



ELCat Open Data: Creating a next generation catalogue of language vitality

GARY HOLTON



This article appears in:

Endangered languages and the land: Mapping landscapes of multilingualism

Proceedings of the 22nd Annual Conference
of the Foundation for Endangered Languages (FEL XXII / 2018)
Vigdís World Language Centre, Reykjavík, 23–25 August 2018

Editors: Sebastian Drude, Nicholas Ostler, Marielle Moser

ISBN: 978-1-9160726-0-2

Cite this article:

Holton, Gary. 2018. ELCat Open Data: Creating a next generation catalogue of language vitality. In S. Drude, N. Ostler & M. Moser (eds.), *Endangered languages and the land: Mapping landscapes of multilingualism*, Proceedings of FEL XXII/2018 (Reykjavík, Iceland), 84–89. London: FEL & EL Publishing.

First published: December 2018

Link to this article: <http://www.elpublishing.org/PID/4013>



This article is published under a Creative Commons License CC-BY-NC (Attribution-NonCommercial). The licence permits users to use, reproduce, disseminate or display the article provided that the author is attributed as the original creator and

that the reuse is restricted to non-commercial purposes i.e. research or educational use. See <http://creativecommons.org/licenses/by-nc/4.0/>

Foundation for Endangered Languages:
EL Publishing:

<http://www.ogmios.org>
<http://www.elpublishing.org>

ELCat Open Data: Creating a next generation catalogue of language vitality

Gary Holton

University of Hawai‘i at Mānoa
1890 East-West Rd., Honolulu, HI 96822, USA
[\[holton@hawaii.edu\]](mailto:holton@hawaii.edu)

Abstract

The Catalogue of Endangered Languages (ELCat) launched in 2012 as a collaborative effort to provide verifiable and reliable information on language vitality, providing citations to multiple sources, allowing users to make their own assessments based on the literature. ELCat incorporates its own vitality metric, the Language Endangerment Index but users are free to make their own judgements based on primary data. This paper reviews some of the features of ELCat, acknowledging several shortcomings and outlining plans for addressing those shortcomings in the next project phase, dubbed ELCat Open Data. The most notable changes proposed include providing easier access to download and deposit primary data; adding a time component so that changes in vitality can be tracked; and tracking non-endangered as well as endangered languages. Taken together these changes will provide a foundation for scholars to develop better theories of language endangerment and vitality. As a component of the Endangered Languages Project website the enhanced ELCat Open Data project will serve as a resource for the public, scholars, individuals, language communities, and funding agencies, helping these entities to make the best decisions in how to deploy limited resources most effectively.

Background

The endangered languages crisis is widely acknowledged as one of the most pressing problems facing humanity today, posing moral, practical, and scientific issues of enormous proportions. Assessment of language vitality is an important and necessary first step toward addressing this crisis and reversing language shift (Fishman 1991). The Catalogue of Endangered Languages (ELCat) is a freely available crowd-sourced platform which addresses this need by providing information on the world’s endangered languages. The Catalogue provides authoritative information on language status and vitality; informs users about the plight of endangered languages; and encourages efforts to slow the loss. The Catalogue is available to the public via the Endangered Languages Project (ELP), a website dedicated to promoting endangered languages (endangeredlanguages.com). ELCat is a cost-free resource for the public, scholars, individuals, language communities, and funding agencies to help them make the best decisions in how to deploy limited resources most effectively (Hauk & Heaton 2017). The website serves both as a resource for information on the languages of the world, and as an arena to collaborate with others working to document, revitalize, and promote endangered and minority languages.

ELCat was created as an outgrowth of the 2009 Endangered Languages Information and Infrastructure Workshop, an NSF-funded initiative organized by Lyle Campbell, Helen Aristar-Dry, and Anthony Aristar which brought together leading researchers and key stakeholders from across the US and the world. The recommendations from this workshop formed the basis

for the design and operation of the Catalogue, which was further developed with funding from NSF (grants BCS-1058096 and BCS-1057725), the Luce Foundation, Google.org, and the University of Hawai‘i at Mānoa.

