Digital Language Diversity: Seeking the Value Proposition

Martin Benjamin
LSIR, École Polytechnique Fédérale de Lausanne
IC, BC 114, Station 14, 1015 Lausanne, Switzerland
E-mail: martin.benjamin@epfl.ch

Abstract

This paper is a response to the CCURL workshop call for discussion about issues pertaining to the creation of an Alliance for Digital Language Diversity. As a global project, Kamusi has been building collaborative relationships with numerous organizations, becoming more familiar than most with global activities and the global funding situation for less-resourced languages. This paper reviews the experiences of many involved with creating or using digital resources for diverse languages, with an analysis of who finds such resources important, who does not, what brings such resources into existence, and what the barriers are to the wider development of inclusive language technology. It is seen that practitioners face obstacles to maximizing the effects of their own work and gaining from the advances of others due to a funding environment that does not recognize the value of linguistic resources for diverse languages, as either a social or economic good. Proposed solutions include the normalization of the expectation that digital services will be available in major local languages, international legal requirements for language provision on par with European regulations, involvement of speaker communities in the guided production of open linguistic resources, and the formation of a research consortium that can together build a common linguistic data infrastructure.

Keywords: multilingual, philanthropy, funding, participation, open data, language infrastructure

1. Introduction

While the availability of digital resources for a myriad of languages might strike language technology professionals as an obvious good, the topic does not even enter the consciousness of most people in the world. Knowledge of and attitudes toward multilingualism in the digital sphere are important, because the existence of technological resources depends on both the demand for them, and enthusiasm for investment in their production. This paper sets as its problematic the diversity of attitudes toward digital language diversity (DLD), and the challenges that these many perspectives pose for those involved in languages and technology to succeed in a less-than-welcoming funding and policy environment.

The paper first investigates the various parties involved in the intersection between technology and languages, asking for whom DLD is a value. Following the assumption that increasing the quantity and quality of digital resources for diverse languages is an important goal, the paper then asks how to demonstrate the value of DLD for both consumers and policy makers. Finally, the paper discusses an initiative to bring together many groups that are working on various aspects of DLD, to harmonize many currently-atomized projects toward the development of a shared linguistic data infrastructure that will be widely recognized as a valuable goal.

The paper is semi-ethnographic, based on hundreds of conversations with people at many levels of involvement with language and technology from dozens of countries. It is intended to provoke discussion about policy, not to present research results regarding any particular language resource or technology, with the aim of contributing toward action that will open digital resources to billions of people who speak diverse languages that currently sit at or outside the margins of technology. Without a concrete plan that can be pursued over the course of ten or twenty years, action will be haphazard and ineffective. Without understanding the motivations and barriers for the people involved, a concrete plan cannot be developed. This paper seeks to address the social considerations of DLD, in order to foster a hospitable environment for it to thrive.

2. Attitudes Toward DLD

When thinking of DLD, it is first important to recognize that people have a great range of involvement with and perspectives on the subject, ranging from passion to indifference to ignorance to hostility. Many types of actors can be identified, with some broad themes emerging that are nevertheless not universal within categories. In this section, the ways people connect to DLD are differentiated, and the considerations of some of the players are noted.

2.1. Speakers of diverse languages. Language diversity in the digital realm is a fuzzy concept. Without arguing which languages sit where on a scale of tools, data, content, and speakers, it can be seen that languages such as English, French, and German have a great many resources, languages such as Polish and Chinese aspire toward the same, languages like Swahili and Vietnamese enjoy some digital presence but do not necessarily see a future with the same ubiquitous lingosystem as exists for the first category, and a great many languages neither have nor expect a notable involvement in the technological realm.

For people toward the top of the scale, all or many of their regular digital interactions can be conducted in a language

they know, even if services such as speech recognition or autocomplete are unreliable. Speakers assume that language technologies are developing, and services will improve over time. Few give deeper thought to language issues, with many holding the implicit assumption that everyone else should have the wherewithal to maneuver through technology in an available language.

