

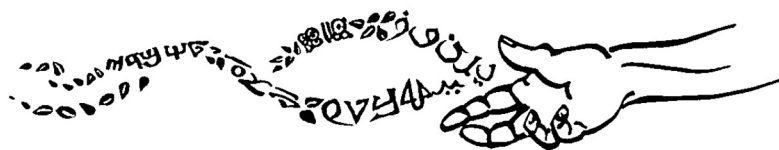
# FEL XXI Alcanena 2017

## Communities in Control

Foundation for Endangered Languages  
CIDLeS  
SOAS World Languages Institute  
Mercator Research Centre







# **FOUNDATION FOR ENDANGERED LANGUAGES**

in association with

**Interdisciplinary Centre for Social and Language  
Documentation (CIDLeS)**

**SOAS World Languages Institute, University of London**

**Mercator Research Centre, Netherlands**

## **FEL XXI**

### **Communities in Control:**

### **Learning tools and strategies for multilingual endangered language communities**

Proceedings of the 21<sup>st</sup> FEL Conference

19 – 21 October 2017

Editors:

Nicholas Ostler, Vera Ferreira and Chris Moseley



**SOAS**  
University of London  
— 100 Years —

World  
Languages  
Institute



Foundation for Endangered Languages



© Foundation for Endangered Languages

[www.ogmios.org](http://www.ogmios.org)

## **FULL CORRECTED VERSION**

Published by

Foundation for Endangered Languages  
129 High Street  
Hungerford RG17 0DL  
England

ISBN 978-0-9560210-9-0

## **FEL XXI – ORGANISING COMMITTEE**

Vera Ferreira (CIDLeS)  
Cor van der Meer (Mercator)  
Mandana Seyfeddinipur (SOAS, University of London)  
Nicholas Ostler (FEL)  
Steven Krauwer (CLARIN)

## **Sponsors**



Fundação  
para a Ciência  
e a Tecnologia



GRUTAS DE MIRA DE AIRE



EUROSOL HOTELS

# Digital resources and language use: Expanding the EGIDS scale for language development into the digital domains

## Sebastian Drude

The Vigdís International Centre for Multilingualism  
and Intercultural Understanding  
School of Humanities, University of Iceland  
[sdrude@hi.is]

## Anton Karl Ingason

Faculty of Icelandic and Comparative Cultural Studies  
School of Humanities, University of Iceland  
[antoni@hi.is]

## Ari Páll Kristinsson

Language Planning Department  
The Árni Magnússon Institute for Icelandic Studies  
[aripk@hi.is]

## Birna Arnbjörnsdóttir

Faculty of Languages and Cultures  
School of Humanities, University of Iceland  
[birnaarn@hi.is]

## Einar Freyr Sigurðsson

Faculty of Icelandic and Comparative Cultural Studies  
School of Humanities, University of Iceland  
[einasig@hi.is]

## Eiríkur Rögnvaldsson

Faculty of Icelandic and Comparative Cultural Studies  
School of Humanities, University of Iceland  
[eirikur@hi.is]

## Iris Edda Nowenstein

Faculty of Icelandic and Comparative Cultural Studies  
School of Humanities, University of Iceland  
[ien1@hi.is]

## Sigríður Sigurjónsdóttir

Faculty of Icelandic and Comparative Cultural Studies  
School of Humanities, University of Iceland  
[siggasig@hi.is]

## Abstract

This paper proposes that the digital domains of language use (DDL) be included in future assessments of language vitality. DDL, including the consumption of online content, engagement with social media and chat which now make an important, and rapidly growing, part of the daily language use in many speech communities. This is true even in communities which were until recently fairly isolated from modern technology. Criteria for determining digital language vitality vary according to the general situation of the language but constitute a separate dimension – languages otherwise in a similar situation can differ considerably in terms of digital vitality. We consider, for instance, advanced language technology and digital content for institutional languages, teaching material and dedicated collaborative resources for vigorous and developing languages, and language documentation material for endangered languages. Crucially, what needs to be assessed is not only quantity, availability, quality, coverage, maturity, sustainability, and adaptability of resources and technology as proposed in earlier frameworks, but also, principally, their usage by the speakers. We find the concept of digital *minoritization* useful to advance the assessment of digital language vitality, as illustrated by the language situation in three West Nordic countries, Iceland, Norway and The Faroe Islands. This is work in progress and is an extension of the MoLiCoDiLaCo project being carried out at the University of Iceland.

