

Digital Language Divide from the Sociolinguistic Perspective: Is There a Chance to Bridge the Gap?

Dubravka Papa

In multilingual environments, the “digital language divide” as the primary barrier preventing the speakers of small languages, the educationally deprived, the elderly, and the disabled, from their full integration into the emerging digital societies is a highly relevant as the grounds for digital inequality for billions of people worldwide. The paper aims to define and critically analyse the digital language divide focusing on how digital language technologies contribute to discrimination and inequality. The research is based on the analysis of the selection of studies and language policy reports related to digital inequality, linguistic diversity and digital language technology mainly from the sociolinguistic perspective. The analysis shows that small languages are excluded from access to information, emerging social inequalities and a growing need for inclusive language policies in further development of the communication based on AI and digital technologies.

Keywords: digital language divide, linguistic equality, multilingualism.

1. Introduction

In the modern globalised world, multilingualism has become one of the key features and a prerequisite for successful communication in societies in which coexistence between indigenous speakers and speakers of minority languages or immigrant populations is necessary. Research on multilingualism began with the founders of the concept itself, namely Uriel Weinreich and Einar Haugen, who published their key works in the 1950s, laying the foundations for later sociolinguistic research, language planning and multilingualism. In the sociolinguistic approach of the 20th century, the cornerstones of research in sociolinguistics were the works of Joshua Fishman on the study of minority languages, the concept of language loyalty and the language preservation policy, as well as the works by John Gumperz, who was one of the pioneers of the analysis of speech interaction in multilingual contexts and one of the founders of sociolinguistics, which deals with issues of language in the environment. Critical and contemporary research in sociolinguistics from the late 20th and in the early 21st centuries addresses issues of language planning and concepts such as translanguaging and bilingualism. There, Suzanne Romaine addresses the qualitative nature of multilingualism, cognitive and social aspects of bilingualism and multilingualism.

Multilingual societies not only share the same physical, communicative space, but since the 21st century, they share a broad digital space. Therefore, multilingualism represents a wealth and advantage in communication, while simultaneously representing a barrier to access to information, labour market and social inclusion. This is particularly emphasised in the modern digital era, when speech communities also face a digital divide due to exposure to digital technologies as a source of potential inequality and discrimination. In this context, multilingual communities face a digital divide, which refers not only to the inability to access the Internet and digital technologies but also to

more complex forms of exclusion including linguistic and cultural exclusion. The linguistic divide in the digital space is based on the uneven availability of content, tools and information in different languages.

Since European Union is a multilingual community, in which 24 languages enjoy the same official status, and can be used on equal basis in the EU institutions, there is a constant need to strike the balance between the language load, multilingual policy and institutional efficiency. Additionally, there are more than 60 regional and minority languages in the EU contributing to its cultural and linguistic abundance but on the other hand requiring appropriate, well-established language policies, planning and management. Moreover, there is a significant imbalance in technological language support, since English is better supported in this regard than other EU languages. The result of this imbalance is a limited access to information, digital services and reduced social-economic activities involvement for the speakers of small languages.

The paper researches into the questions how digital technologies contribute to aggravating or mitigating the digital language gap among speakers of small languages, how language policies at national and international levels affect digital inclusion of speakers of small languages and what the key consequences of the exclusion of small languages from the application of digital communication are. As regards the paper structure, after stating the aims and methodology of the research, an overview of the theoretical framework for the topic is given. Next, the relevant literature related to language vitality, digital and language divide is analysed. The focus is also on the analysis of two case studies and finally, the concluding section of the paper summarizes the analysis results, reflections, implications and gives recommendations for further research.

2. Aims

The aim of the paper is to research the influence of language on digital access and equity based on the analysis of literature sources addressing the digital divide and linguistic diversity. In this context the paper will define the term “digital language divide” from sociolinguistic, economic, and legal perspectives, define the term “language vitality”, analyse and assess the sources on the condition and status of speakers of small languages. It will also focus on the phenomenon of digital language inequality caused by digital technologies as regards small languages. In conclusion, two case studies offering possible recommendations for language policies and further research into the topic are presented.

3. Methodology

The methodological framework of the paper is based on a qualitative content analysis of scholarly works in the fields of sociolinguistics, digital communication, language policy and information technology. The sources include academic articles, reports issued by international organisations (such as UNESCO and ITU), legislation and current case studies that illustrate the impact of the digital divide on speakers of small languages. The paper examines initiatives and attempts to bridge language barriers in the digital space, offering insights into possible directions for the future development of digital inclusion and language equality. Although there are other terms used for a language with small

number of speakers (minority, lesser-used, underrepresented, endangered, indigenous, regional, local etc.) the term “a small language” has been used in this paper as explained in Papa and Omazić (2021). The concept of “smallness” is deemed relative and multifaceted, and it is not limited solely to quantitative indicators, such as the number of speakers. Although the concept of “a small language” is usually equated with “a minority language” in sociolinguistic literature, in this context the term has taken a broader meaning, going beyond the dichotomy of minority and majority languages. Thus, “small languages” are carriers of authentic culture and adaptability, which gives them a strategic prestige in a global society.