Conversely, people who speak languages lower on the scale do not generally expect that they will ever be represented digitally. People with secondary education often do not find this to be an issue, with technology naturally belonging in the same sphere as their school books, in a language they can read well enough for practical purposes. Those who are not literate in a well-endowed language might be aware of the privileges others enjoy, but have just as much expectation of partaking as they would of being chauffeured in a Mercedes they glimpse on the street. Thus, one either does or does not access technology through a language in which it is already well developed, without demanding or conceiving of services in a local language. Most of the perhaps 95% of people in many African countries who are thereby excluded, for example, do not have exposure to the idea that this could change, while the other 5% do not feel the need.

People will use digital tools that make sense to them. For example, Sri Lankans text in the Sinhala language by transliterating to Latin text, because they do not have useable input devices for their script on their mobile phones, but this temporary expedient would certainly be retired if Android had a well-integrated Sinhala keyboard. As WhatsApp spreads across India, a new method of communications is opening up – conversations as recorded voice messages, detached from real time, that require no greater literacy than the ability to turn on a phone and press the record and stop buttons. This technology is language-neutral on the surface, and greatly enhancing to the linguistic diversity of its users, but will never result in immersion in a lingosystem, only passive improvements in the ability to communicate without entering the technological mainstream.

2.2 Researchers. People who work on NLP and HLT are inherently sympathetic to underrepresented languages. However, research remains stacked in favor of English and a select few other languages, as can be seen in conference programs such as those of the Association for Computational Linguistics.[1] By necessity, researchers tend to develop expertise at the intersection of a certain language or set of languages, and a certain technology or set of technologies. The pity is that subject expertise is difficult to transfer to diverse languages – though much research is generalizable and could be shared in principle, opportunities to do so are rare among established research groups, and impossible for languages that have neither funds nor the research teams to pursue them. Instead, researchers do exemplary work developing digital resources in their language communities, such as solid teamwork among the languages of India, that do not resonate to the benefit of local languages elsewhere. Moreover, teams spend inordinate energy reinventing overlapping tools, such as similar software to build Wordnets in different languages. As discussed below, unifying researchers could both strengthen their individual projects beyond their focus of language or topic, and produce an action agenda that promotes DLD as a whole.

2.3. Governments and policy makers. Among people with important positions in governments and international agencies, four major attitudes prevail. First are those who believe that the path of progress lies with the languages at the top of the scale. This is especially the case in the US, where language policy is geared toward assimilation to English, and support for research on languages low on the scale largely falls to a smattering of funds from the National Science Foundation, the National Endowment for the Humanities, and military budget for languages of strategic interest. However, leaders in other countries also express indifference or hostility to their local languages; for example, official Rome scorns Italy's smaller languages, most Colombian authorities turn their backs on indigenous languages, and some people in high positions in India advocate a focus on English. Second are those who recognize the value of diversity, to the extent that they can promote it for the languages within their ambit. Irish, Welsh, Estonian, Icelandic, even the Sami language spoken by 30,000 people in Finland all enjoy the support of their national governments. The European Commission coordinates major activities toward DLD at the top of the scale; however, EC interest beyond their core 24 official member languages is largely restricted to communication with major trading partners. Spain invests heavily in its regional languages, with Catalan one of the better-resourced languages in Europe. Russia takes an active interest in its minority languages, though print resources prevail over digital. Similarly, South Africa devotes considerable resources to its largest eleven languages, but plays almost no role sharing its expertise elsewhere on the continent. Conversely, the African Academy of Languages has the African Union mandate to promote digital resources for the continent, but no mechanisms to do so. UNESCO, with an established program on multilingualism in cyberspace, has the most holistic global view, but no funds to actuate projects. Third are those who see value in major local languages, such as Swahili or Vietnamese, but do not extend their concern to smaller vernacular languages within their countries. Fourth are those who are interested in their nations' mother tongues, but do not see them as candidates for digital inclusion. An interesting example here is Uganda, where the president was actually the active lead author of print dictionaries for Nyakore and Kiga (Museveni et al 2009 and 2012), but where even the national Luganda language remains at the digital periphery. For Africa,

[1] https://www.aclweb.org/website/node/434

support for digital resources for cross-border languages is a stated policy objective of the African Union, but not one that is buttressed with the resources for implementation.

