 (Vincent 2016) based on existing rights of development, expression, and assembly. The Internet Governance Forum of the U.N. holds that accessibility and linguistic diversity are principles 3 and 7 of “The Charter of Human Rights and Principles for the Internet” (Bodle 2011), deriving from these previously mentioned historical charters. Minority language rights and increased internet access are part of the U.N.’s sustainable goals. (UNSDG 2021) The Europeans hardily agree that the right is universal (GlobeScan 2010). The EU guarantees citizens a right to an accessible, affordable, open internet in their language. (Europa 2023) How the EU actualized this goal has significantly affected the multilingual web. Looking at Figure 1, it is clear that much work needs to be done to actualize this universal right.

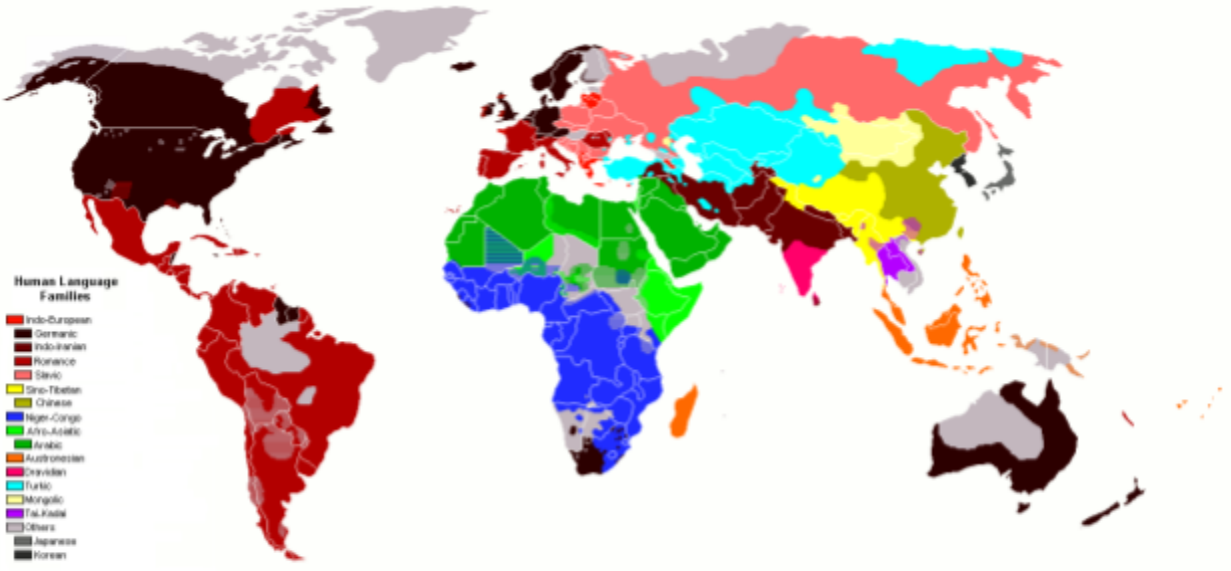

Figure 2: Language Families

Language Rights

Language Advocates are also interested in the digital divide. Using one's native language is another fundamental right established by the Universal Declaration of Human Rights. (United Nations 1948) The EU has also weighed in on language and internet rights, establishing the right to linguistic diversity and forbidding language-based discrimination in Articles 21 and 22 of the Charter of Fundamental Rights. (Dorssemont 2019) It is logical that to use the internet, it must be available in a comprehensible language. The establishment of digital rights in the 21st century has built upon these rights regarding language. Looking at Figure 2 and Figure 1, there seems to be a surface correlation between internet access and language family.

National Economic Health

Creators of national language policy are also invested in ending the digital divide. Within a country, a language may concede use to a more powerful language in education. If new research is not being generated in that language, educated citizens go to the country of the dominant language in a migration pattern called "Brain Drain" (Certo 2014). This weakens the country, as they do not benefit from the work of these highly talented citizens (Lee 2010).

Research Question

The languages of the EU are now firmly established on the internet, and its citizens are much more digitally connected than the global average. However, many European people do not speak these 24 languages and have lower connection rates.

Within Europe, *does a language being an EU official language cause it to have high digital functionality*?

Literature Review

School	Why is there inequality in internet access?
Infrastructure	Lack of electricity, cable, 3g
Literacy	Inability to comprehend the written content
Poverty	Inability to afford connection or devices
Language	Content not available in an understandable language

Researchers have found four segmented barriers to Internet access: infrastructure, poverty, literacy, and language availability (Luxton 2016). All current research into the digital divide (Young 2015) follows one of these four schools (APC 2008). Historically, The UN has focused on the literacy and infrastructure schools bridging the digital divide within the Global South (Thacker 2019). Indeed, one can only access the internet with power and reading skills (Thacker 2019). However, this logic does not make sense in the context of Europe. As of 2022, access to electricity is universal in Europe (Ritchie 2022). 99.6% of the European population can access at least 3g cellular (Taylor 2022). Adult literacy rates in each European Country are universally above 95%, and over 99% of total European adults are literate (CIA 2023). To explain variation in the functionality of European languages on the internet, we can eliminate infrastructure and literacy as causes and must look to economic or linguistic factors.