4. Theoretical Framework and Definitions

This section provides a review of the recent literature and explains the key concepts forming the theoretical basis of the paper. Special emphasis is placed on the role of multilingualism in the society and the factors contributing to language variation and change. It discusses language policy relevant to digital language divide phenomenon, deals with key issues (language vitality, language endangerment and digital vitality) related to language division and language barriers in the digital age. The concept of digital language divide and language inequality is referred to from a sociolinguistic perspective analysing the causes and providing case studies of the digital language divide.

4.1. *Development of the EU Project on Language Equality in the Digital Age*

One major initiative focused on promoting language equality through digital means was the project European Language Equality (ELE), which was funded by the European Union. In 2018, the European Parliament approved a report titled Language Equality in the Digital Age, which included over 40 recommendations from two key committees: the Industry, Research and Energy (ITRE) Committee and the Culture and Education (CULT) Committee. These recommendations emphasized strengthening the institutional framework for language technologies (LT) within the EU, advancing research and education strategies, and ensuring LT benefits are accessible across both public and private sectors. This led to the recognition of the need for a sustained, coordinated research and development programme in LT. The project’s ultimate goal is to achieve complete digital language equality (DLE) across Europe by 2030.

The project deliverables were strategies and recommendations (SRIA) for digital language equality in Europe addressing the problems of digital inequality of the European languages within the framework of the goal of Deep Natural Language Understanding by 2030. It relied on the previous incentives such as META-NET intended to include a wide spectrum of actors i.e. the academic community, industry, linguistic communities and political institutions. It refers to equal technological support for European languages in the context of “the development and application of artificial intelligence (AI), natural language processing (NLP), language understanding (NLU) and language and speech technologies (LTs and STs)” (Giagkou et al. 2023, p. 2). ”

AI and related technologies have the power to enable multilingualism in the digital environment, e.g. through translation, automatic recognition and generation of speech and text. However, the 2012 META-NET study revealed a significant imbalance in technological support across European languages. While English benefits from a well-developed ecosystem of tools, corpora, and data, other languages, in particular small languages, such as Icelandic or Maltese receive minimal support, placing them at considerable risk of digital extinction. The EU is striving to ensure equal rights for all languages, but in technological terms, this equality has not yet been achieved. Thus, Kornai (2013) further warns that many languages are in danger of not surviving in the digital age, because they lack basic resources for development.

4.2. On Language Vitality in View of Language Equality in the Digital Age

The issues of language endangerment and language vitality have been considered key issues both from the sociolinguistic perspective and from the perspective of human rights (Moseley 2010). Language vitality is not a new term and since 1991, significant efforts have been made to assess it. Some of the tools for language vitality assessment are listed in Papa and Omazić (2021); these are the GIDS scale developed in 1991 by Fishman, the 2003 UNESCO LVE tool, the Krauss's Language Vitality Assessment Framework designed in 2007, the 2010 EGIDS scale by Lewis and Simons, and the EuLaViBar compiled in 2013 within the ELDIA project by a group of researchers (Djerf, Spiliopoulou, Kührt, Toivanen, Sarhima and Laakso).

In recent years, new tools for assessing language vitality have emerged. Notably, the Language Endangerment Index (LEI) was introduced in 2014, and in 2015, research on the Linguistic Landscape (LL) was developed, emphasizing the visibility and representation of languages in public spaces among other factors. Despite the tools for determining the level of language endangerment, the language vitality depends on factors that these tools do not include, such as the actual attitudes of speakers towards the language they speak, power relations in a society, trends, cultural values, economic conditions, the preservation of the linguistic environment, or family relationships (Toivanen–Saarikivi 2016).

Language vitality assessing is multi-factor dependant since language communities are specific and complex. Language vitality is not determined solely by the number of its speakers; thus UNESCO 2003 language vitality assessment is based on nine factors that should never be observed alone since it is possible that to one criterion a language ranks well and due to other factors it still might need urgent revitalisation.

4.3. Language Vitality vs. Digital Language Vitality

Language vitality is closely linked to digital vitality due to the importance of digital environment today in the preservation, use and transmission of languages. Simons et al. (2022) state that speakers of endangered languages face serious challenges in the digital age because access to digital tools and technologies is highly unequal across languages. This inequality is called the digital language divide as a term that refers to the differences in the availability of digital language support (DLS) across languages. While some languages have a rich digital system (keyboards, software, machine translation,

voice assistants), others have almost no support at all. As digital ways of communicating and accessing information become increasingly important to everyday life, speakers of languages without DLS are forced to use dominant languages to be able to participate in digital society or otherwise they risk exclusion. While linguists have recognized the potential of digital technology for language revitalisation for decades, and language technology experts are actively working to reduce inequality, it is unrealistic to expect full digital support for every language. In multilingual communities, languages often have specific functional roles, meaning that speakers do not necessarily seek a universal digital presence for all the languages they use.