Precious few international cooperation activities include DLD within their scope; for example, Canada's IDRC supported ITC4D in Africa and Asia for a number of years but has now shifted focus, the British Council is making some investments in supporting mother tongue education at the primary level, and the Swiss SDC has voiced concern for the issue. By and large, however, language remains peripheral to the discourse of international development.

2.4. Donors and foundations. People involved in DLD should recognize that most donors do not find language equity to be a value. Funders have their own agendas, such as curing a disease or saving a forest. Language is rarely on their screens, and may be seen as a hindrance. Getting through doors guarded by program officers who do not see language as part of an organization's mission is almost impossible. Endangered languages do get bits of funding for sentimental reasons, but overall, DLD is seen as unimportant esoterica.

Small private donors largely have no knowledge about language issues. Neither do many DLD projects tread the difficult and poorly trodden path of retail fundraising. Americans in particular are not known for their concern about language, except perhaps for heritage communities that maintain a sentimental attachment to their ancestors or homelands, such as Yiddish. Private donors tend to respond to international concerns when there is a crisis, such as an earthquake or hurricane. People will occasionally respond to heart-tugging appeals about specific endangered languages, but (a) there are too many endangered languages and too few individual donors for that to be an effective strategy toward widespread preservation and documentation activities, and (b) thousands of minority languages that are not on the cusp of vanishing are systematically ignored.

For big donors, language has yet to make a mark as an area of concern. Language barely makes a dent in the grants of the Ford Foundation, for example, with $145,000 spent in 2014 and 2015 on research and development for the emerging Sheng language of Kenya, a $190,000 grant for the Hawaiian language, and $150,000 for a multilingual voter registration platform for Nigeria – not half a million dollars, from an $800,000,000 portfolio [2]. The Gates Foundation has even less interest in language; other than support for English, they have since 2013 granted $100,000 to develop local-language health materials in Burkina Faso, $175,000 for professional development for American teachers of foreign languages, and $100,000 to support language learning for the Makah Nation near their Seattle headquarters, with another $386,000 spent on non-English in prior decades, and no way for prospective grantees to get in the door and make the case for supporting digital language diversity as a path toward the foundations goals of overcoming inequity [3]. For the Hewlett Foundation, language funding equates to English[4]. In an analysis of the grants database of the Foundation Directory[5], Jaumont and Klempay (2015) find that 88% of the roughly 4 billion granted by American philanthropies in Africa over a decade from 2003 went to Anglophone countries, almost entirely for programs conducted in English. Lack of concern for local languages can be further observed in eleemosynary institutions in Europe and elsewhere. Understandably, big donors want projects that can make an immediate, visible impact, whereas language projects have intangible results that might not be evident for decades (if there is ever a way to measure the effect that increased knowledge has on a society, beyond saying that X number of people have used Y resource that contains Z elements). Less benevolently, few philanthropies are amenable to the case that DLD is worth even a moment of their consideration, and neither practitioners nor potential beneficiaries are in a position to demand otherwise.

2.5. Business. The common factor that determines whether a business is interested in DLD is the profit motive, but that can take many forms. Businesses that sell language services often appreciate the value of diversity, though most prefer to focus on languages that promise a bigger return on investment. Other businesses need language resources to communicate with workers, suppliers, or customers. For the first, DLD might have immediate profit motive, such as a translation contract, or might have the long term objective of an expanded usership. However, creating resources for a language for in-house use or external communications, beyond localizing certain material into select languages, is beyond the scope of most businesses. Furthermore, translation agencies have a vested interest in keeping data such as translation memories private. Therefore, companies are often eager consumers of HLT, but not active agents of its production.