## 1. Assessing language vitality / development

Any language policy development, planning for research priorities or language strengthening and maintenance efforts depends on a reliable and comprehensive assessment of the language development<sup>1</sup> of the language(s) in question. The first elaborate conceptual framework was Fishman's *Graded Intergenerational Disruption Scale* (GIDS) (1991). Then, after the call from Krauss (1992) and others, the focus has been on endangered languages. More recently, Lewis and Simons (2010) proposed an adjustment and

expansion of Fishman's 8-level GIDS. Their 13-level EGIDS (*Expanded GIDS*) is arguably the best-known framework for assessing language development, largely because it has been applied to all languages listed in SIL's *Ethnologue* (Simons & Fennig, 2017) since its 17th edition (2014). It also provides means for an informative assessment for languages at all levels of language development, from international (0) to extinct (10).<sup>2</sup>

<sup>1</sup> *Language development* is preferred as a more general term that includes *language vitality* and *language endangerment* but also covers non-endangered languages.

<sup>2</sup> Other proposals have been made, notably by the UNESCO Ad Hoc Expert Group on Endangered Languages (2003), which identified 9 main factors for language endangerment. UNESCO's *Atlas of the World's Languages in Danger* (Moseley, 2010) uses mostly one of these criteria, intergenerational transmission, to arrive at a six-degree scale:

EGIDS offers 13 levels, whose numbering maintains compatibility with Fishman's 8 levels by either adding a letter (6a, 6b, 8a and 8b) or new numbers (0, 9 and 10). The levels and their short labels are the following: 0 International – 1 National – 2 Regional – 3 Trade – 4 Educational – 5 Written – 6a Vigorous – 6b Threatened – 7 Shifting – 8a Moribund – 8b Nearly Extinct – 9 Dormant – 10 Extinct.<sup>3</sup> Often, e.g. in the UNESCO atlas, all languages with an EGIDS level 6a or lower are seen as 'safe'. If 'safe' is defined as 'it is likely that the language will be spoken in a few generations' time', this may be a correct assessment for most languages so classified, but this does not reflect the pressure that even national languages suffer in a diglossic setting, in particular in new domains of language use.

In this paper we argue that new criteria are needed to account for the digital domain of language use, and we make first preliminary suggestions which criteria could be used. These criteria focus on usage more than availability of resources. We also suggest that the main criteria to be examined for assessing digital vitality vary for different groups of languages according to their EGIDS level. We finally illustrate our conception by discussing languages in three West Nordic countries, Iceland, Norway and the Faroe Islands.

—  
The paper reports on work in progress and is an extension of the MoLiCaDiLaCo project ("Modelling the Linguistic Consequences of Digital Language Contact," 2016; Sigurjónsdóttir et al., 2017) which studies digital diglossia, in particular between Icelandic and English.<sup>4</sup>

## 2. Digital language use and language contact

The use of digital forms of communication is rapidly gaining importance in most if not all societies worldwide. Under "digital (forms of) communication" we mean all forms of "computer-mediated communication" in a broad sense, both passive one-way "communication", that is, consuming online content, and active dialogical communication, which can be one-to-one (chatting, e.g., in WhatsApp) or one-to-many (blogging, posting on social media, etc.).

---

safe – vulnerable – definitively endangered – severely endangered – critically endangered – extinct.

<sup>3</sup> For languages in a process of revitalization, the levels 6 to 9 have a slightly different interpretation, indicated by different labels: 6a Vigorous – 6b Re-established – 7 Revitalized – 8a Reawakened – 8b Reintroduced – 9 Rediscovered. We will focus on the primary scale in this paper.

<sup>4</sup> Only towards the conclusion of this paper have we heard of closely related work by C. Soria, V. Quochi and colleagues aiming to establish a Digital Vitality Scale and Indicators. Collaboration is being sought.

In the developed countries, use of these digital forms of communication is growing each year and taking up a considerable part of our daily lives (recent media reports state that US citizens spend five hours daily or more online on mobile devices alone: "U.S. Consumers Time-Spent on Mobile Crosses 5 Hours a Day," n.d.), already surpassing watching television and other forms of activities ("How Much Time ...?," 2016).

The time spent in front of screens (especially on mobile devices) and online (particularly on social media) varies greatly from country to country (Bauer, 2017), different age groups and sectors of the society. In certain regions where many endangered languages are spoken (namely Papua New-Guinea, Central Africa and rural regions of Latin America), internet coverage and access to mobile devices continues to be precarious, but that is often changing quickly. Even in remote indigenous reserves in Brazil, for instance, there are support stations with internet access several hours a day, and many younger indigenous people there have mobile devices with which they go online as soon as they are in areas with access, using their Facebook accounts or Skype, as one of the authors (Drude) observed. (Television, which can be counted as a digital form of communication even if it is analogue, is widespread even in remote areas anyway.)