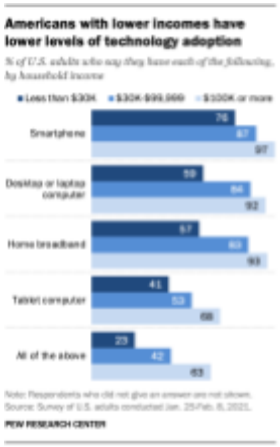

Figure 3: Technology Adoption by income

Rank	Language	Internet Users	Percentage
1	English	1,186,451,052	25.9%
2	Chinese	888,453,068	19.4%
3	Spanish	363,684,593	7.9%
4	Arabic	237,418,349	5.2%
5	Indonesian	198,029,815	4.3%
6	Portuguese	171,750,818	3.7%
7	French	151,733,611	3.3%
8	Japanese	118,626,672	2.6%
9	Russian	116,353,942	2.5%
10	German	92,525,427	2.0%

Figure 4: Top Languages of the Internet

Economic Explanation

In studies of the digital divide, income has proven to be a critical divider between those who can afford internet service and those who cannot, providing a more robust explanation than gender, geographic location (urban-rural), age, skills, awareness, political, cultural and psychological attitudes (Hilbert 2010). As seen in Figure 3, disposable income limits Americans' access to internet technologies. To realize digital rights, citizens must be able to bear the price of broadband and computers. Those who have disposable incomes have more free time for creating content.

Human Development Index

The Human Development Index (HDI) is a highly used measure of holistic economic development. Researchers have found that differences in the HDI between language groups can lead to disparities in Internet functionality in those languages (Thacker 2019). All European countries are rated very high or high on the Human

Development Index (UNDP 2022). However, the less wealthy Eastern European States have a lower HDI than the rest of Europe. Within Europe, the Gini coefficient, or measure of difference, of HDI is still relatively low, compared to global levels. The economic school argues that language groups with high HDI will have high internet functionality.

Possible Linguistic Explanations

There a several competing explanations brought up to explain differences in Internet functionality from the linguistic perspective. However, I feel that two stand out, the size of Wikipedia and whether a language is a national language of the EU.

Multilingual Europeans

As seen in Figure 4, the top 10 languages of the internet make up 76.5% of the content but represent under 50% of speakers. There is a misconception that Europeans are mostly multilingual and know these popular languages. Most EU residents do not speak English (Keating 2020). Only 16% of the population knows German or French as a second language. (Keating 2020). Additionally, citizens prefer to access the internet in their language and do not feel comfortable using many functions in their second language (Young 2015). Because the majority of Europeans attempt to use their native language on the internet, explanations must not assume high multilingualism.

Colonialization

Among the largest internets, English, French, Spanish, Portuguese, and Russian have spread through colonization. Colonial languages underwent early standardization, which was needed to administer expansive overseas governments. Additionally, the colonial languages have a long history of economic dominance over subjugated peoples. For example, the majority of current French speakers live within West and Central Africa, contributing to the French internet's large presence and the exclusion of indigenous languages there. This supports the idea that these languages would be dominant on the internet due to colonial legacy. However, languages such as Finnish and Estonian are highly connected yet have never been used in colonialism.

Latin Script

Another potential barrier to access is less support for non-Latin scripts, which could limit languages that use either the Greek or Cyrillic alphabets in Europe. Keyboards were historically available only with Latin letters, and to this day the market is significantly smaller for non-Latin layouts. Webpages and applications often require additional plugins to use non-Latin scripts. Software is exclusively written in Latin, all of which may contribute to non-Latin script languages having lower internet functionality. However, Russians are highly connected despite using Cyrillic.

Wikipedia

As discussed in the corpus section, the non-official languages in the EU have been left out of much of the gains in modern translation and have smaller Wikipedias.

These include EU regional, Eastern European, and Russian minority languages. Previous research has shown that gaining a large Wikipedia is essential in ascending to an internet language (Kornai 2013). These authors asserted that no language has even become highly used on the internet without first developing an organic Wikipedia. The free online encyclopedia provides a base of citable knowledge used to build up the corpus of content within your language. These authors claim that the size of the language's Wikipedia will have an outsize impact on the language's functionality on the internet.

EU Parallel Corpus

A corpus is a large body of written work in a language, forming an encyclopedia of thought, and bilingual corpora, such as the Rosetta Stone, form the basis for traditional translation. The European parallel body of translations is another significant development in the Internet's history. The EU spends 1% of its budget each year translating everything it publishes into 24 official languages (Ginsburgh 2022). Twenty years of these documents are publicly available to all EU citizens, and two-way communication between government and citizens occurs within all these tongues (Ginsburgh 2022). This body of text is human history's most significant parallel corpus, entirely in digital format (Euro-Lex 2016). Google, Wikipedia, and ChatGPT have broadly used this corpus to power neural network translation and general parallel bodies of knowledge. The neural networks within Google Translate and ChatGPT can now speedily and accurately translate these 24 languages at unparalleled rates. Wikipedia creators have used this bevy of digital sources to create chains of references for their

digital encyclopedias. Paired with the above machine translators, each EU Language Wikipedia continued to grow.