According to Kornai (2013), several areas have to be taken into consideration when speaking of the interdependence of specific language functions and language vitality. Firstly, the language use in digital media is a significant indicator of its vitality. If a language is actively used in digital media (on the internet, in applications, social networks, translation tools and software) this increases its relevance and presence, especially among younger generations. Secondly, the interdependence is structured by tools and resources accessible in specific languages. Languages that have accessible digital tools (e.g. keyboards, spell checkers, speech recognition systems) have a better chance of long-term survival because they enable easy everyday language usage. Thirdly, digital vitality enables the creation and sharing of educational and cultural content, i.e. knowledge transfer that strengthens the intergenerational transmission of the language. Additionally, prestige and visibility play a significant role. Languages present in the digital space are perceived as more “thriving” and relevant, which influences speakers’ attitudes about their value and future. Finally, standardization and documentation are processes that digital tools facilitate. Standardizing and archiving languages is particularly important for minority and endangered languages.

Language vitality is increasingly measured today by the degree of digital integration and accessibility of the language. Simons et al. (2022), analyse the extent to which the world’s languages are represented and supported in the digital environment. The results of their quantitative assessment of digital language support using indicators such as language presence on the Internet, support in operating systems, availability of digital tools, language technologies, and language use on social networks show which languages are digitally well-supported, which are endangered emphasizing those that urgently need digital revitalization. Languages are categorized according to their level of digital vitality, from those with full support to those completely neglected in the digital space. The authors emphasize that the preservation of linguistic diversity in the digital age is possible only through the collaboration of linguists, computer scientists, and policymakers, with the goal of enabling all languages to have a digital presence and functionality.

Although Kornai (2013, p. 6) notes that the level of vitality set by EGIDS is the best “predictor of a language’s digital status”, Simons et al. (2022, p. 4300) emphasize that digital vitality is not the same as social or linguistic vitality. They state that methodologically, the digital vitality of a language should be independent “(orthogonal) of its non-digital vitality”. This means that when assessing digital language support (DLS), one should not automatically take into account the number of speakers or linguistic vitality in the traditional sense (e.g. in everyday language usage or intergenerational transmission). They support this statement by an illustration of Latin

which, although a “dead” language (with no native speakers), according to their methodology ranks as the 80th most digitally vital language, because of its strong presence in digital resources. Conversely, Aimaq, a language with almost two million speakers in Afghanistan, shows very low digital vitality, because it has almost no digital infrastructure. Therefore, it can be concluded that the number of speakers and everyday use do not suffice as indicators for understanding the digital status of a language and a separate analysis of digital factors is required.

5. The Economic and Legal Aspects of Digital (Language) Divide

A digital divide can be observed interdisciplinary from sociolinguistic, legal and economic perspectives. Gomes (2025, p. 150) according to Haight et al. (2014) states that in the 1990s, the term digital divide referred mainly to the difference in rates of internet access. Nowadays, the meaning extended to quality of Internet connections and equipment, user knowledge, and social support.

Broadening of meaning in terminology is usually facilitated by broadening of the fields of language functions. In this context, Van Dijk (2020, p. 6) believes that the digital divide is an issue of the “relationship between information and communication technology (ICT) and social inclusion”. It is marked not only by “physical access to computers and connectivity but also by access to the additional resources that allow people to use technology well.” (*ibid*). However, the digital divide extends beyond mere access to hardware and internet connectivity; it also encompasses challenges related to digital content, language accessibility, education, literacy levels, and the availability of community and social support resources. Van Dijk (2020) also notes that the digital divide is a complex social concept that not only refers to access to technology, but comprises a number of other meanings. Van Dijk (2020) redefines the concept of the digital divide, emphasizing that access to technology is not just a matter of device ownership, but involves a range of interrelated factors.

Rather than focusing solely on a binary divide between those who have and those who do not have access to technology, Warschauer (2003) proposes a more complex framework that encompasses different dimensions of access to information and communication technologies (ICT). Three common models of ICT access according to Warschauer include devices, connectivity, and literacy. The first model refers to physical access to technology, such as owning a computer, smartphone, or other digital device. However, simply owning a device is not enough to use technology effectively. Connectivity includes access to the Internet or other network infrastructures that allow devices to connect to the global information network. Without adequate connectivity, even the most advanced devices lose their functionality. However, literacy encompasses the ability of users to use technology, including digital skills, critical thinking, and the ability to adapt to new technologies. Without adequate literacy, access to devices and the internet does not result in true inclusion in the digital society.

