

TO BE DIGITALLY ALIVE OR NOT: THE CASE OF AN ENDANGERED OFFICIAL LANGUAGE*

*Gülin DAĞDEVİREN KIRMIZI***

Abstract: Today, the issue of Digital Language Vitality proves to be one of the central discussions in linguistic vitality, minority and endangered languages. In this study, the Gagauz, which is an endangered language spoken in the Gagauzia Autonomous Territorial Unit, Moldova, has been investigated on the basis of the digital language vitality. Digital Performance of the Gagauz language was explored under the categories of the Available Internet Services, Localised Social Networks, Localised Software, Machine Translation Tools/Services and Dedicated Internet Top-level Domain of a language. According to the independent web search, the Gagauz language has a varying degree of digital representation. The language can be labelled as none for the categories of Localised Social Networks, Localised Software and Dedicated Internet Top-level Domain. On the other hand, in terms of the available internet services, the language falls into the label of limited as there are some web pages in the Gagauz language. Similarly, Machine Translation Tools/Services in the Gagauz language can be labelled as medium. Lastly, the use of the Gagauz language in Wikipedia shows that the language is digitally labelled as Incubator which refers to a small Wikipedia. The findings are in the same line with the previous language vitality studies.

Keywords: Digital Language Vitality, Endangered languages, the Gagauz Language, the Russian Language.

Dijital Olarak Canlı Olmak ya da Olmamak: Tehlike Altındaki Resmi Bir Dilin Durumu

Öz: Günümüzde Dijital Dil Canlılığı konusu dilbilimsel canlılık, azınlık ve tehlike altındaki dillerle ilgili temel tartışmalardan biri olmuştur. Bu çalışmada Moldova Gagauzya Özerk Bölgesinde konuşulan tehlike altındaki Gagauzca dijital dil canlılığı açısından değerlendirilmiştir. Gagauzcanın dijital performansı mevcut internet hizmetleri, yerleştirilmiş sosyal ağlar, yerleştirilmiş yazılım, makine çevirisi araçları/hizmetleri, tahsisli üst düzey alan adı uzantısı başlıkları altında incelenmiştir. Serbest internet araması sonuçlarına göre Gagauzcanın çeşitli seviyelerde dijital temsili vardır. Dil yerleştirilmiş sosyal ağlar, yerleştirilmiş yazılım ve Tahsisli üst düzey alan adı başlıkları kategorilerinde “yok” olarak sınıflandırılmıştır. Diğer yandan bazı Gagauzca web sayfaları olduğundan dil mevcut internet hizmetleri bakımından sınırlı olarak sınıflandırılmıştır. Benzer biçimde makine çevirisi araçları/hizmetleri açısından Gagauzca orta olarak derecelendirilebilir. Son olarak Gagauzcanın Wikipedia’da kullanımı dilin küçük Wikipedia’ya karşılık gelen kuluçka seviyesinde olduğunu göstermektedir. Sonuçlar önceki dil canlılığı çalışmalarıyla aynı doğrultudadır.

Anahtar kelimeler: Dijital Dil Canlılığı, Tehlike Altındaki Diller, Gagauzca, Rusça.

* Makalenin Geliş ve Kabul Tarihleri: 29.03.2019 - 11.12.2019

** Dr.Öğr.Üyesi, Başkent Üniversitesi, Eğitim Fakültesi, Yabancı Diller Eğitimi Bölümü, İngiliz Dili Eğitimi Ana Bilim Dalı, Ankara, Türkiye. gdagdeviren@baskent.edu.tr, ORCID: 0000-0002-4675-1040.

Introduction

In the context of endangerment, the digital presence of a language has importance for its vitality. Availability of the digital softwares, services, tools and networks of a language, and the proficiency of digital skills of its speakers is thought to contribute its vitality. Brenzinger et al. (2003, p. 11) asserted that ‘new media, including broadcast media and the Internet, usually serve only to expand the scope and power of the dominant language at the expense of endangered languages’. Taking this into consideration, it can be said that the future of the endangered languages seems rather vague, if they are not supported. While digital technologies are mostly in favour of mainstream languages, endangered languages’ practice are restricted. Together with availability of these services, the readiness and competency of its users is another issue at this point. As Crystal (2000, p. 141) claimed ‘an endangered language will progress, if its speakers can make use of electronic technology’. Therefore, the issue of language vitality and endangerment need to be investigated from various perspectives.

A considerable amount of literature has been published on linguistic vitality since 1970s. The purpose of these studies has been to investigate the contexts of diglossia (Landry and Allard, 1994), a majority language (Young, Bell and Giles, 1988), indigenous languages (Florey, 2009) etc. In order to assess language vitality various methodologies have been developed and used so far. Some of them are Fishman’s (1991) Graded Intergenerational Disruption Scale, Florey’s (2009) Linguistic Vitality Test, Harmon and Loh’s (2010) The Index of Linguistic Diversity, McConvell and Thieberger’s (2001) Language Endangerment Indicators, and UNESCO Ad Hoc Expert Group’s (2003) Language Vitality and Endangerment Scale. As can be clearly understood, the path from language vitality to digital vitality has a relatively recent history.