Theory

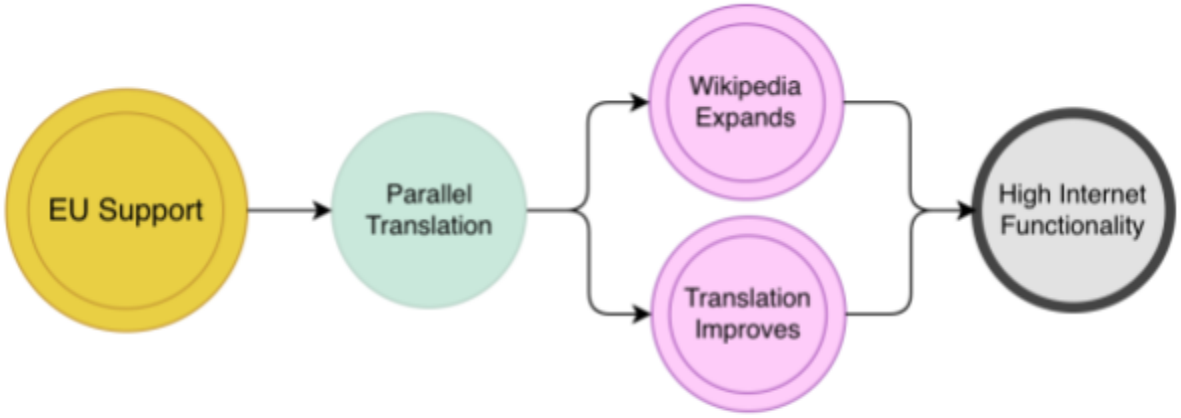

Figure 5: Theory of Highly functional Internet Languages

Timeline of Theory

EU language support will lead to higher functionality in those languages. The internet was over 80% in English during the 20th century. Then, the EU parallel corpus gave rise to large Wikipedias in the 2000s and powerful translators in the 2010s. New sections of the internet, not written in English grew. Bilingual internet users switched to their mother tongues and began creating native content. Following this, monolingual speakers adapted to the Internet, and the language solidified its functionality as an Internet language (Kornai 2013).

Alternative Explanations

There could be other factors about the language, either supporting or hindering the amount of content. Latin script has dominated computing technology, and languages with other scripts may produce less content. A colonial language may have higher internet functionality. Finally, corpus size may cause variation within content production, but EU support may not. The relative Wikipedia size will measure this effect. The economic school argues for the inclusion of HDI as a control.

Design

Methodology

I will use simple and multiple variable linear regression to determine whether a language being part of the EU is responsible for being useful on the internet. Regression analysis is a good choice, as I am modeling the conditional probability between my IV, EU status, and DV, internet functionality. The population parameters of these two variables are unknown, so they will be estimated with a linear model based on the available data. Linear regression allows for the estimation of the expected value for internet functionality, given its status in the EU. I will also account for the alternative linguistic explanations of Wikipedia Size in a language, whether a language was written in the Latin Script, and whether a language has a colonial history as control variables. Additionally, HDI will control for economic differences between the countries.

Case Selection

The selected languages are the 43 European Languages with over a million speakers. National and regional designations determine what is a dialect and what is a language. Italy and Spain have recognized the regional languages spoken at home and in less formal settings as distinct from the national languages. However, French is considered the only language in France, and Langue D'oil and Occitan are dialects. A single macro language encodes the German, Estonian, Albanian, and Serbo-Croatian language groups. Minority languages in European Russia are included if they are official in one Oblast. Selecting countries only within strengthens the internal validity of this study, while weakening the external validity.

Bias

The cutoff for inclusion at 1 million speakers is arbitrary, based only on the currently available data. Additionally, it is difficult to define how many people speak a minority language. Citizens may inflate or deflate their stated usage of a minority language based on social or political pressures (Crystal 1987). I have excluded some "dialects" of Germanic languages based on the common written standards of these tongues, but these could be considered languages in their own right. Turkish is excluded, even though some consider it within Europe, as are Caucasian languages such as Armenian. The geographical boundaries of Europe are not agreed upon, as Europe is primarily a political and cultural region within Eurasia. Other researchers may decide to include these two regions. Finally, languages such as Arabic and Vietnamese are spoken within Europe by populations descended from more recent immigrants.

However, they are not included because these languages have historically been used exclusively elsewhere. The Romani people are excluded from all aspects of European life (Amnesty 2011) and reliable data is impossible to come by about the population.

Operation:

Variable	Type	Relationship	Description
Official Language of the EU	Binary	Key IV	Of the 47 languages selected, 22 are official languages of the EU, and 25 are not.
Internet Functionality	Ratio (.5, 1.5)	Key DV	Sliding scale, represents what user can do on the internet in that language. % *Contents* divided by % *Internauts*.
Internauts	Integer	DV component	# users of the internet in a given language. One person can be an internaut in multiple languages.
% Internauts	Percentage	DV component	Internauts in a language divided by the total number of internet users in all languages.
Content	Integer	DV component	Measures presence of Wikipedia, Twitter, # webpages, and software support, serves as a comprehensive measure of virtual presence.
% Content	Percent	DV component	The total amount of content in a given language divided by the total amount of content in all languages. Normalized for multilingualism
Wikiscore	Weighted Score (0, 30)	Control	Takes # of articles, users, active users, edits, and admins as arguments and returns a number representing the deviations from English.
Latin_Script	Binary	Control	Whether a language used Latin Script
Colonial_Language	Binary	Control	Whether a language has been used in a colonial setting
HDI	Range (0, 1)	Control	Wholistic measure of development

Figure 6: Operation of Variables

A holistic score is used to determine the relative size of a language's Wikipedia. The score is the distance of a language's Wikipedia to English. To measure internet functionality, a ratio of % contents divided by % internauts is used. This variable has also been defined as *content production*.

Data Sources

Daniel Pimienta, a researcher at the Observatory of Linguistic & Cultural Diversity on the Internet, has measured virtual linguistic diversity since 1998. His most current data was released in 2023, responding to previous biases within his work

(Pimienta 2023). It is the first dataset to include languages with 1-5 million speakers. The more utilized w3school.com dataset consistently overcounts English, making it unsuitable for this research.