Warschauer (2003) highlights the importance of other resources that influence digital inclusion. Physical resources related to availability and quality of hardware and software, digital resources comprising availability of relevant and localized digital

content, human resources involved in education, training and support for users in acquiring digital skills and finally, social resources referring to the role of communities and institutions in supporting digital inclusion, including schools, libraries and NGOs.

In defining the causes of digital divide, Warschauer (2003, p. 31) notes that they may lie within the socioeconomic status (income, education), age (elderly people often excluded), geographic location (rural vs. urban), language (minority and non-digitalized languages), and disability (technical inaccessibility). Globally, developed countries have better access to technology and the internet whereas underdeveloped countries, the speakers of “digitally neglected” languages, and people with disabilities are often excluded from the digital society. Some typical cases of digital divide in practice are inability of children without internet during the COVID-19 pandemic to follow online classes or speakers of languages without digital tools (keyboards, translation tools) lacking access to basic information, or inability of senior citizens to use e-services (e.g. e-Citizens, e-banking).

There are several factors, that affect digital divide according to Hernandez (2023) and among these are language barriers, age, geography, disability, discrimination and other factors. According to the 2016 report on language equality in the digital age by the Scientific Foresight Unit of the European Parliamentary Research Service (Science and Technology Options Assessment – STOA), the digital divide encompasses concepts such as digital literacy and the competences required to effectively use technology. Digital exclusion arises as a direct outcome of this divide, while digital equality represents the intended objective, achieved through targeted policies, educational initiatives, and the development of digital infrastructure.

From the economic perspective, language barriers in the European Union create multiple negative consequences especially in the context of the free movement of people, goods, services and information within the Digital Single Market (DSM). The language gap within Europe is growing, especially between speakers of dominant languages and speakers of other languages. Labour mobility is limited and partly due to language barriers, only 5.8% of EU citizens have moved to another member state for work. Access to cross-border public services is difficult, as public administrations rarely offer information in migrants’ languages. Citizens’ political participation is weakened if they cannot communicate in their own language, which further encourages political inactivity and a sense of exclusion. The e-commerce in the EU is divided into 6 language groups, and countries with lower language barriers have four times the number of cross-border online shoppers.

Legal perspective of digital (language) divide derives from legislation or case law governing digital divide. The European Parliament resolution of 11 September 2018 on language equality in the digital age (2018/2028(INI)) is a political document that highlights the importance of preserving and promoting linguistic diversity within the European Union, particularly in the context of the digital transformation of society. The resolution recognises significant impact of digital technology on linguistic equality, in particular for lesser-used, regional and minority languages. In this context, the European Parliament expresses concern about the lack of appropriate policies to prevent the widening technological gap between languages with adequate resources and those without them. It stresses the need for developing language

technologies in order to facilitate communication for people with disabilities, including deaf and hard-of-hearing people. The report also advocates for robust legal safeguards for regional and minority languages, along with the acknowledgment of the collective rights of national and linguistic minorities within the digital domain. Moreover, it recommends the establishment of a dedicated centre for linguistic diversity to raise public awareness of the value of lesser-used languages. It further urges member states to formulate inclusive language policies and to allocate sufficient resources to promote multilingualism and language diversity in digital contexts. The resolution addresses the European Union institutions, academia and research centres, member states, businesses (especially SMEs) and other relevant stakeholders. The aim is to encourage cooperation and coordination between these entities in order to ensure linguistic equality in the digital age. As a resolution of the European Parliament, this document does not have legally binding force. However, it represents a political statement and recommendation that can influence the shaping of policies and legislation within the European Union.

While the European Parliament Resolution of 11 September 2018 on linguistic equality in the digital age [2018/2028(INI)] is not legally binding, it draws on and builds on several key legally binding instruments of the European Union documents relating to linguistic equality and the protection of linguistic diversity. Among these is the Charter of Fundamental Rights of the European Union (2000), specifically Article 22, which mandates the Union to cultural, religious and linguistic diversity respect. This Charter acquired binding legal status following the implementation of the Treaty of Lisbon in 2009. Additionally, the Treaty on European Union (TEU) reinforces this commitment; the Article 3(3) stipulates that the Union shall respect its rich cultural and linguistic diversity and ensure the preservation and promotion of Europe's cultural heritage. Lastly, Regulation No 1/1958 determining the use of languages within the European Economic Community defines the official and working languages of the EU institutions, ensuring that all official languages have equal status in legislation and communication within the Union. Furthermore, it should be noted that the European Charter for Regional or Minority Languages (Council of Europe, 1992) is not an EU instrument but EU member states have ratified it. It obliges the signatory states to the protection and promotion of regional and minority languages as integral components of Europe's cultural heritage.