1. The Scales of (Digital) Language Vitality

The first systematic assessment of language vitality was conducted by Fishman (1991), who developed Graded Intergenerational Disruption Scale (hereafter, GIDS), which emphasizes the role of intergenerational transmission in the vitality of languages. GIDS provides 8 levels of intergenerational disruption. Level 1 defines the situation where the language is used in various domains such as education and mass media; on the other hand, Level 8 is used for the languages whose speakers are of grandparent generation. In their study Lewis and Simons (2010, p. 106) discussed GIDS and listed shortcomings of the scale. These are a) description, in static terms, b) inadequate description of the statuses, c) emphasis on home domains and local community and d) limited descriptions of the levels of disruption. Having revised GIDS, Lewis and Simons (2010) added three categories (namely Levels 0, 9 and 10) and developed Extended Graded Intergenerational Disruption Scale (hereafter, EGIDS). Moreover, they divided

two levels into two other categories such as the Level 6a and 6b. Consequently, EGIDS become a more detailed and sensitive scale for the investigations of language maintenance on the basis of institutional and community levels.

Having inspired from GIDS and EGIDS, Kornai (2013) proposed four-level categorization of digital language death which are *thriving* (T), *vital* (V), *heritage* (H), and *still* (S). Before explaining this categorization, it is necessary to consider the factors which are influential in the digital language maintenance. According to Kornai (2013, p. 1), “(i) the size and demographic composition of the language community; (ii) the prestige of the language; (iii) the identity function of the language; (iv) the level of software support; and (v) Wikipedia” are the factors leaning the language death. Community size was investigated by Kornai (2013, p. 2) on the basis of the number of digital natives. Secondly, with regard to prestige, it can be said that the use of a language in digital communication is prestigious. Thirdly, identity function of a language refers to a crucial distinction between passive and active web presence. Active web use which includes social networks, business/commerce, live literature etc. is a good indicator of digital vitality (Kornai, 2013, p. 3). Software support is another factor to facilitate the spread of a language in digital realm. Lastly, the use of a language in Wikipedia which is one of “the very first active digital language communities” is a good indicator on the basis of language maintenance and endangerment. Taking these into consideration, Kornai (2013) used the term *thriving* to refer languages which are frequently available in digital domains in other words these are the languages with high level of digital use (Gibson, 2016). Kornai (2013) exemplifies *thriving* as 100.000 digital natives while fewer of them (more than zero) is considered to be *vital* (p. 5). The languages in *thriving* category are English, French, German, etc. The category of *heritage* is defined and exemplified by Kornai (2013, p. 5) as “those with zero L1 speakers but more than n2 (say, 100) L2 speakers”. The digital natives in the category of *heritage* are created by the outsiders, namely language specialists therefore mostly they are not accessible by native speakers. Lastly, *still* refers to the no use of language in digital domains.

Gibson (2016) revised Kornai’s classification and emphasized the importance of micro-uses of such as text messaging, computer-mediated communication. Gibson’s (2016, p. 48) main argument is that “text messaging does not follow formal norms for English, and is a sociolinguistic domain where nonstandard and non-normative practices are encouraged”. Therefore, he proposed categories of *emergent* and *latent* and excluded Kornai’s (2013) category of *heritage*. The status of *emergent* refers to the situation where speakers use the language in personal text messaging and social media. On the other hand, Gibson (2016, p. 49) differentiates *latent* from *emergent* on the basis of “stable intergenerational transmission, an available model of writing the language, the availability of appropriate technology and fonts in which to write the language in the desired;

and communal desire to see the language used digitally”. However, instead of the categories of *latent* and *emergent*, Gibson’s category of *heritage* was criticized, afterwards.

Soria et al. (2017) criticized Kornai and Gibson’s category of *heritage* as it refers to a language which does not exist but still available in digital domains. Soria et al. (2017) offered DLVS (Digital Language Vitality Scale, hereafter), which includes six categories; *Pre-digital*, *Dormant*, *Emergent*, *Developing*, *Vital* and *Thriving*. These categories are explained below:

- A language which is in *Pre-digital* stage cannot expand digitally as its speakers have no or low proficiency in using digital media technologies. Moreover, in this stage character encoding or writing system is not available in digital domains and connectivity is not affordable.
- The second level, *Dormant*, describes the context where some of the speakers are literate in digital domains while there is no technological support for the language.
- In the next stage, which is named *Emergent*, texting, messaging and the use of social media is rare and limited.
- Soria et al.’s (2017) next level is *Developing* where more frequent use of social media and medium-sized Wikipedia are available.
- On the other hand, languages which have regular online communication and a variety of resources are labelled as *vital*. These languages have localized software and online translation services.
- The last category of Soria et al.’s (2017) classification is *Thriving*. The term is used for the languages which are used in all digital domains widely (including communication and transaction). The speakers of these languages have no digital barriers and they use the latest technology.