Analysis

Figure 7: EU Languages

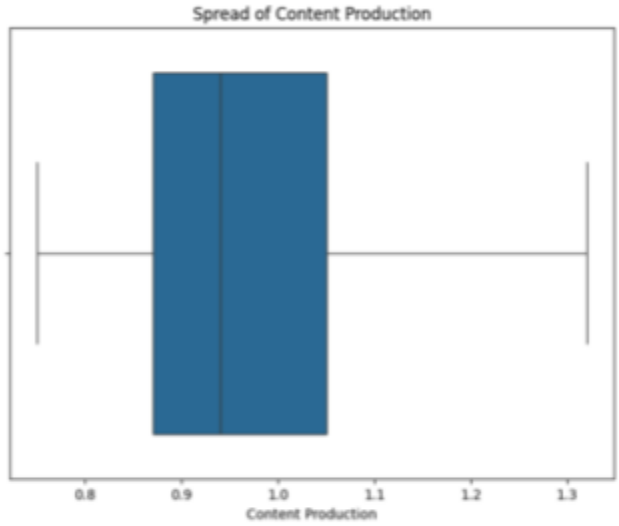

Figure 8: Content Production Univariate Analysis

Univariate Analysis

EU Membership

There are 27 members of the EU, and between them 24 languages are official. Figure 7 shows the 24 official languages along with their home countries. Austria, Belgium, and Luxembourg use the German, Dutch, and French languages but are independent EU nations.

Internet Functionality

As seen in Figure 8, Internet Functionality ranges from .8 between .65 and 1.45, with an average *Internet functionality* score of 0.96. A handful of well-represented languages cause the long tail to stretch to 1.45. However, the median language within Europe has an internet functionality of .93. This reflects the divide between the high status of the major languages on the internet compared to the lower status of the majority of languages. The IQR is 1.05 to .87, which means that a second shorter tail accounts for the lower end of the languages, with 76 % of the data falling within one standard deviation of the mean.

Wikipedia Score

The Wikiscore represents equivalent distances between the quality of the English Wikipedia and the 328 other current Wikipedias. The difference between English(1st) and Swedish(11th), Swedish and Tatar(33rd), and Tatar and Kabardian(44th, lowest in the dataset) is ten deviations. To illustrate the relative differences, active users are one component of the Wiki_Score, with English having 122,000, Swedish having 2,100, Tatar having 81, and Kabardian having 16. Figure 9 shows a sample of Wikipedia Scores for languages included in the study.[1]

Human Development Index

HDI for all language groups within Europe is uniformly above .75, excluding the highly ostracized Romani people. Figure 10 shows the continuum between Western and

[1] Full data can be found at https://github.com/souvorinkg/Internet_Languages

Eastern Europe. Turkey, Kazakhstan, and the Caucasian nations were not included in this study.

Selected Languages	Wikipedia Score
English	0
French	5.71
Swedish	10.01
Hungarian	12.28
Slovak	15.26
Macedonian	18.32
Chechen	20.58
Chuvash	22.88
Sardinian	25.57
Kabardian	29.36

Figure 9: Selected Wikipedia Score

Figure 10: HDI within Europe

Bivariate Analysis

In bivariate regression, EU membership, Colonial Language, HDI, and Wiki_Score were all significant. However, Membership in the EU only accounted for 26.6% of the variation in internet functionality.

Impact of all variables on Content Production
OLS Regression Results

Dep. Variable:		C_PROD	R-squared:			0.820
Model:		OLS	Adj. R-squared:			0.801
Method:		Least Squares	F-statistic:			43.30
Date:		Sun, 12 Nov 2023	Prob (F-statistic):			1.17e-13
Time:		14:47:29	Log-Likelihood:			66.757
No. Observations:		43	AIC:			-123.5
Df Residuals:		38	BIC:			-114.7
Df Model:		4				
Covariance Type:		nonrobust				
	coef	std err	t	P>|t|	[0.025	0.975]
const	0.6142	0.131	4.692	0.000	0.349	0.879
EU	-0.0064	0.023	-0.280	0.781	-0.053	0.040
Col_Lang	0.0860	0.028	3.123	0.003	0.030	0.142
Wiki_Score	-0.0094	0.002	-4.433	0.000	-0.014	-0.005
HDI	0.5503	0.143	3.851	0.000	0.261	0.840
Omnibus:		1.003	Durbin-Watson:			2.188
Prob(Omnibus):		0.582	Jarque-Bera (JB):			1.105
Skew:		0.335	Prob(JB):			0.575
Kurtosis:		2.592	Cond. No.			395.

Figure 11: Regression 1, EU -> Content Production

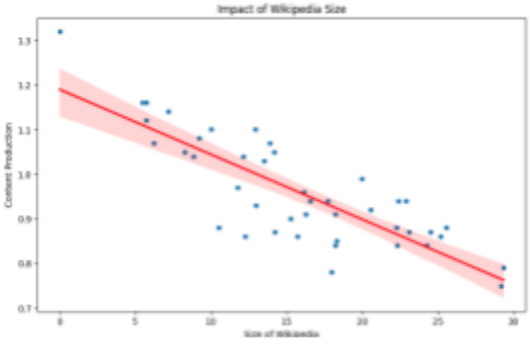

Figure 12: Impact of Wikipedia Size

Multivariate Analysis

EU Membership -> Content Production

In the Multivariate Analysis shown in Figure 11, the HDI and Wiki Scores of language communities become highly significant, as well as whether the chosen language is a colonial language. However, whether a language is within the EU loses explanatory power. Colonial Language is a binary indicator shown above. As seen by the correlation coefficients, languages used in a colonial context scored .08 points higher in internet functionality. To contextualize a jump of .08, this is equivalent to the jump in internet functionality from Chuvash to Danish or Finnish to French.