After its adoption, Resolution 2018/2028(INI) served as a political impetus to raise awareness of the need to develop language technologies for small languages, encourage research and funding of projects aimed at the digital inclusion of linguistic minorities, and strengthen cooperation between member states in promoting linguistic diversity in the digital environment. The Resolution is the official position or recommendation of the European Parliament on a specific issue e.g. linguistic equality, human rights, the environment, foreign policy, etc.

5.1. Digital Language Divide

Gaspari et al. (2021, p. 4) provide in ELE project-WP1: European Language Equality – Status Quo in 2020/2021 the definition of digital language divide as “the state of affairs

in which all languages have the technological support and situational context necessary for them to continue to exist and to prosper as living languages in the digital age."

Bella et al. (2023, p. 1) define the notion of digital language divide as "referring to the gap between languages with and without a considerable representation on the Web and within the worldwide digital infrastructure." The authors tackle the widening disparity in digital support across the world's languages by emphasising the structural linguistic bias in the development of language technologies. The authors argue that dominant models of natural language processing grounded in artificial intelligence and machine learning, systematically favour a few globally dominant languages, such as English, while the majority of the world's languages remain neglected. This bias is not only technical, but also socio-cultural and epistemological, as it contributes to the exclusion of language communities from the digital sphere of knowledge, communication and identity. Digital linguistic bias results in epistemological injustice, a process in which the knowledge and linguistic perspectives of speakers of small languages are marginalized or completely excluded from the digital world. This undermines equal access to information and the possibility of expressing linguistic and cultural identities within a digital society.

In response to these challenges, Bella et al. (2023, p. 11) introduce the LiveLanguage initiative, an interdisciplinary project that aims to develop language resources and technologies in collaboration with local communities. The project is based on ethical principles, community involvement and cultural sensitivity, and proposes a methodology that not only recognizes linguistic diversity, but also actively integrates it into the development of language tools. It aims to create a more equitable digital system that does not support homogenisation, but rather encourages multilingualism and inclusivity. In conclusion, a profound change is called upon in the way language technologies are developed and distributed, emphasizing that access to language in the digital age is not only a technical issue, but also a political, social and ethical one. It is necessary to establish inclusive models that actively work to bridge the digital language divide through openness, collaboration and technological innovation.

The success of DSM and the building of European integration depends on reducing language barriers through linguistic inclusion in public services, multilingual digital communication, and support for minority languages, technical and political measures that enable the availability of content in citizens' mother tongues. Language barriers are not only a technical challenge, but also a strategic problem that affects unity, mobility, the market and democracy in the EU. Without an active multilingualism policy, the goal of creating an inclusive, integrated and competitive European area remains limited.

5.2. Digital Status of Small and Underrepresented Languages

Bella et al. (2023, p. 4) suggest that the language we speak influences our digital experience and language barriers shape access to information, our participation in digital communities, and opportunities for our online expression. A language can be a tool or a barrier to access information. Although the Internet is often presented as a universal platform for accessing information, the availability of content varies

significantly depending on the language. For example, English-speaking users have access to more search results and resources on platforms compared to speakers of less represented languages. This imbalance limits access to knowledge for many language communities.

Among the effects of language on digital participation is that spoken language significantly shapes our online experiences, influencing access to information, participation in digital communities, and opportunities to express opinions. Achieving digital equality requires recognizing and addressing these language barriers through the development of inclusive technologies and policies that support linguistic diversity. Users often communicate within language groups, which can limit cross-cultural interaction. The lack of content and tools in certain languages can discourage users from actively participating in digital discourses, thereby reducing their visibility and influence online. English language dominance on the internet creates digital inequality and makes access to information and services difficult for speakers of other languages. For example, many online services and tools are not available in minority languages, limiting their use and usefulness for these communities.

Digital technologies are aimed at reducing language barriers and technological advances, such as machine translation and multilingual platforms, offer opportunities to reduce language barriers. However, the effectiveness of these tools is often limited for less resourced languages. Additional investment and development are needed to ensure equal digital participation for all language communities. The problem of inequality of information is shown in the unequal representation of languages on the Internet, with an emphasis on Google as the dominant search platform and its limited linguistic inclusivity.

Pimenta et al. (2009) point out a striking fact that a well-known search engine recognizes 30 European languages but only one African language. This statement serves as a warning about the systematic neglect of thousands of languages other than the globally dominant ones, which creates a deep digital linguistic inequality. Although Google states that it aims to increase the number of supported languages, it faces serious challenges, especially when it comes to languages that are not standardized, languages that exist only in spoken form, and languages with a small number of speakers and no digital resources. Despite an estimated 7,000 languages in the world, Google supports just 130 languages for search, representing only 2% of the world's languages.