Soria et al. (2017, p. 6) used a set of indicators to assess the digital vitality: digital capacity, digital presence and use and digital performance. *The digital capacity* is investigated on the basis of the “evidence of connectivity, digital literacy, internet penetration or digital population size, character/script encoding and availability of language resources”. The second set of indicators, *digital presence and use*, is divided into the language’s “use for e-communication, use on social media, availability of Internet media and Wikipedia” (Soria et al., 2017, p. 6). Being investigated in the current study, the last set of indicators, digital performance, include “available Internet services, localised social networks, localised software, machine translation tools/services, dedicated Internet top-level domain” (Soria et al., 2017, p. 6). These indicators were explored in the context of the Gagauz language in the current study.

2. The Gagauz Language and Gagauzia Autonomous Territorial Unit

The Gagauz language is member of the south western branch (West Oghuz Turkic) of Turkic languages (Johanson, 1998, p. 82). Today, it is spoken in various parts of the eastern Europe such as Ukraine, Moldova and Bulgaria. In this study, the variety in Bessarabia, classified as a definitely endangered language by Moseley (2010), was investigated. Geographically, the area between Dniester and Prut rivers is called Bessarabia; however politically it lies within the borders of Ukraine and Moldova today. Moldova has a critical role in the vitality of the Gagauz language. The Gagauz people demanded for and established Gagauzia Autonomous Territorial Unit (hereafter, ATU) of Moldova in 1994. Being composed of three administrative districts, -Comrat, Ceadâr-Lunga and Vulcănești- Gagauzia is located in the south-eastern Moldova. According to the National Bureau of Statistics (2014), the population of Gagauzia ATU was 134,535. Previous census held in 2004 showed that the autonomous territory has a multi-ethnic population. The National Bureau of Statistics of the Republic of Moldova (2004) showed that apart from the Gagauz people, Bulgarians, Moldovans/Romanians, Russians and Ukrainians are living in the area. Therefore, the Law on the Languages of Gagauzia ATU was regulated to guarantee the rights concerning the Gagauz, Russian and Moldovan Languages in legal procedure, legislation, personal names, topographical indications and education (Sirkeli and Lisenco, 2012). However, in practice, the field observations show that Russian is the dominant language in various domains of life (Dağdeviren-Kırmızı, 2015; Menz, 2003). In the same vein, having analysed the legal framework, Sirkeli and Lisenco (2012, p. 8) asserted that Russian is used dominantly in the “Internet resources of public authority bodies, local public administrations and institutions of Gagauzia”. What Sirkeli and Lisenco (2012) claimed was statistically supported by the Nantoi et al. (2016), who found that 78,5% of the respondents consume information in the Russian language in the Gagauzia ATU. On the other hand, according to their findings, only 0.9% of the respondents use websites in the Gagauz language. It can be said that the wide gap between the percentages of the languages may indirectly evidence the lack of the web content in the Gagauz language.

3. The Present Study

In the current study Soria et al.'s (2017) categories of *Digital Performance* and *Wikipedia as an indicator of Digital Presence and Use* were adopted to assess the digital vitality of the Gagauz language. According to Soria et al. (2017, p. 18), the indicators in Digital Performance clearly show “what can be digitally done with a language”. Therefore, Available Internet Services, Localised Social Networks, Localised Software, Machine Translation Tools/Services and Dedicated Internet Top-level Domain of a language, were chosen to assess the presence of the Gagauz language. In addition to Digital Performance category,

the existence of Wikipedia in the context of digital vitality, which is strongly emphasized by (Kornai, 2013; Miola, 2013), was taken as an indicator in this study. The data collected through web crawling as an independent research. It is beyond the scope of this study to examine the a) evidence of connectivity, b) digital literacy, c) Internet penetration or digital population size, d) digital capacity, e) character/script encoding and f) availability of language resources, as there has not been any statistics about Gagauz internet users. Similarly, as the categories about the use of the language in g) e-communication, h) social media, and i) the availability of Internet media, require the surveys or censuses, this study cannot take them into consideration.

4. Results

4.1. Digital Performance and Use

4.1.1. Available Internet Services

Soria et al. (2017, pp. 17-18) investigated localized digital presence in terms of the “online newspapers, online news, search engines, edutainment products and services, entertainment, video subtitling, health services, e-Commerce services, public Administration/e-Government, e-Banking, cultural services, tourist information and services, advertising, promotion and marketing, customer care”. The lack of registries or catalogues about the available internet services of the Gagauz language has made imperative independent research.

According to the results, the services of edutainment, subtitling, online health, e-Commerce/e-Banking, public administration, cultural events of places, online advertisement or marketing are not offered in the Gagauz language. According to Diozu (2015, p. 307) the number of pages in the Gagauz language in 2012 and 2013 was below 20 and in the following year, 2014, there were no Gagauz pages in governmental websites. The official website of Gagauzia ATU is mostly in Russian. Similarly, as Diozu (2015, p. 308) stated there were no pages of commercial websites in the Gagauz language between 2006 and 2014. On the other hand, there was an increase in 2012 and then again a decrease in 2013 in the pages in the Gagauz language for the non-commercial web sites (Diozu, 2015, p. 308).