Since its launch in 2012 the ELCat user base has grown enormously. In November 2017 ELCat website visitors initiated over 27,000 sessions totaling nearly 700,000 page views. A unique feature of the Catalogue is the ability for site users to upload language “samples”—text, image, audio, or video resources which illustrate the language in use. Site users are increasingly creating content specifically for upload to the site, such as annotated descriptions of how to make baskets in Xaad Kil (North America) and transcribed recordings of traditional stories in Galoli (East Timor).¹

ELCat has a unique three-tiered governance structure which includes representatives of academic, corporate, and non-profit entities. Though based at the University of Hawai‘i at Mānoa (UHM), the project is overseen by an International Board of Directors consisting of representatives from linguistic regions across the world. ELCat is embedded within the Endangered Languages Project (ELP), a non-profit entity which is overseen by the Alliance for Linguistic Diversity and has its own Governance Council.

ELCat now includes nearly 7000 language samples, and this number is growing daily. Users regularly contact the ELCat team with requests to add languages to the database so that they can upload language samples. The following request from a Sardinian speaker, received Jan 2, 2018, is typical.

¹ www.endangeredlanguages.com/lang/2050/samples/7131;
www.endangeredlanguages.com/lang/9169/samples/7022

I've read that C`ademia Siciliana with your help managed to add sicilian among YouTube's subtitled languages ... and since we need to have sardinian too among these languages, I thought to join the Endangered Languages Project (I've created my account) and to ask you which is the path to follow. Can you help us?

In March 2015 ELCat hosted an NSF-funded workshop which brought together the ELCat International Board of Directors to discuss plans for the long-term sustainable maintenance of the project. A key recommendation of this workshop was the consolidation of project information technology infrastructure at UHM under the direction of the Department of Linguistics (Campbell 2015). In 2016 all web infrastructure for ELP, including ELCat, was migrated from Google servers to UHM. This migration required a significant amount of recoding of the site, due to: (i) deferred maintenance to upgrade the Django web framework and associated packages to the latest stable release version; and (ii) the need to rewrite proprietary API's which were only supported by the Google server (particularly the, Google App Engine). Much of this recoding was undertaken on a pro-bono basis by the Information Technology Service Cyberinfrastructure (CI) group at UHM. Migration was completed on December 12, 2016, and the site has been entirely maintained and served by UHM since that date. Following the web site migration ELCat hired a programmer to interface with UH Cyberinfrastructure group and continue site maintenance. Over the past year and a half we have been able make many improvements to site usability and repair many of the broken features identified by participants in the 2015 ELCat Sustainability Workshop.

Maintaining a modern web-based resource such as ELCat requires significant investment and presents numerous challenges. In the remainder of this paper we describe some of the remaining challenges and discuss plans to overcome them, focusing in particular on increased interoperability and data transparency. Before proceeding, in the following section we discuss some of the issues surrounding the measurement of language vitality.

Assessing language vitality

ELCat implements a quantitative metric of language vitality known as the Language Endangerment Index, or LEI. The LEI combines four factors: intergenerational transmission, number of speakers, speaker number trends, and domains of use. Each factor is assessed on a six-point scale; these values are then combined linearly to obtain an LEI score, with double weight applied to the first factor. Missing factors are not included in the score but are instead used to calculate a level of certainty for the LEI score. The resulting LEI score may range from 0 to 100%, with associated certainty of 0 to 100%, depending on the number of available factors.

In ELCat the LEI is calculated on a source by source basis. That is, a particular language may have several LEI

scores, each drawing on different sources. For example, Tanacross (ISO 639-3 tcb) is assessed as "severely endangered" (LEI = 61-80%) with 20% certainty based on Rogers et al. (2010), but as "critically endangered" (LEI > 80%) with 100% certainty based on Arnold et al. (2009). The difference in certainty for the LEI scores stems from the fact that the first source reports only one LEI factor, number of speakers, while the second source reports all four factors. ELCat also contains information on five additional sources for Tanacross language vitality.

Users are free to submit additional sources on language vitality through the web interface, provided a resolvable citation is supplied. Where the authenticity of a source is in question, the source is vetted by the ELCat Regional Directors.

Like any quantitative measure of language vitality, the LEI is inherently reductionist and thus fails to capture the full range issues affecting language shift and language reclamation. As Grenoble puts it: "All the measurements of vitality are quantitative and reduce complex social dynamics to a single number or set of numbers" (2016: 293). On this point the ELCat team and the Alliance for Linguistic Diversity very much agree. The numbers provided by ELCat should be seen as a first step, a kind of triage for understanding language vitality. Uniquely, ELCat does not purport to provide a single version of the "truth" about language vitality. Rather, it attempts to provide access to all known sources about language vitality, and it invites users to add additional sources. Moreover, ELCat is embedded within the ELP platform which allows for sharing of additional resources relating to language vitality, such as language learning videos.