Zook and Graham (2007) conducted research that illustrates the concrete consequences of language inequality on the internet. The authors analysed Google search results in the West Bank using three languages: Hebrew, Arabic, and English. Results show that Arabic in the Palestinian territories yields only 5% to 15% of the results compared to the same search in Hebrew. English yields four to five times more results than Arabic, even though Arabic is the first language of the majority of the population. These data indicate the technical and content dominance of certain languages, which directly affects access to knowledge, information and digital services for speakers of digitally “weak” languages. It can be concluded that although technology publicly declare goals of expanding language coverage, the practice shows a pronounced digital language gap. Languages of the Global South and indigenous

communities remain outside the digital sphere, thus limiting their right to information and expression in the digital world.

5.3. Language Technology and Role of Multilingual Users

Hale (2013) suggests that the issue of the digital language gap may be bridged through translation technology and the empowerment of multilingual users. The focus is on the role of technology companies and online platforms in enabling multilingual communication and access to content in multiple languages. Automatic translation technologies offer one of the potential solutions for bridging language barriers on the Internet. Despite currently being available for a limited number of languages, there are already significant examples of their application. In 2021, Microsoft launched Skype Translator, a tool for simultaneous speech translation during video calls; Facebook and Twitter have integrated Bing Translator to enable users to automatically translate posts and content. These technologies not only encourage inclusivity, but also open up new business opportunities by entering multilingual markets.

Hale (2013) researched the role of multilingual users. The research results show that Internet platforms should more actively use the potential of multilingual users. He states that current platforms often ignore content in other languages because sites like TripAdvisor and Google Play only display reviews in the language of the user interface, while the reviews and content in other languages are often hidden or deprioritized. Hale (ibid.) suggests as an example Wikipedia, which is technically adjusted to enable the search of a topic through several language editions at the same time; this would encourage the multi-directional exchange of knowledge. Hale's data reveal that only 11% of Twitter users and 15% of Wikipedia users use multiple languages on these platforms. However, these multilingual users are significantly more active since they post more tweets, edit and write content on Wikipedia more often.

Their role is particularly significant because they can decentralize information and spread local news and cultural content to other linguistic and geographic communities. By doing so, they potentially reduce the fragmentation of information along linguistic or national borders. Visualisations by the Global Language Network from the Massachusetts Institute of Technology (MIT) further confirm the role of multilingual users. The interactive grid shows which languages are most connected through user behaviour on platforms like Twitter and Wikipedia. On Twitter, users who speak Malay, Portuguese and Spanish often tweet in English at the same time. On Wikipedia, users from a variety of linguistic backgrounds contribute to content in English, with English serving as a bridge between communities. Translation technologies and multilingual users are key tools in bridging the digital language divide. However, to realise their role fully, online platforms must actively recognise and foster multilingualism, rather than remaining limited to the user interface language or dominant languages.

6. Case Study

6.1. Digital Language Vitality Case Study by Simons et al.

Kornai (2013) introduced the categorisation of language vitality in the digital environment, developing a framework to assess digital vitality. Using this model, he classified languages into four distinct levels: Digitally thriving, Vital, Heritage and Still. Simons et al. (2022, p. 4303) present a classification of the world's languages according to their digital vitality i.e. the level of their digital language support (DLS). In their study, Simons et al. (2022) divide languages into five categories: Thriving, Vital, Ascending, Emerging, and Still reflect the degree of their presence and functionality in the digital sphere, from those with full support to those that are practically digitally excluded. Table 1 (Simons et al. 2022) shows that most languages are digitally neglected since the largest number of languages (almost 4,000) belong to the Still category (e.g. Aimaq-on the upper range and Yurok-on the lower end of the range), which means that they lack even basic digital infrastructure. This confirms a serious imbalance in digital presence among the languages worldwide. Very few languages have developed digital support, namely only 33 languages (e.g. English-on the upper range, Hungarian-on the lower end range) are categorised as Thriving, which indicates their full digital integration, including support in software, presence in digital media, and language technologies (e.g. speech recognition). Giving two examples of languages per level (one from the upper and one from the lower end of the range) illustrates the internal diversity within each category. For example, the Greenlandic and Hunsrik languages are both in the category of Ascending languages, but with different levels of digital development.

Table 1. Number of languages per DLS level

Level	Languages	Examples
Thriving	33	English, Hungarian
Vital	95	Nepali, Tongan
Ascending	401	Greenlandic, Hunsrik
Emerging	3304	Dogri, Michif
Still	3996	Aimaq, Yurok

Source: Simons et al. (2022, p. 4300)

The numbers indicate that few languages are digitally vital and specific digital revitalisation actions are required. The total of Emerging and Still categories together encompasses over 7,000 languages, meaning that more than 90% of the world's languages have little or no digital support. In conclusion, this case study is not only an informative, but also an alarming evidence of the large digital divide between languages, and of the need for targeted intervention to support so-called "low-resource languages."