However, it is seen that there are websites of online newspapers and news in the Gagauz language. *Anasözü* is the only print and online newspaper in this language at the time of writing. Published monthly in print since 1988 and online since 2011, *Anasözü* has completely Gagauz content while *Meydan Gazetası* (Meydan Newspaper) has content in the Gagauz language and in Anatolian Turkish. It is observed that the website has not been continually and fully updated. The official website of *Gagauziya Radio Televizionu* (Gagauz Radio and Television) broadcast news in video format which are mostly in Russian and then in the

Gagauz language. In conclusion, the lack of many internet services and the insufficiency of monolingual news websites show **limited (Grade 3)** representation of the Gagauz language. This categorization refers to the situation where one or two digital services available.

Table 1. Soria et al.'s (2017, p. 19) classification of Available Internet Services

Label	Grade	Internet Services
None	2	No digital services available in the language (= less than 1)
Limited	3	Some digital services available in the language (= 1 to 2)
Medium	4	Some digital services available in the language (= 3 to 5)
Strong	5	a considerable variety of digital services is available (= 6 to 8)
Advanced	6	a wide variety of digital services is available (≥ 8)

4.1.2. Localised Social Networks

Although Scannell (2012, p. 2) asserted that “social media will not save your language”, it is a fact that it plays one of the pivotal roles in language maintenance and revitalization. Localized social networks help speakers to use the services in their native languages. Supported languages in social networks can serve as evidence of digital salience. However, there have been a limited number of studies on the effects of localized social networks on language vitality. In one of these, Cunliffe (2007) investigated the use of Welsh on *Facebook* and found a correlation between the existence of localized language and young bilinguals’ language behaviour. As there are not survey results or statistics regarding the number of the users living in the Gagauzia ATU, Moldova, the most practical way to investigate localized social networks is web crawling, more specifically checking out the language options offered by the network.

It is seen that the search engines like Google and Yandex do not offer interface in the Gagauz language to their users. According to Alexa (2019), apart from the search engines, the most visited websites of Moldova include social networks such as *YouTube*, *Vkontakte* and *Odnoklassniki*, respectively. *YouTube*, which is a video sharing platform, facilitates its users to create and watch the video content. It supports 80 languages which include Russian and Romanian; however, there is no support for the Gagauz language yet. Being the most popular social network in Eastern Europe, *Vkontakte* is likely to be the most commonly used social network in the Gagauz context. According to *Vkontakte*, it has 97 million monthly active users. However, the Gagauz language is not one of the supported 86 languages. *Odnoklassniki* (Ok.ru), another website for social networking, is popular in Russian and the Post-Soviet context. The number of the languages that

the website supports is 11 and the Gagauz language is not one of them. Similarly, although it has 2.27 billion monthly active users Facebook does not support the Gagauz interface (*Facebook*, n.d.). *Facebook*, at the time of the writing, supported 108 languages.. According to Twitter (n.d.), a micro-blogging service, with 326 million monthly active users in the last quarter of 2018 that supports 34 languages, there is no interface in the Gagauz language.

In conclusion, the web research shows that the Gagauz language is underrepresented in most popular social network services. Additionally, as there has not been any statistics about the numbers of the Gagauz speaking users, little has been known about active Gagauz users of these networks. As a result, it seems that the label for the Gagauz language would have been **none (Grade 3)** had it not been any localized social media.

Table 2. Soria et al.'s (2017, p. 20) classification of Available Internet Services

Label	Grade	Localised Social Networks
None	3	No social media localised in the language (= < 1)
Limited	4	At least one social media interface localised in the language (= 1)
Medium	5	Some social media interface localised in the language (= 2 to 3)
Advanced	6	Many social media interface localised in the language (= > 3)

4.1.3. Localised Software

According to Soria et al. (2017) localized software and operating systems can be evaluated as the indicators of digital vitality. The authors also emphasize that even the ones who undertake the process of localization are important. Whether they are system developers or local stakeholders can give hints about the sociolinguistic situation (Soria et al., 2017).

Assessment of a language on the basis of localized software and operating systems in this study is based on internet research. In their study, Soria et al. (2017, p. 21) investigated the digital representation in *Windows*, *Mac OS X*, *Linux*, *Android*, *iOS*, *Microsoft Office*, *LibreOffice*, *Firefox*, *Chrome*, *Internet Explorer*, *Thunderbird*, *Adobe Creative Suite*, and *Gimp*. The results in this study show that these software are not localized for the Gagauz language. Therefore, the *iOs*'s *Siri* voice assistant and the *Android*'s *Bixby* are not available in the Gagauz language. Similarly, the keyboard used in *Whatsapp*, a cross-platform messaging service, is not supported in the Gagauz language. A note of caution is due here, since, although Windows 10 provides an option to select the Gagauz language in Latin and Cyril alphabets, handwriting, text-to-speech and display language are not available in the Gagauz language.

Consequently, the label of the Gagauz language would be **none (Grade 2)**. This defines the situation where “neither operating system nor general purpose software localised in the language”.

Table 3. Soria et al.’s (2017, p. 20) classification of Localised Software

Label	Grade	Localised software
None	2	Neither operating system nor general purpose software localised in the language
Limited	3	At least one operating system (either desktop or mobile, either open or commercial) localised in the language
Medium	4	At least one desktop and one mobile operating system (either open or commercial) + some general purpose software (a word processor and a browser) localised in the language
Strong	5	Most used operating systems and general purpose software localised in the language; some specific purpose application software localised.
Advanced	6	Main operating systems and application software localised in the language.