In all forms of digital communication, passive and (inter)active, language is crucially involved on two main levels: (1) as the language of the digital content that is consumed, exchanged or produced, and (2) as the language of the interface (operational system and specific tools used) that provides access to it. Language use on both levels constitutes major new domains (functional areas), the *domains of digital language use* (DDLUs), which are the domains that are characterized by a certain group of intermediating channels which all are based on digital technology (computers, mobile devices, television sets, ...). One can argue that these domains together actually constitute a separate value on the medial dimension of language variation, alongside with written and oral communication (with both of which it shares many properties). In these DDLUs, language contact, and even stronger, diglossia is bound to exist for most users speaking other mother-tongues than English, which continues to be the language in which most content is available, let alone the user interface of tools, many of which are available in English only. So, the whole non-English-speaking world may be seen as being in a diglossic situation with English in the digital domains of language use.

As these domains of digital language use are gaining weight, the choice of languages in these domains gain importance when it comes to assessing language development. Yet, the current scales do not take any DDLUs into account at all. One purpose of the MoLiCaDiLaCo project is to improve on this situation, aiming at building a scale for *digital* language vitality.

This is work in progress, and it builds partly on previous work on specific aspects of digital language presence.

### 3. Availability of digital resources and actual usage: two different dimensions

There have been earlier attempts at measuring the digital strength of languages. Notably, Krauwer (2003) proposed a matrix of a *Basic LAnguage Resource Kit*, *BLARK*, which lists language resources such as corpora, dictionaries, etc. and language technology, such as speech recognition or parsers, for the oral and written domains. BLARK has been offered on an open platform via ELDA (“BLARK: Basic LAnguage Resource Kit,” n.d.), with some uptake (e.g. Prys, 2006).

Building on this framework, in the META-NET Whitepapers (Uszkoreit & Rehm, 2012), 11 different criteria are applied to assess language technology support for 31 European languages. These criteria are divided into two main classes; (1) Language Technology: Tools, Technologies and Applications; and (2) Language Resources: Resources, Data and Knowledge Bases. The criteria listed under the first main class are (i) Speech recognition, (ii) Speech synthesis, (iii) Grammatical analysis, (iv) Semantic analysis, (v) Text generation, and (vi) Machine translation. Under the second main class are listed (vii) Text corpora, (viii) Speech corpora, (ix) Parallel corpora, (x) Lexical resources, and (xi) Grammars. Importantly, the situation in each of these 11 criteria is then further evaluated according to seven features; (a) Quantity, (b) Availability, (c) Quality, (d) Coverage, (e) Maturity, (f) Sustainability, and (g) Adaptability.

There are also more recent attempts, such as by Thomas (2017), looking at the language interface of a number of common digital tools. Similarly, Hammarström (in prep.) is building an inventory that classifies each language for its “documentational status”, from full grammars to word lists and finally to the simple mention of the language. Also, the OLAC catalogue (Simons & Bird, 2003a, 2003b) lists resources, digital and references, for potentially all known languages.

Crucially, however, usage is not one of the criteria applied in any of these works. We propose that usage be added as a major dimension in the evaluation of language vitality. Even though a certain application or resource exists for a certain language, this is only a necessary condition for that it is actually being used by speakers of the language. This is illustrated by the preliminary results discussed in section 6.1, where a large proportion of Icelanders prefers using English as the operating language for their mobile devices, even though Icelandic interfaces are available.

This preference for English in the digital domain is without a doubt motivated by a series of factors, some of which are independent of availability. Thus, raw availability is not enough, since discoverability and

accessibility might play an important role in addition to the quality (e.g. imperfect translations for versions in less known languages) and quantity (number and selection of materials) of the resources available in a given language. Assessment of digital language vitality must therefore include a comprehensive (quantitative and qualitative) assessment of available resources with a focus on how extensively they are actually used. This is particularly important when assuming input-driven models of language acquisition in the context of language transmission, as the following review on input factors in language learning illustrates.

“The more you hear a language, the better you learn it” (Paradis & Grüter, 2014, p. 1). Experience (linguistic input) guides acquisition and guarantees language transmission. Input and experience affect both monolingual and bilingual children’s language development, with bilingual acquisition being influenced even more by input factors than monolingual acquisition (e.g. Paradis & Grüter, 2014:11). Previous research has found a strong correlation between the rate of time spent interacting in a language and resulting language skills (Hart & Risley, 1995, 1999; Pearson et al., 1997; Thordardottir, 2014; Unsworth, 2014).

The term “input” has been used in a variety of ways in second language acquisition research. Corder (1981) defines *intake* as “what goes in”, not what is “available” for going in of the linguistic stimuli provided by the outside world which he refers to as input. Input has to be comprehensible (Krashen, 1982) or comprehended (Gass, 1997) interesting (or relevant), in sufficient quantity, and authentic (see also Cook, 1991) to lead to language development. To the extent that many languages are in fact losing ground to English, it is important to ask whether the English input available to different language communities has changed and is more optimal now in this sense, than it was some years and decades ago. Has the digital revolution created input which is more relevant to speakers of other languages? Findings of recent studies of the spread of English in Iceland suggest that this may indeed be the case (Arnbjörnsdóttir & Ingvarsdóttir, 2017).