A few companies have taken a much longer view toward DLD, with no immediate payoff, but potentially long term value to stockholders. The translation services of Google and Microsoft probably bleed money, requiring vast processing power that is not recovered through sales or advertising revenue. However, as global companies, both understood that most of their potential market does not speak English, so it was logical to start offering services in other languages. Speculatively, as the translation services became increasingly popular, they began to generate their own momentum, and their improvement is now tied as much to the corporations' sense of mission as to any financial aims. Certainly, Google and Bing Translate are espoused as general-purpose public services with unspecified social benefits down the line. At the same time,

[2] https://www.fordfoundation.org/work/our-grants/grants-database/grants-all

[3] http://www.gatesfoundation.org/How-We-Work/Quick-Links/Grants-Database, search term = language

[4] http://hewlett.org/grants/search

[5] https://fconline.foundationcenter.org/

Google's attention to localization, as exemplified by the excellent versions of its software in Swahili produced by its Nairobi office, expands their reach to millions of customers who are gaining increasing access to technology, and who do not have technical literacy in English. While Google could be critiqued for a large range of shortcomings in their language offerings and their approach to sharing data, they and a few other forward-facing companies are helping lead the development of linguistic resources for nearly 100 languages, offering proof positive that a good bankroll and a cutting-edge technological back end can advance development for any chosen language.

3. Constraints on DLD Development

The lack of a profit incentive for languages down the scale means that most DLD efforts are promulgated by people with a greater sense of mission than a budget to implement it. SIL, for example, coordinates the work of numerous dedicated field researchers, from freely available FLEx software for gathering lexical data, to the Webonary system for hosting results in a standard, searchable format. Yet, though each project is bilingual with a major contact language, there is no common core of senses that is shared among projects, and thus no way to link the work that is done on one language with the work that is done on any other, nor to deployment within technologies that build upon linked data. This is an example of how collaboration within the Human Languages Project (HLP) discussed below, particularly mapping emerging sense-specific concept sets that can be used across projects, could save a lot of repetition and confusion.

Academic projects, when funded, also produce results that produce problems. First, the projects are limited to the term for which they receive funding, which means that they might not get all the way through to stated aims, or might reach those objectives – development of a prototype, acquisition of a particular amount of data – and then have to stop. Second, electronic resources need a perpetual host, or they disappear, and digital results all too frequently vanish when funding runs out, or the researcher moves to another university and their original server account is deleted. Additionally, many academic projects are not conceived to integrate with wider efforts, for example as data that can be used for downstream applications, or run into insurmountable barriers regarding copyrights or the expense and time needed to share results beyond the articles that describe them.

The recent growth of technological hotbeds in places such as Nairobi and Accra has not resulted in major new resources for the languages of their countries. Bright young techies have little financial incentive to pursue projects for local languages. As with IT professionals everywhere, they take jobs that have a good chance of financial reward. Usually, that means working on business or e-governance projects that do not include language concerns. As an example, the Kenya Revenue Authority's online tax filing service was an expensive investment that employed skilled programmers, but is not available in any Kenyan language[6]. Meanwhile, adventurous entrepreneurs face enough risk launching startups, without venturing into unproven language markets. While one could argue that localized shells and local content could be profitable, for example with an Android action game, that is not an argument that has attracted many risk-takers in Johannesburg or Bangalore.

Wikimedia's forays into diverse languages demonstrate that creating content and data in a language requires more than an open platform. Though they list Wikipedias in nearly 300 languages[7], far fewer have enough articles or information to attract readers or count as original linguistic content. A large percentage of multilingual Wikipedia content is generated by robots, usually stub articles with formulaic translations, such as this random entry describing some asteroid, typical of the Yoruba Wikipedia: "3585 Goshirakawa jẹ́ plánẹ̀tì kékeré ní ibi ìgbàjá ástẹ́rọ̀ìdì"[8]. Wiktionary has similarly established shells for 172 languages[9], but close inspection shows that much of the content for many languages is useless at best. The utility to the speaker communities is therefore minimal, and uptake for most languages negligible. Nevertheless, the existence of workspaces for the languages sends a dangerous signal that the languages are already taken care of, and that the community will take control of its own resources with no need of further external effort or concern.