While presence in the EU had no explanatory power on Content Production, Wikipedia Score had an exceptional impact on content production. The bivariate relationship between these two variables is highly linear, as seen in Figure 12. In the initial analysis shown in Figure 11, this relationship is both strong and leads to significant changes in Internet functionality, as decreasing Wiki_Score by 1 (moving towards the English Wiki) leads to almost a full .01 increase in Internet Functionality.

HDI also has high predictive power in several cases. An increase in HDI by .05 is associated with a jump in .03 for Internet Functionality. The difference in status can be seen comparing the Albanian or Macedonian language's internet functionality (Two countries with lower HDI) with the higher status of Romanian and Hungarian (two countries with higher HDI). However, HDI does not have high explanatory power in bivariate analysis. For example, Regional Dutch and Northern Italian Languages have some of the highest HDI among populations in the world, but all have some of the lowest Internet Functionality in the study.

OLS Regression Results

Dep. Variable:	Wiki_Score	R-squared:	0.669
Model:	OLS	Adj. R-squared:	0.643
Method:	Least Squares	F-statistic:	26.23
Date:	Sun, 12 Nov 2023	Prob (F-statistic):	1.86e-09
Time:	14:47:29	Log-Likelihood:	-119.83
No. Observations:	43	AIC:	247.7
Df Residuals:	39	BIC:	254.7
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	23.9482	9.132	2.623	0.012	5.477	42.419
Col_Lang	-8.0457	1.636	-4.917	0.000	-11.355	-4.736
HDI	-3.3380	10.798	-0.309	0.759	-25.179	18.503
EU	-6.7145	1.369	-4.906	0.000	-9.483	-3.946

Omnibus:	1.906	Durbin-Watson:	1.870
Prob(Omnibus):	0.386	Jarque-Bera (JB):	1.010
Skew:	-0.305	Prob(JB):	0.603
Kurtosis:	3.438	Cond. No.	32.8

Figure 13: Regression 2, EU -> Wikipedia Score

Selected Languages	Internet Functionality
English	1.32
French	1.12
Swedish	1.1
Hungarian	.86
Slovak	.9
Macedonian	.85
Chechen	.92
Chuvash	.94
Sardinian	.88
Kabardian	.79

Figure 14: Selected Languages with Internet Functionality

EU Membership -> Wikipedia Score -> Content Production

EU membership did not cause high Internet Functionality directly. My theory held that EU membership caused a jump in Internet Functionality through the intermediary variable of Wikipedia Size, so I constructed two additional tests, measuring the impact of EU membership on Wikipedia size, and Wikipedia size on Internet Functionality.

In this second Regression Model shown in Figure 13, being in the EU was an excellent predictor of Wikipedia's success, even when controlling for other variables. Based on the correlation coefficient, we can interpret that Languages of the European Union had Wikipedias that were roughly six times larger than other European Languages. The impact of a decrease in six points of a language’s Wikipedia score can be illustrated in Figure 9. All 22 EU languages observed were in the top 26 of 43 European Wikipedias. This relationship makes sense, as the EU spends 1% of its budget translating documents annually. This corpus provides a base of reliable, in-language citations for Wikipedia on various subjects. The general translation noted

ability in these languages makes transferring knowledge easier, leading to large Wikipedias.

Following this, having a large Wikipedia leads to significant jumps in that language's general functionality on the internet. Figure 13 shows that EU languages have Wikipedias six times larger than non-EU languages within Europe. Based on the analysis found in Figures 12 and 13, a sixfold increase in Wikipedia size would be a jump of .06 in the model predicting content production ratio. To demonstrate the impact of a .06 increase in internet functionality, this represents the jump from Karbadian to Hungarian, Danish to French, or Chechen to Czech. There are notable differences in these pairs of languages in international prestige, use in research, support in products, and as a mode of instruction. Figure 14 shows relative differences in internet functionality among selected languages, further illustrating the impact of a .06 jump.

Conclusion

Summary of Results

The languages of the EU have large Wikipedias and languages with large Wikipedias are significantly more usable on the Internet. Work in the literature supported the thesis that a language having official status in the EU would lead to more functionality of that language on the internet. The EU creates a parallel internet corpus in the official languages, which powers neural network translation used since 2016. High-quality translations likely helped knowledge transfer among languages. Content creators in these languages would have a wealth of tools available, and the portion of the internet written in that language would rise. There is no direct relationship between

an official language in the EU and having high internet functionality. However, being an official language of the EU caused tremendous increases in those languages' Wikipedias. In turn, languages with large Wikipedia were more usable on the internet. EU policy shapes the present and future languages of the internet.

Importance to Minority Language Advocates

The relationship between Wikipedia and the utility of a language is essential to the Minority speakers advocating for greater autonomy. In 2001, Catalan Nationalists authored the 1st non-English Wikipedia (Wikipedia:Multilingual Monthly Statistics 2001). Today, Catalan has achieved historic levels of autonomy in Spain, and many Catalonians now use their language on the internet, a triumphant story of 21st-century language revitalization (Novella 2010). The Cebuano language is spoken natively by 34% of Filipinos, a group that has been fighting for the ability to use their language in their homeland. After English, the Cebuano Wikipedia is the second-largest, powered by the work of machine translating bots. (News 2020) When a Telugu Wikipedia editor announced that he had written 365 articles in 365 days (ఆఫీసుకి 2017), he was celebrated as a hero in the movement for Telugu autonomy in South India and later congratulated by the Indian VP (Jagdeep 2017). He has now written 1000 articles in 1000 days. All three autonomous movements are linguistic minorities fighting to make their language more used online. When languages cede utility in the global sphere, they also lose many of their most talented members, who go on to study and work in dominant language spheres in a process called brain drain (Certo 2014). Using

Wikipedia to boost a language's internet presence can fight brain drain and support minority languages (Lee 2010).