4.1.4. Machine Translation Tools/Services

Open source Machine Translation tools or services help users to translate texts from a source language to a target language. The importance of Machine Translation in the context of minority or endangered languages arises from the availability of digitalized form of language. According to Soria et al. (2017) the availability of Machine Translation tools is an indicator of multilingual corpora. Forcada (2006) states that Machine Translation tools have a positive effect on normality, literacy levels and visibility; additionally, it contributes to standardization of a minority language.

The popular search engines like *Google* and *Yandex*, offer open source translation services. Google Translate which supports over a hundred languages does not include the Gagauz language. Similarly, although it supports 95 languages, Yandex Translate does not offer any translation service for the Gagauz language.

Apart from the search engines’ translation services, there are small-sized open source translation websites for the Gagauz language. The first of these online translation tools is *Freelang* developed by Figueiredo (2014). It has 2,254 words from Gagauz to English and 1,747 words from English to Gagauz. Although users can add or modify an entry, the last update was done in March 21st 2014. It is clear that the website has not been actively used since. The second online open source website for Gagauz translation is the bilingual *Glosbe* (n.d). It has 787 translated phrases and 71 sentences from English to the Gagauz language. It is also possible to use its Russian to Gagauz interface which includes 790 phrases and 70 sentences. Similar to *Freelang*, *Glosbe* has a publicly edited content.

According to the Wikipedia (2019) report which compares language features of machine translation programs, Gagauz is not supported by *Pairaphrase*, *Xerox Easy Translator*, *Omniscien Technologies Language Studio*, *IBM Watson*, *Apertium*, *OpenLogos*, *Slate Rocks*, *Moses for Mere Mortals*, *Babel Fish*, *Bing Translator*, *GramTrans*, *SYSTRAN*, *IdiomaX*, *SAIC Omnifluent* and *lingenio translate*. Thus, considering these findings, the label of the Gagauz language for Machine Translation Tools and Services according to Soria et al.'s (2017, p. 22) categorization would be **medium that is**, “at least one (online) service / tool, at least two language pairs in both directions” is available.

Table 4. Soria et al.'s (2017, p. 22) classification of Machine Translation tools/services

Label	Grade	Machine translation tools/services
None	3	No MT for the language
Basic	4	at least one (online) service/ tool, at least one language pair or one direction
Medium	5	at least one (online) service / tool, at least two language pairs in both directions
Advanced	6	more than one (online) service /tool, more than 5 language pairs

4.1.5. Dedicated Internet Top-level Domain

According to IANA (Internet Assigned Numbers Authority) (2018), top-level domains are “the highest level of subdivisions with the domain name system [...] divided into two distinct categories, generic top-level domains and country-code top-level domains.” The glossary on IANA's (2018) website defines generic top-level domains as the ones used for general purposes, while country-code top-level domains are assigned to countries which achieve certain standards. As for the role of top-level domains in digital language vitality, Soria, et al. (2017, p. 23), emphasized that “languages having their own top-level domain name present a non-negligible digital activity, indicating that there is some level of correlation between the existence of the domain and digital vitality.” However, this perspective can be misleading as these codes are assigned to countries as mentioned below, however, for the case of the Gagauz language, the variety investigated in this study is spoken within the territories of the Autonomous Territorial Unit of Gagauzia in Moldova. Although Moldova has a country-code (.md), it is seen that the official websites of the Gagauzia uses the top-level domain of .md. More specifically, it would be impossible to discuss the representation of a language by taking the availability of its top-level domain. According to Soria et al.'s (2017, p. 23) classification, the Gagauz language falls into the category of **Grade 3 (No)** as there are “no dedicated Internet top-level domain”.

Table 5. Soria et al.'s (2017, p. 23) classification of Dedicated Internet top-level domain

Label	Grade	Dedicated Internet top-level domain
No	3	No dedicated Internet top-level domain
Yes	5	There is a dedicated Internet top-level domain

4.2. Wikipedia as an Indicator of Digital Presence and Use

Although there are inconsistencies between active interest and digital use (Soria et al., 2017), *Wikipedia* can be considered a good indicator of digital vitality. Kornai (2013, p. 3) emphasized that “*Wikipedia* is always among the very first active digital language communities, and can be safely used as an early indicator of some language actually crossing the digital divide”. *Wikipedia* also provides statistics about the language-based uses.

The data about the number of articles, total users vs. active users and the statistics of page views are available at *Meta Wikimedia* and updated on a daily basis. According to the *Meta Wikimedia* (2018), the Gagauz language has 2759 *Wikipedia* articles. Therefore, the Gagauz language falls into the category of languages that have more than one thousand articles. When the number of the users in Gagauz language is taken into consideration, it becomes apparent that there are 8304 users. However, the number of the active users is 14. Taking similar numbers about the Gagauz language in *Wikipedia* into consideration, Kornai (2013) asserted that the language can be classified as “borderline”. This refers to languages that “the statistical evidence is not strong enough to make an individual determination with high confidence” (Ács, Pajkossy and Kornai, 2017, p. 333).