Despite the existence of diverse Wikimedia shells, as well as free blogging platforms that support any UNICODE script (though without localized interfaces), most individual speakers do not see themselves in a position to do anything about their own languages. Non-specialists cannot take responsibility for difficult infrastructure; few people install their own water pipes, or write their own word processors, and none can take on all the work necessary to create their own lingosystem. Standard users do not control the technology, cannot localize a piece of software, and cannot issue data into the void. While they could in principle add Wiki or blogging content, few know this is even a possibility, and there is no well-trod path that starts and keeps people involved in content or technology creation.

4. Normalization

DLD will not come about of its own accord. There is too much of a gap between the interests in language by the people who create digital resources, and the people who

[6] https://itax.kra.go.ke/KRA-Portal/
[7] https://en.wikipedia.org/wiki/List_of_Wikipedias
[8] https://yo.wikipedia.org/wiki/3585_Goshirakawa
[9] https://meta.wikimedia.org/wiki/Wiktionary#List_of_Wiktionaries

speak diverse languages who are not in positions to effectively demand services. However, no great technological leaps are required to create a full panoply of resources for any given language. The heavy lifting in HLT that is undertaken for languages at the top of the scale can be applied to other languages at relatively high speed and low cost. For instance, speech recognition technology does not need to be invented anew, but rather have existing technology trained with data from a new language. DLD is a matter of the time invested in gathering data, building linguistic models, and creating content. Features inherent to a language, such as diverse writing systems and grammars, are relatively surmountable challenges. However, few people are aware of the pathways, fewer are passionate about the desirability of following them, fewer yet are in a position to work toward implementation, and nobody with money will fund any rigorous effort to address the situation.

What is needed is the normalization of the expectation that each language should have a digital existence. So far, there has been no effort to create public awareness about the possibility for linguistic equity, so people who might wish for good resources think that they are about as likely as their traveling to the moon, and therefore worth about as much time investigating. People who do not know that it is possible to make resources for their language will certainly not demand it. For most people, technology is something that one takes as it comes, without thought of going to the manufacturer and asking for new features. Without economic power to exert, and no political groundswell to demand change, linguistic communities do not even dream of a meaningful presence in the digital sphere.

Beyond the persistent efforts of language technology developers to demonstrate that digital resources can be brought into existence whenever the funds and personnel are available, people interested in DLD can pursue two strategies:

5. Advocacy

The first strategy toward digital inclusion is aggressive advocacy. Ordinary citizens cannot demand language services, but their governments can. However, for governments to make such demands, policy makers need to believe that they are both reasonable and achievable. The case can be made in a few areas, which do not all involve digitization. In most instances, regulations can be adopted directly from existing European directives, both because the wording has been well hammered by lawyers, and because no European country could object to trading partners elsewhere in the world imposing exactly the same linguistic conditions as they demand for themselves.

Language advocates should advance model legislation for interested nations, requiring that corporations provide information and services in major local languages, in areas such as food labels, medicines, aviation, product safety instructions, and any product purchased under government contract.

Aviation is the showcase for decision makers, who tend to fly frequently on international carriers, to recognize the desirability of services in national languages. As Air Canada puts it, "Safety is always our number one concern. For this reason, … earphones are not allowed during critical phases of flight as they would prevent you from hearing safety announcements."[10] Airlines cannot argue with their own insistence that it is essential for passengers to understand instructions from the flight crew, and having those instructions in a language that can be understood by the citizens of the country they are flying to fits directly within that logic. Further, there is almost no additional cost for an airline to train a flight attendant with native language skills; with a short grace period, Air France, British Airways, and other carriers could have speakers of national languages on flights serving their entire route system in a few months.