In sum, while we acknowledge the limitations of quantitative assessments of language vitality, we believe many of these limitations can be at least partially overcome by addressing the recommendations of the 2015 ELCat Sustainability Workshop, as discussed below.

Remaining challenges for ELCat

The ELCat Sustainability Workshop resulted in several important recommendations, including:

- overcoming barriers to interoperability
- harmonizing language identification
- increasing transparency

Below we discuss proposed solutions to these challenges which involve fundamental changes to the operation of the ELCat database. We also discuss an additional issue, namely, the need to expand ELCat coverage to include all languages, not just those deemed by current metrics to be endangered.

Interoperability and language identification

ELCat currently provides the most robust and reliable information on the status and viability of the world's endangered languages. While information on language

status is useful in isolation, the true potential of ELCat lies in the exploration of possible correlations between language status and other factors, both linguistic and non-linguistic. Is there a link between language vitality and culture? Between intergenerational transmission and kinship system? Between language vitality and morphological complexity? Between speaker population and political economy? Investigating these kinds of questions requires interoperability between ELCat and other large data sets. Data interoperability in turn facilitates data exploration, potentially providing answers to questions which have yet to be posed.

An additional advantage to interoperability is that it allows different linguistic databases to focus on their respective strengths. Associated data need not be stored in the database itself but can be connected via links with other databases. ELCat is an authoritative resource for language vitality information. ELCat also includes information about language classification and language location, but it cannot be said to be authoritative in those areas. Maintaining this ancillary, non-authoritative information thus detracts from the core competence of the ELCat project. Improved interoperability would allow ELCat to draw on related databases for this ancillary content rather than maintaining data which is not directly relevant to language vitality.

The current ELCat database presents several barriers to data interoperability. First, the database structure combines both backend information (language vitality and status) and frontend information (website content). This makes it difficult to provide direct database access to users. As a workaround we currently provide users with a spreadsheet (.csv file) representing a simplified version of ELCat data. However, the spreadsheet contains data from only a single (“preferred”) source, rather than from all extant bibliographic sources.

A second barrier to interoperability is the use of ad hoc language codes. ELCat relies on its Advisory Board for language classification information. Where the ELCat classification matches the ISO 639-3 classification, ELCat adopts those 3-letter codes. Where the classification differs, ELCat assigns its own ad hoc numeric code. For example, the language identified in ELCat as Ocotlán corresponds to three different ISO 639-3 codes (zac, zpv, zpn), so it has been assigned an ad hoc ELCat code of 8791.² Many of the ad hoc ELCat codes crosscut the language classifications in other databases, including not only ISO 639-3 but also Glottolog, MultiTree, and others. Hence, there is currently no easy way to map ELCat data to these other linguistic databases—a point to which we return below.

To overcome these barriers and increase interoperability we propose two modifications to the ELCat database. The first modification involves making ELCat compliant

with Cross Linguistic Linked Data (CLLD) database standards. This migration will facilitate interoperability with other CLLD compliant databases, including Glottolog (glottolog.org), D-Place (d-place.org), Glottobank (glottobank.org) and Gelato. As CLLD is increasingly adopted as a standard, the number of interoperable databases is expected to grow considerably. Likewise, the adoption of CLLD by ELCat will help to inspire greater use of this standard. In addition, adoption of CLLD standards will facilitate solutions to transparency, version control, and sustainability.

The second proposed modification is the adoption of Glottolog codes (“glottocodes”) for language classification. As discussed above, ELCat currently employs a mix of ISO 639-3 codes and ad hoc codes. Deferring to Glottolog facilitates interoperability with the many databases using Glottolog codes and removes the burden of maintaining its own language identification code. Cases of disagreement between Glottolog classifications and those advocated by ELCat Advisory Board can be resolved through regular communication with Glottolog staff and through submission of issues to the Glottolog git repository.³ Changes to classifications can be incorporated as part of regular updates to Glottolog. In this way Glottolog also benefits from the expertise of the ELCat Advisory Board, while changes to classifications are rendered in a structured and well-documented fashion.

Deferring to Glottolog for language identification has the additional advantage that changes to language identification and proposed changes can be tracked via the git repository. At present there is no mechanism to track changes to language identification within the ELCat database. That is, if changes are made to the classification—such as the creation of new language codes or the merging of existing codes—these changes are not tracked in the database, and there is no way for users to see what changes have been made.