According to the statistics from *Meta Wikimedia* (2018), the Gagauz language is found in the list of languages which have zero local media files. The last upload in the format of media file was done in 2016-03-20. There has been no new upload for more than two years.

The list of *Wikipedia* classifies article sizes into four categories as *absent*, *stubs*, *articles* and *long articles*. The classification is done through multiplying language weigh by the number of the characters in the texts. *Absent* class has no article, while the category of *stubs* includes articles with characters less than 10,000. The term ‘*articles*’ is given to the groups with a number of characters between 10,000 and 30,000. The last category ‘*long articles*’ has the largest size compared to the others. The languages in this category have article sizes more than 30,000 characters. The article size of the Gagauz language is 4,389 and the numbers of texts under the categories of *absent*, *stubs*, *articles* and *long articles* are given in Table 6.

Table 6. The article size of the Gagauz language in *Meta Wikimedia* (2018)

Absent	Stubs	Articles	Long Articles
900	86	14	0

As seen from above, there are 900 entries in the Gagauz language. However, these entries do not include texts; therefore, they are grouped as absent. The numbers of texts less than 10.00 and 30.000 characters are 86 and 14 for the Gagauz language, respectively. Lastly, there are no texts more than 30,000 characters. The numbers given above show a low-level of representation of the language.

According to the data of *Wikimedia*, there are 8,347 registered users in the Gagauz language. However, the number of active users is reported to be 12 *Gagauz Vikipediya*. The term “active user” refers to the ones who have made at least one edit in the last thirty days (Meta-Wikimedia, 2018).

Therefore, the grading of the Gagauz language for the scales would be **incubator (Grade 2)** which refers to languages with less than a hundred articles.

Table 7. Soria et al.’s (2017, p. 18) classification of *Wikipedia*

Label	Grade	Wikipedia
None	1	no Wikipedia
<i>Incubator</i>	2	<i>a small Wikipedia is available (less than 100 articles)</i>
Small	3	between 100 and 10,000 articles
Medium	4	between 10,000 and 100,000 articles.
High	5	between 100,000 and 1,000,000 articles
Big	6	over 1,000,000 articles

Discussion and Conclusion

In this study Soria et al.’s (2017) indicators of digital vitality were adopted to investigate the representation of the endangered Gagauz language. The assessment was done utilizing the categories of digital performance and the use of Wikipedia. According to the findings of independent web research, the Gagauz language has a varying degree of representation on the internet and in digital technologies.

The results of available internet services show that the Gagauz language has **limited (Grade 3)** digital vitality. Except from an online version of a printed newspaper, there are no other internet services such as edutainment or e-commerce/e-banking. On the basis of the localized social networks, the Gagauz language can be labelled as **none (Grade 3)**. Similar results were obtained for localized software such as a localized operating system or application software in the Gagauz language. The language is not represented, thus, it can be labelled as **none (Grade 2)**. Machine translation tools and services is the category that the Gagauz language is relatively better represented, compared to other indicators. Independent web search showed that although there is a lack of language options

for translation provided by the mainstream search engines, there are small-scaled online dictionaries for the Gagauz language. Therefore, it can be labelled as **medium (5)**. The last indicator assessed in the category of digital performance is the dedicated Internet Top-level domains. According to IANA (2018), there is no dedicated top-level domain of the Gagauz language. Therefore, the language can be labelled as **no (Grade 3)**. Lastly, the indicator about Wikipedia showed that the Gagauz language can be labelled as **Incubator (Grade 2)**, if the articles in this language is taken into consideration.

Within the context of the Gagauz language, on the other hand, it is a fact that access to internet is practically difficult in the area. Moldova, where the services of telecommunications are provided by the Ministry of Transport and Communications, has one of the lowest rates of mobile phone and internet users (Latham, Buckmaster, Hitchins and Nikolayevich Sukhopara, 2019). Consequently, as Mooko (2006, p. 122) stated, it is expected that “such technology is most accessible to the affluent and educated members of the society, and not necessarily to the majority of the ordinary people who speak the endangered languages.” Taking into consideration that, the Gagauz is spoken mostly by the elderly in rural areas (Dağdeviren-Kırmızı, 2015) and that young Gagauz people are more proficient in digital skills and literacy, it is apparent that the Gagauz language would be poorly represented in digital domains.

On the other hand, the existence of a realistic demand by the speakers of the community could also help the increase of the digital representation of the language. The attempts of the communities or volunteered groups like Kanienkeha, Inupiaq, Chitimacha (Galla, 2009), Aymara, Chatino, Mixteco, Mixe (Jany, 2018) have resulted in creating tools or orthographic conventions in these languages. In the case of the Gagauz language, there are no major demands by the speakers. It is clear that the Gagauz people find the use of Russian more functional in cyberspace as well as in many domains of life. Thus, the issue of sustainability arises at this point. Since the mere existence of the digital services and tools do not support the vivacity of an endangered language, only speakers who achieve digital literacy and have desire to use the facilities could play a major role in vitalizing the language.