Government contracts are the next step, with a proven pathway to success. The Brazilian government, for example, will not purchase any product that is not available in Portuguese, so major software manufacturers localize their products to that language without question as part of their normal development cycle. The costs of localization are extremely low versus potential sales to government agencies, whether in software where the purchaser might otherwise be tempted to FOSS solutions, for light bulbs where a competitor could easily claim the market by printing a few extra words on their packaging, or for SUVs where a hundred-page user manual could break the sale of a fleet of expensive vehicles. Of course, if light bulbs are packaged for the government in the local language, the same packaging will make it to ordinary store shelves – which is the wider objective.

EU Regulation (EU) No 1169/2011, on the provision of food information to consumers, states that, where labelling is required, it should be "in a language easily understood by the consumers of the member states where a food is marketed". For medicines, Directive 2001/83/EC of the European Parliament, on the Community code relating to medicinal products for human use, states that, "The package leaflet must be clearly legible in an official language or official languages of the Member State where the medicinal product is placed on the market, as specified, for the purposes of this Directive, by that Member State". Similar regulations exist for medical devices and for other products. It should be beyond dispute that the same rules apply for the first languages of billions of other people around the world.

Beyond software localization, these mandates would promote DLD in two ways. First, many of these

[10] http://gofar.aircanada.com/en/go-far-answers/question/earphones-during-ttl/

requirements will be best provisioned with digital intervention, such as the development of translation systems that can produce results acceptable in a legal context. Second, the growing presence of local languages within national markets will lead to increasing expectations that they will become ubiquitous, including within the digital sphere; if tinned tomato labels can be understood by local purchasers, why not tax filing services or voice commands to a mobile device? Realistically, regulations can only enforce the improvement of resources in select non-European languages, but those few dozen will both satisfy many existing deficiencies, and open the door to DLD for languages even farther down the scale.

6. Production

Where policy makers and the public agree that languages have value, and funds and interest can be mobilized, the expansion of DLD depends on the production of resources within each language. This is not straightforward, because there are many more languages than there are existing advocates, researchers, or business cases.

For an example of production possibilities, I point to the design of the Kamusi Project to enroll speaker communities in the production of data for their own languages. Such data can be used for future technologies, with the goal of digital lingosystems far along the scale; online systems can work for languages with a critical mass of networked speakers, a threshold that has not yet been explored. The systems for community participation have been described elsewhere (Benjamin 2015, Benjamin and Radetzky 2014). In short, games and mobile activities elicit consensus-validated data through targeted microtasks that are designed to be fun and compelling. The tasks are built on premises discussed above, that people do not have the individual ability to develop their own language resources, but will contribute if doing so is easy and well explained, and does not require their own technical or financial investment.

Several incentives are posited to give value to community members to participate in the DLD production process. The first is the creation of resources that can make their own lives easier, for example by producing terms that they see will go directly on product labels for the foods they buy. Second is producing something for their children, including data that can be used in L1 education. Third is producing something for the community; this is expected to be a particularly strong motivation among diasporic populations who wish to give back to their homelands. Fourth are intrinsic rewards, such as pride in seeing one's language grow online, and the recognition within social networks that one is taking an active role in advancing language development. Finally, many people find language play to be inherently enjoyable; people pay for games like Scrabble for languages high on the scale, so there is every reason to suppose that people will enjoy passing time with free games as have never before been available for less resourced languages. Unfortunately, despite extensive development on the back end, the systems have not yet been released publically at the time of writing to test these hypotheses, due to a technical constraint that can be represented thusly in UNICODE: ~~money~~.