Increasing transparency of ELCat data

Access to data. ELCat regularly receives requests from the public to access underlying ELCat data. Currently there are two ways to for users to access the data, neither of which provides full access. The first method is via the filters in the website search interface.⁴ Here users may filter by various predetermined categories, such as country, region, speaker numbers, and vitality level. In order to allow users greater access to the data, in 2017 we created a second method which allows users to download a snapshot of the database as a csv file. The process is labor intensive and requires intervention by the web administrator, and as noted above it still does not allow users full access to all of the data. Adopting a CLLD compliant database format will allow users to easily download ELCat data so that they can repurpose it. This

² endangeredlanguages.com/lang/8791

³ github.com/clld/glottolog/issues

⁴ www.endangeredlanguages.com/lang/search

facility has already been implemented by Glottolog and other CLLD databases. The ability to access underlying data is not only a matter of convenience for users but also key to reproducible research (Berez-Kroeker et al. 2018). While in theory all ELCat data are publicly accessible via the current web interface, in practice it is extremely difficult for users to access data in aggregate.

Version control. A second aspect of the transparency issue is the need to implement a version control system. Currently, there is no easy way to track changes to the database over time. Occasional snapshots of the database are stored locally (but not archived), and portions of the database are now available as spreadsheet downloads (also not archived). But in general, if a language or source is added or changed in the ELCat database, the previous information is not retained. Not only does this make it difficult to keep track of administrative changes, it also deprives researchers of a potentially rich source of data, namely, changes in language vitality through time.

We propose to implement a database update process at regular intervals (annually or biannually). Proposed changes will be logged in a shared Git repository and then written to the database at regular intervals. Each database update will be assigned a version number and archived as a published output of the ELCat project. These archived versions will be freely available for public download.

Tracking non-endangered languages

From its inception ELCat was designed as a catalog of *endangered* languages. In this sense it differs significantly from other resources such as Ethnologue and Glottolog. If a language is not considered endangered, then it is not included in the ELCat database. Limiting the scope to endangered languages has allowed ELCat to concentrate resources where they are needed most and where they are most relevant. However, this approach has always been inherently problematic from a technical standpoint because it relies on a single assessment of vitality, when by design ELCat allows for multiple views and multiple sources on vitality. (Recall the seven different sources for Tanacross noted above.) If two different sources differ in their assessment of vitality for a particular language, then which one should be used to decide whether to include the language in ELCat? This question suggests that in order to understand endangered language vitality we actually need to understand vitality for all languages.

From a technological standpoint the inclusion of all languages in the database is not difficult. The challenge is rather an emotional one. The concept of “endangered” is built in to the name of the project and into the website domain name.⁵ Nevertheless, we see at least three important reasons why a catalogue of endangered

languages like ELCat needs to also track non-endangered languages. We discuss these in turn.

Understanding language endangerment. The first reason to include non-endangered languages is that to be able to distinguish between languages which are endangered and those which are not, it is necessary to know more about the non-endangered ones and what it is that qualifies them as non-endangered (or “safe”). For the 3413 languages currently included in the ELCat database, it is possible for users to retrieve LEI scores based on at least one and often several sources, along with full bibliographic references to the sources. This allows for a certain amount of transparency, in that users of the database can readily see on what basis the language in question was determined to be endangered (and hence included in ELCat). However, if a language is determined not to be endangered, then it is not included in the database and this vitality information is not reported to users. Hence, users have no way of determining on what basis the decision was made to classify the language as “safe.”

For several hundred or so of the world’s major languages the question of endangerment is trivial. These languages are clearly “safe” according to all four factors evaluated by the LEI and would probably be assessed as such by any measure of language vitality. That is, these languages are spoken by all members of the community, have hundreds of thousands of speakers, have stable or increasing speaker populations, and are used in all domains. However, the vast majority of languages not currently included in ELCat are *not* completely “safe” as measured by these factors. Many, if not all, of these languages are likely to become endangered with time. Typically, languages which are borderline endangered lose speakers over time and become more endangered. At some point these languages become endangered and suddenly appear in the ELCat database. This transition from “safe” to “endangered” is a critical point for languages, yet ELCat contains information for only one side of the critical point. This makes it difficult to use ELCat data to understand the transition and to investigate the endangerment process. Likewise, a language undergoing revitalization or reclamation would in theory at some point be removed from the current ELCat database once it achieves “safe” status. Tracking the various vitality factors for all languages contributes to a better understanding of the process of endangerment and language shift.