Additionally, the digital revolution has possibly made English input available in more voluminous and comprehensive ways, with an added dimension of relevance and interactivity for younger users than before. If so, the proportion of English increases within the overall relevant input a child receives, to the detriment of that of other languages.

The data being gathered within MoLiCoDiLaCo will allow us to compute indices of input quantity and quality for individual speakers, which can hopefully be correlated with language skills and variables of linguistic change.

Indeed, we hope that this interaction of availability and input quantity and quality might predict usage. In fact,

quantity, quality and usage in a way correspond to Kornai's (2013) signs of imminent language death (loss of function, of prestige, and of competence), building on earlier work (Dorian, 1992; see also Sasse, 1992).

#### **4. Digital minoritization**

The terms “minoritization” and “minoritized languages” have sometimes been used to refer to languages that have been marginalized or even banned in their language community (DePalma, 2015). A minoritized language is usually a minority language in its language community, even though that is not necessarily the case. One example that has been quoted in the literature is Galician, which is reported to be losing ground with respect to Spanish, especially in the cities and among younger speakers (*ibidem*). In times of the digital revolution and globalization, a different kind of minoritization of majority languages might be developing – what has been called “digital minoritization” within the MoLiCoDiLaCo project. This term will be used here to refer to a development where a majority language in a given society loses ground to another language (usually English) in the DDLUs, due to changes in society and technological advances even if some of the most obvious indicators for language vitality and endangerment do not clearly signal imminent language death (see also Deumert, 2014).

Many languages are currently losing ground to the globally dominant English even though they fulfil all requisites for language vitality according to criteria such as the EGIDS and the UNESCO Language Vitality and Endangerment scales. Kornai (2013) claims that only about 5% of the world's languages will survive the digital age, and Uszkoreit & Rehm (2012) and Rehm et al. (2016) that at least 21 of Europe's official languages are in danger of “digital extinction” (see sec. 5.3). Increased exposure to English in new domains and media results in digital minoritization of these languages, but the precise nature of this development remains understudied. What does it mean, for example, if access to a modern technology, like speech recognition, hinges on abandoning the use of the native language in parts of daily life? Is digital minoritization transferred to the native competence through loss of native elements and/or adaptation of elements from English, and/or does it lead to bilingualism, which might or might not be sustainable?

Digital minoritization, where a major part of linguistic input through new domains and media is in English and not in a local/traditional language, plays a substantial role in a prospective revision of vitality assessment methods.

### **5. Criteria for digital language vitality**

#### **5.0 The character of these criteria and their relation to EGIDS**

In this section, we will propose criteria that can be used to assess how ‘well’ a language is doing in respect to digital language use. As said initially, this is based on work in progress; this is a first effort and the criteria are tentative, explorative and rather illustrative than definitive. We welcome feedback and suggestions for additional or refined criteria. The major objective at this point is to make our purpose clear and why we believe that it is significant to evaluate languages by including a digital usage dimension.

In principle, criteria as we propose here are independent of EGIDS; they are rather orthogonal to it. That means, two languages which have generally very similar results when assessed for language development according to EGIDS (or similar systems like the UNESCO degrees) can still be doing very differently in terms of digital language use. However, we believe that there are strong correlations between EGIDS and the many possible criteria for assessing digital language vitality, so that it mostly does not make any sense to look for certain criteria given the respective language belongs to a certain EGIDS class – as an illustration, using extreme cases, the existence or not of a language documentation corpus is not significant in the case of Spanish, nor would anyone expect automatic speech recognition technology for a language with fewer than fifty speakers. That is why we discuss the criteria in the following sections separately for (1) endangered, (2) developing & vital, and (3) institutional languages.

#### **5.1 Criteria mostly relevant for endangered languages (EGIDS 6b and higher)**

For most endangered languages, advanced or even basic language resources and technology such as those listed in BLARK are out of reach: at least we know of no endangered language which has, for instance, acoustic or language models for speech, or sentence synthesis technology. Even if such technology was achievable, investing time and efforts in building them is arguably not the most sensible thing to do to strengthen such languages; other activities arguably have higher priority.

Possibly the most important digital resource for endangered languages are recordings in audio and video of the language in natural use. Such material can be applied in different ways – notably in language teaching and as an alternative to the consumption of mass media in a majority language. WHETHER they are used in such ways, if they exist, is certainly the most important criterion. Even if the language is not spoken any more, the existence of such material organized in language documentation corpora creates a valuable linguistic resource – and again, whether such corpus is consulted,

by researchers or by members of the group which formerly spoke that language, makes a huge difference.

Some other language resources may be of relevance for languages in this group, such as digital dictionaries as they are being developed for a number of smaller and endangered languages, especially if they are used by the speakers for strengthening their language skills, or by foreigners to learn the language.