Lack of money is the most consistent obstacle for DLD. That is, technology does not present barriers, because most languages can piggyback on prior work for other languages. Nor does DLD necessarily require the agreement of policy makers, although that would help lubricate the finances; as academic researchers and business initiatives show, digital resources can appear whenever someone takes the initiative to create them, regardless of official support. This paper therefore closes by inviting interested readers to participate in an emerging consortium to create a Human Languages Project, along the lines of the Human Genome Project or the Human Brain Project, that unites groups from around the world in the development of tools to produce language data, the development of the data itself for a great diversity of languages, and the development of tools to deploy that data in advanced HLT knowledge and NLP applications. Instead of competing for non-existent funding, banding together within HLP can make the case that digitization of the world's languages is an important and realistic goal, that can be achieved by a network of competent partners with a modicum of philanthropic and intergovernmental support.

7. Conclusions

While those active in producing resources for DLD assume the value to be obvious, the case has not yet been made to the powers of the purse. Researchers can discuss success rates for L1 education (Ouane and Glanz 2011), humanists can wax sentimental about the heritage at the cusp of disappearing in endangered languages (Kornai 2013), and activists can bewail the deep and enduring inequities caused by grossly imbalanced language resources (Osborn 1997). However, for EC funding within Horizon 2020, the argument boils down to one consideration: markets. H2020 calls for cross-lingual data development are entirely focused on "data value chains" of "industrial importance"[11]. Sentiment plays no role. Use value to marginalized people is of no relevance. Self-interest, in terms of European trade and security benefits, are the important features for gaining EC support.

The question of how to gain philanthropic support for DLD, particularly from US foundations, is one for which no answers are evident. No foundation currently expresses DLD as a value, beyond limited support for endangered languages, and they do not entertain proposals that seek to convince them otherwise.

[11] Big Data PPP: cross-sectorial and cross-lingual data integration and experimentation, http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/5093-ict-14-2016-2017.html#

What can be attempted is a united approach by like-minded parties, under the auspices of an organizing framework such as HLP. This is in keeping with the objectives laid out in the "Roadmap toward UNESCO's World Atlas of Languages"[12] to the safeguarding of linguistic diversity through the effective application of ICTs. Languages cannot sell themselves, especially languages with few speakers, that have calculably lower economic value. However, pooling resources can lead to much lower costs per language, creating economies of scale that might just tip the balance toward funding support for languages across the board. Such a consortium could create the enabling environment in which DLD thrives – and is thus offered as the value proposition for funding agencies to create a linguistic data infrastructure for languages at all points along the scale.

8. Bibliographic References

Benjamin, M. 2015. Crowdsourcing Microdata for Cost-Effective and Reliable Lexicography. Proceedings of AsiaLex 2015, Hong Kong

Benjamin, M., and Radetzky, P. 2015. Multilingual Lexicography with a Focus on Less-Resourced Languages: Data Mining, Expert Input, Crowdsourcing, and Gamification. LREC, 2014

Jaumont, F. & Klempay, J. 2015. Measuring the influence of language on grant-making by U.S. foundations in Africa. Reconsidering Development, 4(1), 51-65.

Kornai, A. 2013. Digital Language Death. PLoS ONE 8(10): e77056. doi:10.1371/journal.pone.0077056

Museveni, Y., Muranga, M.J.K., Muhoozi, A., Mushengyezi, A., and Gumoshabe, G. 2009. Runyankore/Rukiga-English Dictionary. Kampala, Uganda: Institute of Languages, Makerere University, 213 pp

Museveni, Y.K., Muranga, M.J.K., Muhoozi, A., and Gumoshabe, G. 2012. Katondoozi y'Orunyankore-Rukiga. Thesaurus of Runyankore-Rukiga. Kampala: Fountain, 504 pp.

Osborn, D. 1997. Ultimate Development Participation: Institutionalizing Indigenous Language Use in Education and Research. 23rd annual Third World Conference, Chicago, Illinois

Ouane, A. and Glanz, C. 2011. Optimising Learning, Education and Publishing in Africa: The Language Factor. UNESCO/ Association for the Development of Education in Africa/ African Development Bank.

[12] http://unesdoc.unesco.org/images/0024/002438/243852e.pdf