The problem of negative data. A second, related, problem is that if a language is not found in the ELCat database there is no easy way to determine whether it was: (i) accidentally omitted/overlooked; (ii) determined to be not endangered and therefore excluded; or (iii) identified differently in ELCat and thus subsumed under

⁵ It is worth noting, however, that the ‘Alliance for Linguistic Diversity’, the umbrella organization under which ELCat operates, explicitly avoids the term “endangered” in its name.

a different language. For example, consider the case of Gane (ISO 639-9 gzn), an Austronesian language reported by Ethnologue to have 2900 speakers in 1982. Searching ELCat by either the name Gane or its ISO code returns no results. Should we interpret this to mean that the language is “safe,” or has it simply been overlooked in the database? It could well be that ELCat staff have already evaluated the data on Gane, calculated its LEI score, and determined it to be not endangered (“safe”). However, since there is no entry in ELCat for Gane, we have no way of determining this. Another illustrative example is Sar (Glottocode sar1247). This language lacks an ISO 639-3 code and is instead subsumed under Teiwa (ISO 639-3 twe). ELCat assess Teiwa as “Endangered” but has no references to Sar, either as a distinct language or as a dialect or variety of Teiwa. Hence, there is no way to determine whether Sar was overlooked, determined to be not endangered, or simply classified differently than in Glottolog.

As the old saw goes, “absence of evidence is not evidence of absence,” so the absence of a language from ELCat currently implies no information about the vitality of that language.

Building digital language communities. A third problem with not including non-endangered languages is related to Grenoble’s (2016) point about the complex social dynamics of language vitality. Under the umbrella of the Alliance for Linguistic Diversity, ELCat and the ELP platform were designed to facilitate the development of digital communities for endangered languages:

The mission of the Alliance for Linguistic Diversity is to accelerate, strengthen and catalyze efforts around endangered language documentation, to support communities engaged in protecting and revitalizing their languages, and to raise awareness about ways to address threats to endangered languages.

It was never the intent of the Alliance for Linguistic Diversity to bar participation by language communities simply because they failed to achieve a sufficiently high enough LEI score. As noted in the introduction to this paper, ELCat regularly receives requests from members of minority language communities who would like to make use of the resources on the ELP site but whose language does not technically qualify for inclusion due to its low LEI score. In some cases ELCat staff have included such languages on an ad hoc basis by adding an undocumented category to the LEI score, labelled “at risk.” Generally this has been done in response to specific user requests. For example, ELCat now contains an entry for Ladino (ISO 639-3 lad), a language with over 100,000 speakers. Users have uploaded videos and other Ladino resources to the ELP site.

As currently implemented, a language is considered “at risk” if its LEI score is “safe” (LEI = 0%) but its certainty is less than 100% (i.e., at least one of the four factors is missing data). Though not included in the original conception of LEI (Lee & Van Way 2016), this approach

allows minority communities to participate in ELCat. To date a total of 83 “at risk” languages have been included in ELCat. Including *all* languages in the database—whether currently “safe,” “at risk,” or at some stage of endangerment—will permit more communities to participate directly in ELP without having to go through an ad hoc process to have their language added to the database. Including these larger and possibly less endangered languages will also provide a web-based platform for struggling language communities to combat what Kornai (2013) dubs “digital language death,” threatened by the lack of digital support for all but a handful of the world’s languages. In this way ELP can better fulfill its mission to support communities engaged in maintaining, revitalizing, and reclaiming their languages, regardless of where their languages stand on a particular vitality measurement.

Looking ahead

Implementing the changes described above will not be easy. Maintaining a complex database-driven website requires significant investment, but there are many good reasons to continue to support ELCat. The linguistic community has already devoted enormous intellectual effort to the project since its inception in 2009, and minority language communities across the world have contributed nearly 7000 resources to the site. The changes outlined above will help to ensure a more sustainable future so that ELCat can continue to serve these populations. We call this proposed approach ELCat Open Data, because it focuses critically on data interoperability and data transparency.

The focus on openness is closely tied to open access. ELCat is freely available at no cost to users, without advertising. This is particularly important now that other resources on language vitality, such as Ethnologue, are now behind paywalls. The focus on openness also facilitates reuse, allowing researchers and communities to make greater use of ELCat data.