## **5.2 Criteria mostly relevant for developing and vigorous languages (EGIDS 5+6a)**

Languages that, according to EGIDS, are ‘safe’ for the time being, but not institutional, typically have different criteria as to their digital development than endangered languages on the one hand or institutional languages on the other.

For some stronger languages in this group, some of the “basic language resources” of the BLARK, such as lexical resources or text corpora, may be relevant criteria, while many others are not realistically achievable.

More important for languages of this group may be criteria like the following: Does technology exist that allows one to represent / type the language easily with digital technology. In particular, (1) are all needed symbols represented in UNICODE? (2) Are fonts available that contain the glyphs necessary to represent texts in the language? (3) Are input methods (e.g., keyboard definitions) available that allow typing text in the language efficiently? Again here, the actual usage of these resources in digital communication are arguably more important than the fact that they are available. This leads to criteria such as (4) do speakers write in the language on computers? (5) Do speakers write in the language on mobile devices? (6) Do they chat or use social media in the language? Note that these two groups of criteria are certainly related, but principally independent one from another: even if the supporting resources are not available (negative answers to questions (1) to (3)), speakers can choose to type in the language anyways, using workarounds or simplifications, such as substituting and omitting diacritics. Conversely, even if the resources are available, they may not be used, and speakers may choose not to use other languages instead.

Another set of criteria specifically relevant for languages in this group concerns the (a) existence, (b) creation and (c) usage of online resources such as (1) websites in and about the language, (2) forums, wikis and similar information-sharing platforms in the language, and (3) social media (blogs, Facebook, Instagram, ...) in the language.

## **5.3 Criteria mostly relevant for institutional languages (EGIDS 0–4)**

Institutional languages are not endangered, but they are at least competing, mostly with English, for usage in DDLU. One set of criteria for how “digitally developed” such languages are, concerns primarily language technology.

We earlier referred to the BLARK and META-NET criteria. Looking only at existence (and availability etc., see the seven features (a) to (g) above in section 3) of resources, the META-NET report assessed 31 European languages and then ranked them with respect to language technology support within four main areas; (1) Speech processing, (2) Machine translation, (3) Text analysis, and (4) Speech and text resources. No language is classified as having “excellent support” in any of these fields, and only English has “good support”. Few languages, especially big ones (French, Spanish, German, Italian) have “moderate support”, whereas all the other languages have either “fragmentary support” or “weak/no support”. Four languages (Icelandic, Latvian, Lithuanian, Maltese) are in the lowest category for all of the four abovementioned areas.

These results are telling as they are, even without considering the usage of language resources and technology. Research on the usage of these technologies where they exist is urgently needed. We fear that even languages which are evaluated as relatively strong by META-NET, would score lower when it comes to actual usage.

Another set of criteria that applies for these languages concerns *digital content online* – content in entertainment (e.g., videos on YouTube, Netflix and similar platforms), and informational content (Wikipedia pages, etc.). These can then again be evaluated for the seven features (a) to (g), in section 3, above, and, again importantly, their usage.

Summarizing, we suggest to apply criteria for assessing digital language vitality/development that vary according to the general development level of the languages; but the central point is that most if not all languages are in a situation of diglossia with English in the digital domains of language use (for minority and endangered languages, that may be an *additional* diglossia). We conclude by illustrating this conception commenting on countries in the West Nordic region.

## **6. Digital minoritization: cases from the Western Nordic region**

### **6.1 Iceland**

Icelandic offers an ideal test case for the effect of this novel type of contact with English, because it is the native language of the great majority of citizens in a nation-state, and it has legal status as the official



language of that state. Icelandic is used in government documents and compulsory national-level education and it also has a long literary tradition. Therefore, minoritization of Icelandic cannot be attributed to a politically unfavourable environment. Furthermore, the population of Iceland is only 338,000, and the country is geographically isolated. Throughout its history, Iceland has been a largely homogeneous society, both sociologically and linguistically, even though immigration rates have grown in the last few years.

Recent studies suggest that daily English exposure is extensive in Iceland and Icelanders report high English use and advanced levels of English proficiency (Arnbjörnsdóttir & Ingvarsdóttir, 2017). Concerns about the future of the language and its ability to coexist with global English have been brought up a number of times recently in the media in Iceland and the nature of the reaction across society indicates that Icelanders are not indifferent to the perceived weakened status of their native language. Iceland is large enough to have a vibrant community of native speakers, yet it is also small enough to observe fine-grained nationwide empirical patterns. Thus, Icelandic is the ideal test case for studying digital contact with English in detail, and therefore the MoLiCoDiLaCo project was started.