6.2. Digital Language Vitality Case Study – English Only

English is undoubtedly the most vital language in the digital sphere and digitally well supported. Historically, digital vitality of the English language and its distinct

advantage over other languages in the context of the development and availability of language technologies goes back to the early stages of the computer and Internet history. English was the primary language of science, technology and international communication, which resulted in the fact that the first digital resources and standards were designed specifically for the language. This early advantage created the foundation for the subsequent exponential growth of digital support for the language.

Moreover, Maynard et al. (2022) argue in their report that institutional and industrial investment in language technologies supported the development of language technologies for the English language. In addition, governments, universities and corporations (e.g. Google, Microsoft, OpenAI) continuously invest in building tools and resources, which ensures a leading position for the English language in the digital language infrastructure. English also possesses the most extensive digital language resources, encompassing large-scale text corpora, lexicographic databases, machine translation and speech recognition technologies. This rich infrastructure enables the digital functionality of the English language in diverse contexts from education to artificial intelligence. English is, also a reference language in the development of artificial intelligence i.e. training and evaluation of large language models (e.g. GPT, BERT).

Thus, English is not only a communication tool, but also a framework within which modern language processing technologies are developed and validated. This situation makes it difficult to adapt technologies to languages with less available data. Finally, the impact of the digital dominance of the English language on global language inequality is considerable and a new speaker framework, namely the mobile bilingual, emerges in contexts of sociolinguistics change (Rodríguez-Ordoñez et al. 2022).

Although such a position allows for wide availability of technologies to English speakers, it simultaneously marginalizes languages with limited digital support. There is a need for the development of more inclusive and multilingual language technologies to alleviate the digital language divide.

7. Conclusion

In elaborating on trends in European language policies, Kirchmeier (2020) concluded that language technology could be one of the means to prevent language from becoming extinct and preserve its vitality. However, some languages could become digitally extinct, as stated in the Meta-NET White Paper Series by Rehm and Uszkoreit (2012).

Although the 24 official languages of the EU have equal political status, they are not nearly equally supported technologically. Furthermore, small languages traditionally suffer from limited support to ensure their future use and survival in the digital age. In this regard, the European Language Equality (ELE) project was one of the projects aimed at developing a systematic and inclusive comprehensive strategic agenda for research, innovation, and implementation with the goal of achieving full digital linguistic equality through concrete guidelines and recommendations in Europe by 2030.

Globally, the world's languages do not have equal digital status, which may deteriorate even more, if additional efforts and resources are not invested in the development of language technologies. Every language requires a stable digital position, but this stability is not guaranteed in the long term without increased investment in research and development and clear language policies. Small languages that could fall behind in the digital sphere, require strong support both at national and European levels to finance long-term programmes equally motivating scientific and industrial development of language technologies for the respective small language.

The digital gap between resource-rich and resource-poor languages needs to be bridged by applying new methods and technologies that enable smaller languages to progress. Coordination between national and international initiatives is crucial for the successful development and sustainability of language technology. In conclusion, a strategic, well-funded and coordinated approach is required so that the small languages remain digitally vital and can compete with globally dominant languages in the digital era.

In summary, access to information and digital services is a prerequisite to linguistic equity in digital space in particular regarding small languages. Digital exclusion of small languages can cause their extinction and in this case digital technology can serve as a revitalisation tool for endangered or extinct languages but only under the condition that language technologies for these languages are developed and maintained. Moreover, digital neglect can have economic and educational consequences since it directly influences labour market, access to knowledge and services. Without digital language support, the speakers of small languages have only limited opportunities for education and professional improvement. The analysis results in this paper are indicative of the necessity for further research in the field of digital language support in digital democracy and e-governing. They also point to further research into the efficient patters of digital language revitalisation by means of AI, corpus linguistics and collaboration with speech communities. Economic outcomes of digital language divide and the role of language policies in digital platforms for education need to be further researched, as well.

As emphasized by Warschauer (2003), bridging the digital divide requires more than just providing technology. It is necessary to integrate technology into the wider social context, to provide relevant content, education and training, and build institutional support. Only through such a comprehensive approach is it possible to achieve real digital language inclusion and reduce the consequences of digital language inequalities.

References

- Bella, G. – Helm, P. – Koch, G. – Giunchiglia, F. (2023): *Towards Bridging the Digital Language Divide*, Xiv:2307.13405v1 [cs.CL].
- European Parliament (2016): Science and Technology Options Assessment (STA), Language equality in the digital age EPRS | European Parliamentary Research Service, Scientific Foresight Unit (STOA) PE 581.937, Brussels, 25–33. <https://journals.sagepub.org/index.php/jamm/article/view/60> Date of access: 20 March, 2025.