Importance to Political Science

This study shows that it is possible to institute large-scale knowledge diffusion from highly prestige languages to national languages, with the help of neural machine translators and the development of a corpus on the internet. For language policymakers, this shows actionable steps for how a language can rise in prestige. Raising your language’s status contributes to more education being used in the medium of the language, which can help fight brain drain. For nations whose languages are underrepresented on the internet, language policy is a component of economic policy (Nnaemeka et al 2023).

Limitations to other regions

The exclusive selection of languages from Europe with over a million speakers limits the external validity of this research. The most underrepresented online languages come from linguistically diverse places such as Nigeria, the Philippines, and India (Pimienta 2023). European languages have undergone extensive standardization within speech and orthography (Staiiger 2019). Regional dialects have been decreasing as transportation improves (Auer 2017). Nationalist movements have suppressed minority languages historically to support lingua francas (Barbour 2000). Outside of Europe, languages tend to be more regional, less official, and more oral (Wardhough 2006).

Future Steps

Future researchers may want to further increase internal validity by accounting for dialects and regional languages in a more standardized way. The regionalized approach worked well to isolate linguistic differences from economic differences in the causes of the digital divide. Keeping a regional lens can be exploited to remove confounding variables. This project has demonstrated a possible future for a more inclusive, truly global web. Next, I will work on deploying neural machine translation techniques in under-resourced languages, to replicate the large-scale knowledge diffusion found within the EU.

Bibliography

"Amnesty International – International Roma Day 2011: Stories, Background Information and Video Material." *Amnesty International*, https://www.amnesty.org/en/documents/eur01/005/2011/en/. Accessed 15 Nov. 2023.

APC. *Internet Rights - APC*. 17 Dec. 2008, https://web.archive.org/web/20081217021644/http://rights.apc.org/charter.shtml.

Auer, Peter. "Dialect Change in Europe—Leveling and Convergence." *The Handbook of Dialectology*, edited by Charles Boberg et al., 1st ed., Wiley, 2017, pp. 159–76. *DOI.org (Crossref)*, https://doi.org/10.1002/9781118827628.ch9.

Barbour, Stephen, and Cathie Carmichael. *Language and Nationalism in Europe*. OUP Oxford, 2000.

Bodle, Robert. "IRPC Charter." *Internet Rights and Principles Coalition*, 1 Sept. 2011, https://internetrightsandprinciples.org/charter/.

Brás, Andrea. "Bridging the Linguistic Divide: The Impact of Language Rights on Internet Freedom." *Localization Lab*, 8 Mar. 2019, https://www.localizationlab.org/blog/2019/3/7/hl8xdh6nacw5bpe5v4skhjkv0smeda.

Buck, Carl Darling. "Language and the Sentiment of Nationality." *American Political Science Review*, vol. 10, no. 1, Feb. 1916, pp. 44–69. *Cambridge University Press*, https://doi.org/10.2307/1946302.

Certo, Peter. "Brain Drain and the Politics of Immigration - FPIF." *Foreign Policy In Focus*, 25 Feb. 2014, https://fpif.org/brain-drain-politics-immigration/.

CIA. *Literacy - The World Factbook*. https://www.cia.gov/the-world-factbook/field/literacy/. Accessed 12 Sept. 2023.

Crystal, David. The Cambridge Encyclopedia of Language. Cambridge : Cambridge University Press, 1987. Internet Archive, http://archive.org/details/cambridgeencycl000crys.

Digital Economy and Society Statistics - Households and Individuals. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Digital_economy_and_society_statistics_-_households_and_individuals. Accessed 15 Nov. 2023.

Dorssemont, F., et al., editors. *The Charter of Fundamental Rights of the European Union and the Employment Relation*. Hart Publishing, 2019.

"EUR-Lex Corpus." *Sketch Engine*, 2 June 2016, https://www.sketchengine.eu/eur-lex-parallel-corpus/.

European Commission, editor. *Promoting Language Learning and Linguistic Diversity: An Action Plan 2004-06*. Office for Official Publications of the European Communities, 2004.

EuroStat. "Digital Economy and Society Statistics - Households and Individuals." *Europa*, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Digital_economy_and_society_statistics_-_households_and_individuals. Accessed 12 Nov. 2023.

---. *What Languages Are Studied the Most in the EU?* https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20220923-1. Accessed 29 Aug. 2023.

Ginsburgh, Victor, and Juan D. Moreno‐Ternero. "Brexit and Multilingualism in the European Union." *Metroeconomica*, vol. 73, no. 2, May 2022, pp. 708–31. *DOI.org (Crossref)*, https://doi.org/10.1111/meca.12379.

GlobeScan. *Internet Access "a Human Right."* 8 Mar. 2010. *news.bbc.co.uk*, http://news.bbc.co.uk/2/hi/8548190.stm.

Graham, Mark. "Time Machines and Virtual Portals: The Spatialities of the Digital Divide." *Progress in Development Studies*, vol. 11, no. 3, July 2011, pp. 211–27. *DOI.org (Crossref)*, https://doi.org/10.1177/146499341001100303.