Expanding ELCat to include all languages will support a more holistic approach to language vitality which moves beyond the rhetoric of endangered languages and takes into account all languages, as argued by Mufwene (2017) in a recent discussion piece. While some of Mufwene’s more controversial claims were countered in the responses to his article, one key point merits further attention: namely, the need to integrate studies of language vitality into the mainstream of linguistics. The rhetoric of language endangerment has been questioned by many authors (cf. Dobrin et al. 2007; Hill 2002), and while vitality remains a useful metric, classification as endangered versus safe has less utility today than it did in an era when there was a need to draw attention to the endangerment crisis. There is a great need to understand language attitudes and language vitality across all communities, not just those (currently) deemed endangered.

Hence, the renewed focus on openness inherent in the ELCat Open Data initiative should also be interpreted as

a continued call for community engagement. ELCat represents a unique collaborative research effort to create critical research infrastructure for advancing the science of linguistics, focusing on transparency, open access, interoperability, broad expertise, and community engagement. As a collaborative resource, the value of ELCat lies not so much in the contents of its database or in the expertise of its advisors; rather, the value of ELCat derives from the strength of its user community. As we move forward with ELCat Open Data, we look forward to continuing that engagement, and we welcome proposals for new ways to expand and improve ELCat through collaboration.

Acknowledgements

We gratefully acknowledge the contributions of all of the people who have engaged with ELCat and the Endangered Languages Project by providing information on language vitality and contributing language samples. This project would not have been possible without the generous financial support of the National Science Foundation (grant BCS-1446383), the First Peoples' Cultural Council, the Henry Luce Foundation, and Google.org. The author would especially like to thank Lyle Campbell, Anna Belew, and Shaylene Boechler for sharing their knowledge of the history of ELCat and ELP and their vision for the project's future.

References

- Arnold, Irene, Richard Thoman & Gary Holton. 2009. *Tanacross Learners' Dictionary: Dihtâad Xt'een In Andëg Dinahtl'âa'*. Fairbanks: Alaska Native Language Center.
- Berez-Kroeker, Andrea L., Lauren Gawne, Susan Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, David Beaver, Shobhana Chelliah, Stanley Dubinsky, Richard Meier, Nicholas Thieberger, Keren Rice & Anthony Woodbury. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics* 57(1). 1–18. DOI: doi.org/10.1515/ling-2017-0032
- Campbell, Lyle. 2015. Report of the Workshop: Sustaining the Catalogue of Endangered Languages (ELCat).
- Dobrin, Lise M., Peter Austin & David Nathan. 2007. Dying to be counted: the commodification of endangered languages in documentary linguistics. In Peter Austin, Oliver Bond, & David Nathan (Eds.), *Language Documentation & Linguistic Theory* 6, 59–68. London: SOAS.
- Fishman, Joshua A. 1991. *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Clevedon: Multilingual Matters.
- Grenoble, Lenore. 2016. A response to 'Assessing levels of endangerment in the Catalogue of Endangered Languages (ELCat) using the Language Endangerment Index (LEI)', by Nala Huiying Lee & John Van Way. *Language in Society* 45(2). 293–300.
- Hauk, Bryn & Raina Heaton. 2017. Triage: Setting priorities for endangered language research. In Lyle Campbell & Anna Belew (Eds.), *Cataloguing the World's Endangered Languages*. London: Routledge.
- Hill, Jane H. 2002. "Expert rhetorics" in advocacy for endangered languages: Who is listening, and what do they hear? *Journal of Linguistic Anthropology* 12(2). 119–133.
- Kornai, András. 2013. Digital language death. *PLoS ONE* 8(10). e77056. DOI: doi.org/10.1371/journal.pone.0077056
- Lee, Nala Huiying & John Van Way. 2016. Assessing levels of endangerment in the Catalogue of Endangered Languages (ELCat) using the Language Endangerment Index (LEI). *Language in Society* 45(02). 271–292.
- Mufwene, Saliko S. 2017. Language vitality: The weak theoretical underpinnings of what can be an exciting research area. *Language* 93(4). e202–e223.
- Rogers, Chris, Naomi Palosaari & Lyle Campbell. 2010. Endangered languages of the United States. In Chris Moseley (Ed.), *Atlas of the World's Languages in Danger*, 108–130. Geneva: UNESCO Publishing.