- European Parliament (2018): European Parliament Resolution of 11 September 2018 on language equality in the digital age (2018/2028(INI)) OJ C 433 <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52018IP0332> Date of access: 20 March, 2025.
- Gaspari, F. – Way, A. – Dunne, J. – Rehm, G. – Piperidis, S. – Giagkou, M. (2021): *European Language Equality-D1.1 Digital Language Equality (preliminary definition)*, ADAPT Centre, Dublin City University
- Giagkou, M. – Lynn, T. – Dunne, J. – Piperidis, S. – Rehm, G. (2023): European Language Technology in 2022/2023. In: Rehm, G. – Way, A. (eds.): European Language Equality. Cognitive Technologies. Springer, Cham. DOI: [10.1007/978-3-031-28819-7_4](https://doi.org/10.1007/978-3-031-28819-7_4)
- Gomes, A. – Dias, J. G. (2025): Digital Divide in the European Union: A Typology of EU Citizens, *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, Springer, 176(1), 149–172. DOI: [10.1007/s11205-024-03452-2](https://doi.org/10.1007/s11205-024-03452-2)
- Haight, M. – Quan-Haase, A. – Corbett, B. A. (2014): Revisiting the digital divide in Canada: The impact of demographic factors on access to the internet, level of online activity, and social networking site usage. *Information, Communication & Society*, 17(4), 503–519. DOI: [10.1080/1369118X.2014.891633](https://doi.org/10.1080/1369118X.2014.891633)
- Hale, S. A. (2013): Multilinguals and Wikipedia Editing. DOI: [10.1145/2615569.2615684](https://doi.org/10.1145/2615569.2615684)
- Hernandez, L. (2023): Factors Affecting the Digital Divide Among Underrepresented Groups, *Journal of Artificial Intelligence and Machine Learning in Management*, 7(1), 25–33.
- Kirchmeier, S. (2020): *Trends in European Language Policies With a View to Language Technology*. DOI: [10.35321/bkalba.2020.93.07](https://doi.org/10.35321/bkalba.2020.93.07)
- Kornai, A. (2013): Digital Language Death. *PLoS ONE*, 8(10), DOI: [10.1371/journal.pone.0077056](https://doi.org/10.1371/journal.pone.0077056)
- Maynard, D. – Wright, J. – Greenwood, M.A. – Bontcheva, K (2022): Deliverable D1.11 Report on the English Language. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. <https://european-language-equality.eu/reports/language-report-english.pdf> Date of access: 30 May, 2025.
- Moseley, C. (ed.) (2010): *Atlas of the world's languages in danger*, Paris, UNESCO Publishing.
- Papa, D. – Omazić, M. (2021): *Jezične politike i jezično planiranje za višejezičnost u Europskoj uniji*, Pravni fakultet u Osijeku.
- Pimienta, D. – Prado, D. – Blanco, Á. (2009): Twelve years of measuring linguistic diversity in the Internet: balance and perspectives. Paris, UNESCO. <http://unesdoc.unesco.org/images/0018/001870/187016e.pdf> Date of access: 20 March, 2025.
- Rehm, G. – Uszkoreit, H. (2012): Languages at Risk: A Challenge for Language Technology. In: Rehm, G. – Uszkoreit, H. (eds): *The English Language in the Digital Age. White Paper Series*. Springer, Berlin, Heidelberg. DOI: [10.1007/978-3-642-30684-6_2](https://doi.org/10.1007/978-3-642-30684-6_2)
- Rodriguez-Ordoñez, I. – Kasstan, J. – O'Rourke, B. (2022): Responding to sociolinguistic change: New speakers and variationist sociolinguistics.

- International Journal of Bilingualism*, 26(5), 529–541. DOI: [10.1177/13670069221110381](https://doi.org/10.1177/13670069221110381)
- Simons, G. F. – Thomas, A. L. – White, C. K. (2022): *Assessing Digital Language Support on a Global Scale*. In: Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea. International Committee on Computational Linguistics, 4299–4305.
- Toivanen, R. – Saarikivi, J. (2016): Language Endangerment and New Linguistic and Cultural Diversities. In: Toivanen, R. – Saarikivi, J. (eds.): *Linguistic Genocide or Superdiversity?: New and Old Language Diversities*, 3–6. Multilingual Matters, Bristol. DOI: [10.21832/9781783096060](https://doi.org/10.21832/9781783096060)
- UNESCO Ad Hoc Expert Group on Endangered Languages (2003): Language Vitality and Endangerment. Document submitted to the International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages, Paris.
- Van Dijk, J. (2020): *The Digital Divide*. Cambridge, Polity Press.
- Warschauer, M. (2003): *Technology and Social Inclusion Rethinking the Digital Divide*. The MIT Press Cambridge: Massachusetts London. DOI: [10.7551/mitpress/6699.001.0001](https://doi.org/10.7551/mitpress/6699.001.0001)
- Zook, M. A. – Graham, M. (2007): Mapping DigiPlace: Geocoded Internet Data and the Representation of Place, *Environment and Planning B Planning and Design*, 34(3), 466–482. DOI: [10.1068/b3311](https://doi.org/10.1068/b3311)