Grin, François. *Language Policy Evaluation and the European Charter for Regional or Minority Languages*. Palgrave Macmillan UK, 2003. *DOI.org (Crossref)*, https://doi.org/10.1057/9780230502666.

Guillen, M. F., and S. L. Suarez. "Explaining the Global Digital Divide: Economic, Political and Sociological Drivers of Cross-National Internet Use." *Social Forces*, vol. 84, no. 2, Dec. 2005, pp. 681–708. *DOI.org (Crossref)*, https://doi.org/10.1353/sof.2006.0015.

Hilbert, Martin. “When Is Cheap, Cheap Enough to Bridge the Digital Divide? Modeling Income Related Structural Challenges of Technology Diffusion in Latin America.” *World Development*, vol. 38, no. 5, May 2010, pp. 756–70. *DOI.org (Crossref)*, https://doi.org/10.1016/j.worlddev.2009.11.019.

Jagdeep. “Https://Twitter.Com/VPIndia/Status/905985266011127808.” *X (Formerly Twitter)*, 2017, https://twitter.com/VPIndia/status/905985266011127808.

Keating, Dave. “Despite Brexit, English Remains The EU’s Most Spoken Language By Far.” *Forbes*, https://www.forbes.com/sites/davekeating/2020/02/06/despite-brexit-english-remains-the-eus-most-spoken-language-by-far/. Accessed 10 Sept. 2023.

Kornai, András. “Digital Language Death.” *PLoS ONE*, edited by Eduardo G. Altmann, vol. 8, no. 10, Oct. 2013, p. e77056. *DOI.org (Crossref)*, https://doi.org/10.1371/journal.pone.0077056.

Lee, Jenny J., and Dongbin Kim. “Brain Gain or Brain Circulation? U.S. Doctoral Recipients Returning to South Korea.” Higher Education, vol. 59, no. 5, May 2010, pp. 627–43. Springer Link, https://doi.org/10.1007/s10734-009-9270-5.

Luxton, Emma. “4 Billion People Still Don’t Have Internet Access. Here’s How to Connect Them.” *World Economic Forum*, 11 May 2016, https://www.weforum.org/agenda/2016/05/4-billion-people-still-don-t-have-internet-access-heres-how-to-connect-them/.

Montoya, Juan David. *Quaternary Sector: Definition, Background, Examples - Economic Activity*. 6 July 2017, https://www.economicactivity.org/quaternary-sector/.

News, G. M. A. “Cebuano Wikipedia, the World’s Second-Largest Wiki Edition, Is Almost Entirely Written by a Bot—Report.” *GMA News Online*, 28 Feb. 2020, https://www.gmanetwork.com/news/scitech/technology/727785/cebuano-wikipedia-the-world-s-second-largest-wiki-edition-is-almost-entirely-written-by-a-bot-report/story/.

Nnaemeka, Ogudu & Adeyinka, Busayo & Jimoh, Nurudeen & Bulus, Felicia & Chinaza, Anosike & Joy, Anosike & Chiedozie, Anosike & Onyemachi, Anosike & Blessing, Oladoye & Daniel, Nweke & Uchenna, Ukamba & Onwuha, Darlington & Solomon, Ogunleye & Tijani, Ahmad. (2023). Political Economy of Language: Linguistic Perspectives on Economic Policy. International Journal of Advanced Multidisciplinary Research. Volume 3. 799-803.

Novella, Antonio, and Jordi Aguilera. *El Català, Un Exemple d’èxit a Internet - Espai Internet*. 8 Mar. 2010, https://web.archive.org/web/20100308032029/http://blogs.tv3.cat/espaiinternet.php?itemid=24990.

Ogden, Jeff. *English: A World Map Colored to Show the Level of Internet Penetration (Number of Internet Users as a Percentage of a Country’s Population).* 24 Apr. 2012. Own work based on: figures from the Wikipedia:List of countries by number of Internet users article in the English Wikipedia, which is in turn based on figures from the International Telecommunications Union (ITU) for 2010 (updated to use figures for 2012 on 28 June 2013) (updated to 2016 on 5 Jan 2019) (updated to 2021/2022 on 17 June 2023). This SVG map includes elements that have been taken or adapted from this map: BlankMap-World6.svg., *Wikimedia Commons*, https://commons.wikimedia.org/wiki/File:InternetPenetrationWorldMap.svg.

Park, Sora. *Digital Capital*. Palgrave Macmillan, 2017.

Phillipson, Robert. “*Lingua Franca* or *Lingua Frankensteinia*? English in European Integration and Globalisation [1].” *World Englishes*, vol. 27, no. 2, May 2008, pp. 250–67. *DOI.org (Crossref)*, https://doi.org/10.1111/j.1467-971X.2008.00555.x.

Pimienta, Daniel. “Indicators of Languages in the Internet.” *Observatory of Languages and Culture in the Internet*, Dec. 2019.

---. “The Method behind the Unprecedented Production of Indicators of the Presence of Languages in the Internet.” *Frontiers in Research Metrics and Analytics*, vol. 8, May 2023, p. 1149347. *DOI.org (Crossref)*, https://doi.org/10.3389/frma.2023.1149347.

Ragnedda, Massimo, and Glenn W. Muschert, editors. *The Digital Divide: The Internet and Social Inequality in International Perspective*. 0 ed., Routledge, 2013. *DOI.org (Crossref)*, https://doi.org/10.4324/9780203069769.

Redman, Kate. “40% Don’t Access Education in a Language They Understand | UNESCO.” *Unesco*, https://www.unesco.org/en/articles/40-dont-access-education-language-they-understand. Accessed 9 Sept. 2023.