Taking these into consideration, it is apparent that the Gagauz is not enjoying digital vitality and it needs strategies to encourage its representation. From the perspective of native speakers, the Gagauz people’s willingness and readiness to create audio and visual contents in the Gagauz language would be decisive in the potential practices in digital domains. Secondly, the localization of the most popular social networking websites like *Vkontakte* and *Odnoklassniki* would expand the scope of the domains where the Gagauz language can be used which in return might result in boosting the linguistic confidence of native speakers.

It is clear that further studies should be carried out in order to explore the digital vitality in detail. As mentioned above, due to the lack of data about the internet infrastructure, facilities and rates of availability for use by the Gagauz speakers in Gagauzia ATU, it is not possible to reach solid conclusions. Additionally, further research needs to examine the native speakers' attitudes towards using internet services in the Gagauz language. The lack of confidence about the proficiency in the native language and the desire to be understood (Viernes, 2018); or the undervalued and negatively stereotyped identity as a native language speaker are contributing factors to negative attitudes towards the use of a specific language in a specific domain. This would also shed light on the functions of the Gagauz and Russian language in the context of digital vitality.

References

- Ács, J., Pajkossy, K. and Kornai, A. (2017). Digital Vitality of Uralic Languages. *Acta Linguistica Academica*, 64(3), pp. 327-345.
- Alexa, (2019). *Top Sites in Moldova*. Retrieved 22nd October 2018 from <https://www.alexa.com/topsites/countries/MD>.
- Brenzinger, M., Yamamoto, A., Aikawa, N., Koundioubá, D., Minasyan, A., Dwyer, A., ... Grinevald, C. (2003). *Language Vitality and Endangerment*. Paris, UNESCO Ad Hoc Expert Group Meeting on Endangered Languages. Retrieved in 9th August 2013 from <http://www.unesco.org/culture/en/endangeredlanguages>.
- Crystal, D. (2000). *Language Death*. Cambridge, UK: Cambridge University Press.
- Cunliffe, D. (2007). *Minority Language Media. Concepts, Critiques and Case Studies*. Clevedon, UK: Multilingual Matters.
- Dağdeviren-Kırmızı, G. (2015). *Emotional and Functional Attitudes of Native Speakers towards Gagauz as an Endangered Language*. (Unpublished doctoral dissertation). Hacettepe University, Ankara.
- Diozu, V. (2015). Linguistic Preferences in the Moldavian National Cyberspace: Reflection of Political, Economic and Migration Processes in the Society. Paper presented at 3rd *International Conference*. Retrieved 23rd January 2019 from http://ifapcom.ru/files/2015/khanty/yak_mling_2015.pdf.
- Facebook. (n.d.). Number of Monthly Active Facebook Users Worldwide as of 3rd Quarter 2018 (in millions). In Statista - The Statistics Portal. Retrieved 17th December 2018, from <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>.
- Figueiredo, R.B. (2014). *Features of this dictionary*. Retrieved 14th March 2019 from <https://www.freelang.net/dictionary/gagauz.php>.
- Fishman, J. A. (1991). *Reversing language shift*. Clevedon, UK: Multilingual Matters Ltd.
- Florey, M. (2009). Supporting Indigenous Languages and Indigenous Knowledge: Developing Community Training Approaches for the 21st Century. In P. Bates, M. Chiba, S. Kube and D. Nakashima (Eds.), *Learning and Knowing in Indigenous Societies Today* (pp. 25-37). Paris: UNESCO.