The first part of this project is an online survey among 5,000 randomly selected native speakers of Icelandic. The preliminary results show for instance that around 75% of speakers from 13–20 years of age could envision using English for the voice command of various devices (Siri etc.). In contrast, only 20% of speakers from 46–60 years of age and only 10% of speakers over 60 years of age could envision using English for these purposes. More than 60% of speakers from 16–20 years of age prefer to have the interface of their devices (computers, smartphones, etc.) in English, whereas only 10% of speakers over 60 years of age say they prefer English (Sigurjónsdóttir et al., 2017). Since the project has just started, it is not known at this time whether Icelandic is already taking early steps towards digital language death, but the results of the project will without doubt contribute to a more informed discussion about the future of Icelandic and relevant maintenance strategies.

In this context, it is relevant to examine the general online access and usage patterns in Iceland. The percentage of internet usage in Iceland is the highest in Europe – 100% in 2017, whereas the average figure for Europe is 80.2% (Internet World Stats, 2017). Also, a recent study showed that 92% of 5–8 year old Icelandic children use computers or tablets, 71% of 2–4 year old children, and 36% of children younger than 2 years old (Ólafsdóttir, 2017). These figures show that Icelanders are progressive and quick to adapt new technologies, which can affect Icelandic in various ways.

## 6.2 Norway

Norway is a very interesting case with respect to digital minoritization. Norwegian has two officially recognized variants, Bokmål and Nynorsk, which enjoy equal status according to law. Around 87% of pupils in Norwegian primary and secondary schools use Bokmål, while 12% use Nynorsk (Statistics Norway, 2011). Out of 430 Norwegian municipalities, 159 use Bokmål in administration, 114 use Nynorsk, while 157 are neutral with respect to language variety. However, around 40% of the population live in municipalities which use Bokmål, whereas only 11% live in municipalities which use Nynorsk (Statistics Norway, 2011). By law, all official application forms must be available in both variants, and at least 25% of all government documents are required to be in Nynorsk.

Given these facts, one would expect the proportion of text in Nynorsk (vs. Bokmål) on the internet to be at least 10–12%. However, by crawling the .no domain, Kornai (2013) found that “the actual proportion of user-generated Nynorsk content is well under 1%”. He concluded that “the Norwegian population has already voted with their blogs and tweets to take only Bokmål with them to the digital age”.

However, the situation appears to be more complicated than Kornai’s study indicates. Other sources claim that on average, around 5% of Norwegian internet pages have been in Nynorsk for the last 20 years or so. According to a count from 2015, 1.6% of Norwegian language pages indexed by Google and 4.6% of pages indexed by Bing were in Nynorsk (Grepstad, 2015). But this is not the whole story since many Norwegians use their own dialect in informal texts such as chats and on social media, and these texts cannot easily be classified as either Bokmål or Nynorsk. In a recent study, 75% of 142 secondary school pupils from the Nynorsk area said that they wrote in their own dialect on Facebook, 90% of them use their dialect on Snapchat, and 100% when writing SMS text messages (Rotevatn, 2014). Obviously, the digital status of Nynorsk merits further research.

## 6.3 Faroe Islands

The situation in the Faroe Islands is quite different from both Iceland and Norway. The number of Faroese speakers residing in the islands is only around 45,000. Even though Faroese is the first language of most islanders, Danish is also widely used and it was not until 1938 that Faroese was accepted as a language of instruction in schools. Faroese didn’t get full authorization as a church language until 1939, and it was only with the autonomy law of 1948 that Faroese finally became the language used in all affairs. Even today, most courses in upper secondary schools are taught in Danish and the majority of programs shown on Faroese television are in Danish or with Danish subtitles.

The percentage of internet users in the islands is very high – 98.3% in 2017 (Internet World Stats, 2017). We do not have any statistics on language use on the internet in the Faroe Islands, but according to our Faroese sources, English is prominent in the digital domain in the Faroe Islands just like in Iceland. We don't either have any information on Faroese speakers' attitudes towards the use of English in their smartphones and computers, or for voice command. Kornai (2013) claims that Faroese could have a better chance of surviving than many languages which have a much higher number of speakers, since Faroese is spoken in a relatively closed community and has a high quality Wikipedia. As pointed out above, however, the mere existence of digital resources does not guarantee their actual use and thus we don't know to what extent they contribute to the language's digital vitality.

However, there is one potentially important distinction between Iceland and the Faroe Islands with respect to the increased use of English. Icelanders are used to being able to use their native language in all circumstances, but English is now depriving Icelandic of some of its domains. In the Faroe Islands, Faroese and Danish have lived side by side for a long time and the Faroese people are quite used to having to use a foreign language in certain areas or in certain contexts. The increased use of English might to some extent be at the expense of Danish, not Faroese. Instead of the native language being replaced by a foreign language, a foreign language which has been in close contact with Faroese for a long time might be in the process of being replaced in part by another foreign language. Thus, the effects of the digital revolution on Faroese might be less drastic than the effects on Icelandic, for instance. Whether this makes a difference in the end remains to be seen. At this point, we can only speculate on this difference, but it obviously merits research.