Reilly, Colleen A. “Teaching Wikipedia as a Mirrored Technology.” *First Monday*, Dec. 2010. *DOI.org (Crossref)*, https://doi.org/10.5210/fm.v16i1.2824.
Renan, Ernest. *What Is a Nation? And Other Political Writings*. Columbia University Press, 2018.
Richter, Felix. “Infographic: English Is the Internet’s Universal Language.” *Statista Daily Data*, 21 Feb. 2022, https://www.statista.com/chart/26884/languages-on-the-internet.
Ritchie, Hannah, et al. “Energy.” *Our World in Data*, Oct. 2022. *ourworldindata.org*, https://ourworldindata.org/energy-access.
Straaijer, Robin. “Language Standardization.” *Obo*, 2019, https://www.oxfordbibliographies.com/display/document/obo-9780199772810/obo-9780199772810-0250.xml.
Taylor, Petroc. “3G Mobile Network Coverage by Region 2022.” *Statista*, https://www.statista.com/statistics/1228787/3g-mobile-network-coverage-worldwide-by-region/. Accessed 10 Sept. 2023.
Thacker, Scott, et al. “Infrastructure for Sustainable Development.” *Nature Sustainability*, vol. 2, no. 4, Apr. 2019, pp. 324–31. *DOI.org (Crossref)*, https://doi.org/10.1038/s41893-019-0256-8.
Tollefson, James W. *Planning Language, Planning Inequality: Language Policy in the Community*. Longman, 1991.
UNDP, editor. *Uncertain Times, Unsettled Lives: Shaping Our Future in a Transforming World*. United Nations Development Programme, 2022.
United Nations. “Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities.” *OHCHR*, https://www.ohchr.org/en/instruments-mechanisms/instruments/declaration-rights-persons-belonging-national-or-ethnic. Accessed 9 Sept. 2023.
---. “Universal Declaration of Human Rights.” *United Nations*, 1948, https://www.un.org/en/about-us/universal-declaration-of-human-rights.
UNSDG. *Speaking Your Language: Paving the Way for More Harmonious Communication through Multilingualism*. https://unsdg.un.org/latest/announcements/speaking-your-language-paving-way-more-harmonious-communication-through. Accessed 9 Sept. 2023.
version, Industrius at English Wikipedia Later. *English: This Map Shows the World’s Language Families.* 14 Feb. 2005. Image:BlankMap-World.png by User:Vardion, *Wikimedia Commons*, https://commons.wikimedia.org/wiki/File:Human_Language_Families_Map.PNG.
Vincent, James. “UN Condemns Internet Access Disruption as a Human Rights Violation.” *The Verge*, 4 July 2016, https://www.theverge.com/2016/7/4/12092740/un-resolution-condemns-disrupting-internet-access.
Vogels, Emily A. “Digital Divide Persists Even as Americans with Lower Incomes Make Gains in Tech Adoption.” *Pew Research Center*, https://www.pewresearch.org/short-reads/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/. Accessed 15 Nov. 2023.
w3techs. *Usage Statistics and Market Share of Content Languages for Websites, August 2023*. https://w3techs.com/technologies/overview/content_language. Accessed 29 Aug. 2023.
Wardhaugh, Ronald. *An Introduction to Sociolinguistics*. Malden, Mass., USA : Blackwell Pub., 2006. *Internet Archive*, http://archive.org/details/introductiontoso00ward.
Weber, George. *Top Languages*. 7 May 2013, https://web.archive.org/web/20130507110651/http://www.andaman.org/BOOK/reprints/weber/rep-weber.htm.
Wiggers, Heiko. “Digital Divide: Low German and Other Minority Languages.” *Advances in Language and Literary Studies*, vol. 8, no. 2, Apr. 2017, p. 130. *DOI.org (Crossref)*, https://doi.org/10.7575/aiac.alls.v.8n.2p.130.
“Wikipedia:Multilingual Monthly Statistics 2001.” *Wikipedia Foundation*, 29 Dec. 2022. *Wikipedia Foundation*, https://en.wikipedia.org/w/index.php?title=Wikipedia:Multilingual_monthly_statistics/2001&oldid=1130341658.
Wilson, Jenny. “United Nations Report Declares Internet Access a Human Right.” *Time*, 7 June 2011. *techland.time.com*,

https://techland.time.com/2011/06/07/united-nations-report-declares-internet-access-a-human-right/.

World Data Lab. *World Poverty Clock*. https://worldpoverty.io. Accessed 12 Sept. 2023.

Wright, Scott. “*Digital Citizenship: The Internet, Society, and Participation* , by Karen Mossberger, Caroline J. Tolbert, and Ramona S. McNeal: Cambridge, MA: MIT Press, 2008, 221 Pages.” *Journal of Information Technology & Politics*, vol. 5, no. 2, Aug. 2008, pp. 262–64. *DOI.org (Crossref)*, https://doi.org/10.1080/19331680802290972.

Young, Holly. *The Digital Language Divide*. http://labs.theguardian.com/digital-language-divide/. Accessed 9 Sept. 2023.

Your Europe. “Using and Accessing the Internet in the EU.” *Your Europe*, https://europa.eu/youreurope/citizens/consumers/internet-telecoms/internet-access/index_en.htm. Accessed 10 Sept. 2023.

ఆఫీసుకి హాయిగా వెళ్లిరండి! 15 Oct. 2017, https://web.archive.org/web/20171015051625/http://www.eenadu.net/magazines/sunday-magazine/sunday-magazineinner.aspx?catfullstory=16749.