- Forcada, M. (2006). Open-Source Machine Translation: An Opportunity for Minor Languages. Paper Presented at *The 5th SALT MIL Workshop on Minority Languages Strategies for Developing Machine Translation for Minority Languages*. Retrieved 17th November 2018 from <https://pdfs.semanticscholar.org/8e9d/b064a62261ba10152c11cd6d05bb306e7cc3.pdf>.
- Galla, C. (2009). Indigenous Language Revitalization and Technology: From Traditional to Contemporary Domains. In J. Reyhner and L. Lockard (Eds.), *Indigenous Language Revitalization: Encouragement, Guidance and Lessons Learned*: pp.167-182. Flagstaff, AZ: Northern Arizona University.
- Gibson, M. (2016). *Assessing Digital Vitality: Analytical and Activist Approaches*. In C. Soria, et al. (Eds.), *Proceedings of the LREC 2016 Workshop, CCURL 2016 - Towards an Alliance for Digital Language Diversity* (pp. 46-51).
- Glosbe, (n.d.). *İngilizce Gagavuzca Sözlük Çevrimiçi*. Retrieved 8th December 2018 from <https://glosbe.com/en/gag>.
- Harmon, D. and Loh, J. (2010). The Index of Linguistic Diversity: A New Quantitative Measure of Trends in the Status of the World's Languages. *Language Documentation & Conservation*, 4, pp. 97-151.
- Internet Assigned Numbers Authority. (nd). *Root Zone Database*. Retrieved in 17th September 2018 from <https://www.iana.org/domains/root/db>.
- Jany, C. (2018). The Role of New Technology and Social Media in Reversing Language Loss. *Speech, Language and Hearing*, 21(2), ss. 73-76.
- Johanson, L. (1998). The History of Turkic. In L. Johanson and É. A. Csató (Eds), *The Turkic Languages* (pp. 81-125). London: Routledge.
- Kornai, A. (2013). *Digital Language Death*. PLoS ONE 8(10), Retrieved 11th November 2018 from <http://www.plosone.org/article/info:doi/10.1371/>.
- Landry, R. and Allard, R. (1994). Diglossia, Ethnolinguistic Vitality, and Language Behavior. *International Journal of the Sociology of Language*, 108, pp. 15-42.
- Latham, E., Buckmaster, B., Hitchins, K.A. and Nikolayevich Sukhopara, F. (2019). Moldova. Retrieved 12th February 2019 from <https://www.britannica.com/place/Moldova>.
- Lewis, M.P. and Simons, G. (2010). Assessing Endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique*, 55(2), pp. 103-120.
- Mail.ru. (n.d.). *Number of Monthly Active Mobile VKontakte Users from March 2014 to December 2017 (in millions)*. In *Statista - The Statistics Portal*. Retrieved 17th December 2018, from <https://www.statista.com/statistics/425429/vkontakte-mobile-mau/>.
- McConvell, P. and Thieberger, T. (2001). State of Indigenous Languages in Australia. *Australia State of the Environment Second Technical Paper Series No. 2 (Natural and Cultural Heritage)*. Canberra: Department of the Environment and Heritage.
- Menz, A. (2003). Endangered Turkic Languages: The Case of Gagauz. In M. Janse and S. Tol (Eds.), *Language Death and Language Maintenance: Theoretical, Practical and Descriptive Approaches* (pp. 143-155). Amsterdam: Benjamins.
- Meta Wikimedia. (2018). *List of Wikipedias*. Retrieved 13th November 2018 from https://meta.wikimedia.org/wiki/List_of_Wikipedias.

- Miola, E. (2013). A Sociolinguistic Account of Wikipiedmontese and Wikilombard. *Sociolinguistica*, 27, pp. 116-131.
- Mooko, T. (2006). Counteracting the Threat of Language Death: The Case of Minority Languages in Botswana. *Journal of Multilingual and Multicultural Development*, 27(2), pp. 109-125.
- Moseley, C. (Ed.). (2010). *Atlas of the World's Languages in Danger* (3rd ed.). UNESCO Publishing: Paris. Retrieved 23rd August 2018 from <http://www.unesco.org/culture/en/endangeredlanguages/atlas>.
- Nantoi, O., Cantarji, V., Botan, I., Gremalschi, A. and Sirkeli, M. (2016). *Moldova between East and West: Views from Gagauzia and Taraclia*. Chisinau. Retrieved 24th September 2018 from http://ipp.md/wp-content/uploads/2016/09/Moldova-between-East-and-West-Views-from-Gagauzia-and-Taraclia_EN.pdf.
- National Bureau of Statistics of the Republic of Moldova. (2004). *2004 Population Census*. Retrieved in 23rd January 23 2018 from <http://www.statistica.md./pageview.php?l=en&idc=350&id=2208>.
- National Bureau of Statistics of the Republic of Moldova. (2014). *Population and Housing Census in the Republic of Moldova*. Retrieved in 12th March 2018 from <http://www.statistica.md/pageview.php?l=en&idc=479>.
- Odnoklassniki. (2019). *Languages*. Retrieved 25th January 2019 from <https://ok.ru/>
- Scannell, K. (2012). *Translating Facebook into Endangered Languages*. Retrieved 21st December 2018 from <https://borel.slu.edu/pub/fel12.pdf>.
- Sirkeli, M. and Lisenco, S. (2012). *Policy Brief: Implementation of Linguistic Rights of the Gagauz of Moldova. Integration of the Gagauz Community into the Society of Moldova*. Retrieved in 14th April 2014 from <http://www.edemocracy.md/files/elections/gagauzia2012/policy-brief-pilgrim-demo-gagauzia-2012-en.pdf>
- Soria, C., Quochi, V., Russo, I., Gurrutxaga, A. and Ceberio, K. (2017). *A Digital Language Vitality Scale and Indicators*. Retrieved 17th March 2019 from https://www.academia.edu/34966423/A_Digital_Language_Vitality_Scale_and_Indicators.
- Twitter. (n.d.). Number of Monthly Active Twitter Users Worldwide from 1st Quarter 2010 to 3rd quarter 2018 (in millions). In Statista - The Statistics Portal. Retrieved 17th December 2018, from <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/>.
- 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, 10-12 March 2003.
- Viernes, G.B. (2018). *Minority Languages On Social Media: Multilingual Practices and Digital Presence*. Paper presented in Conference on Frisian Humanities: Multilingualism, Ljouwert.
- VKontakte (n.d.). *Company Info*. Retrieved 18th February 2019 from <https://vk.com/about>.
- Young, L., Bell, N. and Giles, H. (1988). Perceived Vitality and Context: A National Majority in a Minority Setting. *Journal of Multilingual and Multicultural Development*, 9, pp. 285-289.