## 7. Conclusion

In this contribution, we envisage a scale of digital language vitality, which indicates how well languages are doing in the digital domains of language use. Crucially, the actual usage of the language in different digital domains has been added as it is more important to actual language vitality and maintenance than the mere existence of digital resources or technology; the latter is just a necessary, but not a sufficient condition for high digital vitality. Even national languages such as Icelandic or Nynorsk can be digitally weak in their diglossic relation with English, despite national support and a number of important existing online tools.

The actual assessment of digital language is a challenge and depends on data on real usage, which only to a certain degree can be obtained by means of surveys; direct and statistical significant observation is the ideal.

This contribution is a first step in the development of a general scale of digital language vitality and the authors

invite critical and constructive comments and proposals for refinement and improvement. Also, the exact relationship to existing language vitality / development scales such as EGIDS may need more conceptual development and empirical research. Some of this will be pursued within the MoLiCoDiLaCo project in the coming years, focusing on Icelandic.

## Acknowledgements

This research has been partially supported by the Icelandic Research Fund, IRF, grant number 162991-051 allotted to the project "Modeling the Linguistic Consequences of Digital Language Contact", Principal investigators Sigríður Sigurjónsdóttir and Eiríkur Rögnvaldsson.

We also thank the Norwegian Language Council for providing us with current statistics and information on the language situation in Norway.

## References

- Arnbjörnsdóttir, B., & Ingvarsdóttir, H. (2017). *Language development across the life span: The impact of English on education and work in Iceland*. Berlin: Springer.
- Bauer, R. (2017, January 3). Media (R)evolutions: Time spent online continues to rise [Text]. Retrieved September 20, 2017, from <https://blogs.worldbank.org/publicsphere/6-2016-media-revolutions-time-spent-online-continues-rise>
- BLARK: Basic LAnguage Resource Kit. (n.d.). [Web page]. Retrieved September 22, 2017, from <http://www.blark.org/index.php>
- Cook, V. J. (1991). *Second Language Learning and Language Teaching*. Hodder Arnold.
- DePalma, R. (2015). Learning a minoritized language in a majority language context: student agency and the creation of micro-immersion contexts. *International Journal of Bilingual Education and Bilingualism*, 18(4), 426–442. <https://doi.org/10.1080/13670050.2014.916653>
- Deumert, A. (2014). *Sociolinguistics and mobile communication*. Edinburgh: Edinburgh University Press.
- Dorian, N. C. (Ed.). (1992). *Investigating Obsolescence: Studies in Language Contraction and Death*. Cambridge: Cambridge University Press.
- Fishman, J. A. (1991). *Reversing language shift: theoretical and empirical foundations of assistance to threatened languages*. Clevedon; Philadelphia: Multilingual Matters.
- Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Erlbaum.

- Grepstad, O. (2015). *Språkfakta 2015. Ei forteljing om språk i Noreg og verda gjennom 850 tabeller*. Oslo: Nynorsk kultursentrum. Retrieved from [http://www.aasentunet.no/iaa/no/sprakfakta\\_2015/](http://www.aasentunet.no/iaa/no/sprakfakta_2015/)
- Hammarström, H., Forkel, R., Castermans, T., Speckmann, B., Verbeek, K., & Westenberg, M. A. (in prep.). Visualizing Language Endangerment and Language Description Hand in Hand.
- Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brooks.
- Hart, B., & Risley, T. (1999). *The Social World of Children: Learning to Talk*. Baltimore, MD: Brooks.
- How Much Time Do We Spend On Social Media? [Infographic]. (2016, December 15). Retrieved September 20, 2017, from <http://mediakix.com/2016/12/how-much-time-is-spent-on-social-media-lifetime/>
- Internet World Stats. (2017). Europe Internet Usage Stats Facebook Subscribers and Population Statistics. Retrieved November 10, 2017, from <http://www.internetworldstats.com/stats4.htm>
- Kornai, A. (2013). Digital Language Death. *PLOS ONE*, 8(10), e77056. <https://doi.org/10.1371/journal.pone.0077056>
- Krauss, M. (1992). The world's languages in crisis. *Language*, 68(1), 4–10. <https://doi.org/10.1353/lan.1992.0075>
- Krauwert, S. (2003). The basic language resource kit (BLARK) as the first milestone for the language resources roadmap. *Proceedings of SPECOM 2003*, 8–15.
- Lewis, M. P., & Simons, G. F. (2010). Assessing endangerment: expanding Fishman's GIDS. *Revue Roumaine de Linguistique*, 55(2), 103–120.
- Modeling the Linguistic Consequences of Digital Language Contact. (2016, December 15). Retrieved September 25, 2017, from <http://malvis.hi.is/node/2129>
- Moseley, C. (Ed.). (2010). *Atlas of the World's Languages in Danger* (3rd ed.). Paris: Unesco Publishing. Retrieved from <http://www.unesco.org/culture/en/endangeredlanguages/atlas>
- Ólafsdóttir, S. (2017, October 3). Börn á fyrsta ári nota tölvur [The Icelandic National Broadcasting Service, RUV]. Retrieved from <http://www.ruv.is/frett/born-a-fyrsta-ari-nota-tolvur>
- Paradis, J., & Grüter, T. (2014). Introduction. In J. Paradis & T. Grüter (Eds.), *Input and experience in bilingual development* (pp. 1–14). Amsterdam: Benjamins.
- Pearson, B., Fernández, S., Lewedag, V., & Oller, D. K. (1997). Input factors in lexical learning of bilingual infants (ages 10–30 months). *Applied Psycholinguistics*, 18, 41–58.
- Prys, D. (2006). The BLARK Matrix and its relation to the language resources situation for the Celtic languages. *Strategies for Developing Machine Translation for Minority Languages*, 31.
- Rehm, G., Uszkoreit, H., Ananiadou, S., Bel, N., Bielevičienė, A., Borin, L., ... Zabarskaitė, J. (2016). The strategic impact of META-NET on the regional, national and international level. *Language Resources and Evaluation*, 50(2), 351–374. <https://doi.org/10.1007/s10579-015-9333-4>
- Rotevatn, A. G. (2014). *Språk i spagaten. Facebook-språket. Om normert språk og dialekt blant vestlandselevar*. Høgskulen i Volda. Retrieved from <https://brage.bibsys.no/xmlui/bitstream/handle/11250/2357530/Samla%20vedlegg.pdf?sequence=2>
- Sasse, J. (1992). Theory of Language Death. In M. Brenzinger (Ed.), *Language Death, Factual and Theoretical Explorations with Special Reference to East Africa* (pp. 7–30). Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110870602>
- Sigurjónsdóttir, S., Nowenstein, I. E., Björnsdóttir, S. M., Angantýsson, Á., Ingason, A. K., & Rögnvaldsson, E. (2017, September). *Language contact without contact: A nationwide study of digital minoritization*. Presented at the Workshop on Language Variation and Change in Contact Situations at DiGS 17, Stellenbosch, South Africa.
- Simons, G., & Bird, S. (2003a). Building an open language archives community on the OAI foundation. *Library Hi Tech*, 21(2), 210.
- Simons, G., & Bird, S. (2003b). The Open Language Archives Community: An infrastructure for distributed archiving of language resources. *Literary and Linguistic Computing*, 18(2), 117–128.
- Simons, G.F., & Fennig, C.D. (Eds.). (2017). *Ethnologue: Languages of the World* (20th ed.). Dallas, Texas: SIL International. Retrieved from <http://www.ethnologue.com>
- Statistics Norway. (2011). Statistical Yearbook of Norway 2013, Table 236: Municipalities, by language used in administration. Bokmål and nynorsk. Retrieved September 25, 2017, from <https://www.ssb.no/a/english/aarbok/tab/tab-236.html>
- Thomas, A. (2017, July). *Assessing Digital Language Vitality in Digital Support Resources*. Poster presented at the LSA Institute.
- Thordardóttir, E. (2014). The typical development of simultaneous bilinguals: Vocabulary, morphosyntax and language processing in two age groups of Montreal

preschoolers. In J. Paradis & T. Grüter (Eds.), *Input and experience in bilingual development* (pp. 141–160). Amsterdam: Benjamins.

UNESCO Ad Hoc Expert Group on Endangered Languages; Matthias Brenzinger, Arienne M. Dwyer, Tjeerd de Graaf, Colette Grinevald, Michael Krauss, ... Ofelia Zepeda. (2003). *Language vitality and endangerment*. Presented at the International Expert Meeting on the UNESCO Programme “Safeguarding of Endangered Languages,” Paris. Retrieved from [http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/pdf/Language\\_vitality\\_and\\_endangerment\\_EN.pdf](http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/pdf/Language_vitality_and_endangerment_EN.pdf)

Unsworth, S. (2014). Comparing the role of input in bilingual acquisition across domains. In J. Paradis & T. Grüter (Eds.), *Input and experience in bilingual development* (pp. 181–201). Amsterdam: Benjamins.

U.S. Consumers Time-Spent on Mobile Crosses 5 Hours a Day. (n.d.). Retrieved September 20, 2017, from <http://flurrymobile.tumblr.com/post/157921590345/us-consumers-time-spent-on-mobile-crosses-5>

Uszkoreit, H., & Rehm, G. (Eds.). (2012). *White Paper Series*. Retrieved from <http://www.springer.com/series/10412>