

DISCUSSION

Assessing levels of endangerment in the Catalogue of Endangered Languages (ELCat) using the Language Endangerment Index (LEI)

N A L A H U I Y I N G L E E

*National University of Singapore Department of English Language and
Literature Faculty of Arts & Social Sciences Blk AS5, 7 Arts Link,
Singapore 117570
nala.lee@gmail.com*

J O H N V A N W A Y

*University of Hawai'i at Mānoa
jvanway@hawaii.edu*

ABSTRACT

The Catalogue of Endangered Languages (ELCat) is the central feature of the Google-powered Endangered Languages Project (endangeredlanguages.com), which is a venue for sharing information and resources on the world's endangered languages and the knowledge contained in them. One key feature of ELCat is a quantitative measure that can be used to understand the level of endangerment of any language. Quantitative measurements are needed to compare language vitality across a variety of contexts around the globe, and can be used as a parallel to measurements of other forms of threats to biocultural diversity. This article addresses the development of the Language Endangerment Index (LEI). Based on four factors (intergenerational transmission, absolute number of speakers, speaker number trends, and domains of use), this index is different from other methods of assessment in several ways, especially as it can be used even if limited information is available. (Language endangerment, vitality, intergenerational transmission, speaker numbers, domains of language use)*

INTRODUCTION

The Endangered Languages Project (ELP, endangeredlanguages.com) is a Google-powered platform that seeks to be a hub for information on all of the world's endangered languages and to raise awareness about language endangerment. Predicted rates for language extinction vary. They range from the worst-case scenario,

where ninety-five percent of all the world's languages will be extinct or doomed within 100 years (Krauss 2007) to less catastrophic rates at which one language dies about every three months (Campbell, Lee, Okura, Simpson, Ueki, & Van Way 2013). Regardless of how precise these numbers are, language loss is highly consequential for at least five compelling reasons: (i) the loss of cultural or ethnic identity (Tsunoda 2005); (ii) the loss of knowledge of prehistory by losing the only means of reconstructing words about a culture's past (Evans 2010); (iii) the loss of linguistic diversity (Hale 1992); (iv) the loss of part of the sum of human knowledge (including traditional ecological knowledge) (Crystal 2000); and (v) the loss of languages themselves, which compromises linguists' ability to understand the full range of what is possible in human language and cognition.

ELP features the Catalogue of Endangered Languages (ELCat), which aims to provide reliable and up-to-date information on the endangered languages of the world, including the extent of endangerment for every language listed in the Catalogue, as well as resources for community members, scholars, language activists, and the general public. Considering this, we discuss one of the most important questions facing endangered languages, a question that was key in ELCat's implementation—How should the level of endangerment of any language be assessed? For those concerned with preserving the world's fragile linguistic diversity, it is desirable to be able to quantify language vitality. Quantitative measures of a language's vitality have at least five advantages: (i) they allow a lay audience the ability to understand the scope of the problem of language endangerment from a bird's-eye view; (ii) they raise awareness of language endangerment in communities where it is an issue; (iii) they give researchers and language activists a tool to assess the urgency of attention needed for certain languages; (iv) they enable language vitality assessments to be compared across languages in different contexts and, in that way, allow funding agencies to compare and assess both the urgency and likelihood of success for proposals to document and/or conserve endangered languages; and (v) they allow better understanding of whether or not linguistic diversity can be correlated with biodiversity.

In recent years, the observed correlation between linguistic diversity and biodiversity has garnered widespread attention across various fields of study. That linguistic diversity and biodiversity are linked in some form or other has been proposed by Sutherland (2003), Maffi (2005), and Gorenflo, Romaine, Mittermeier, & Walker-Painemilla (2012). Yet, the notion of biocultural diversity has also been questioned by researchers such as Haspelmath (2012), who wonders if this correlation is spurious, since no clear explanations have been provided for it. Appropriate quantitative measures such as the system proposed here can shed light on this issue if they are used in parallel with measurements of other forms of threat to biocultural diversity. Great impetus exists, therefore, for developing and refining a quantitative system that can be used to assess language endangerment, and that can be understood and implemented easily by other researchers and community members.

Drawing on previous research, ELCat has developed the Language Endangerment Index (LEI), which can be used to assess the level of endangerment of any language of the world, regardless of how much or how little information is actually available on it. This feature makes the LEI different from any of its predecessors.

OTHER METHODS OF ASSESSING LANGUAGE VITALITY

The need for a good method of assessing language vitality is well recognized. The LEI is not the first of its kind. It has as its predecessors such tools as the eight-level Graded Intergenerational Disruption Scale or GIDS (Fishman 1991), UNESCO's nine factors assessing language vitality (UNESCO 2003), Krauss' schema for assessing language viability (Krauss 2007), and the Expanded Graded Intergenerational Disruption Scale or EGIDS (Lewis & Simons 2010). These methods are laudable and each has certain advantages, but they all also have disadvantages that make them unsuitable for use on ELCat. The highly regarded GIDS (Fishman 1991) was one of the very first scales devised for the purposes of assessing language endangerment. It emphasizes intergenerational transmission, and identifies different domains of language use. There are eight levels representing different stages of disruption to the domains of language use. The safest language at level one would be used in most social domains, including education, work, mass media, and government at the national level, while the language most susceptible to loss would be a level eight language, which has lost most domains of use and is maintained only among members of the grandparent generation. Language shift is assumed to have gradually taken place across the different domains, to the extent that a language may possibly die out.

UNESCO uses a different method of assessing language vitality (UNESCO 2003). This method was used to assess the statuses of different languages in the 2010 edition of *Atlas of the world's languages in danger* (Moseley 2010). The language vitality index is assessed through a combination of nine factors (see Table 1). All factors are taken to be equally important and each factor is rated on a scale of zero (worst-case scenario) to five (best possible situation). UNESCO states that these nine factors can be applied to different languages to assess the state of a language, and the type of support needed for maintenance, revitalization, perpetuation, and documentation.

A third approach that classifies languages according to their viability is Krauss' (2007) framework, represented in Table 2, adapted from Krauss (2007). Krauss distinguishes between languages that are 'safe', 'endangered' (at varying degrees), and 'extinct'. At polar ends are the 'safe' languages, which have at least a million speakers or are officially supported regional or state languages, and the 'extinct' category, which includes languages with no speakers. The 'endangered' category is the most elaborated, with five divisions of its own; these divisions are distinguished by how much of each generation speaks the language in question. Krauss (2007) states that

TABLE 1. UNESCO's nine factors (adapted from UNESCO 2003).

Factor number	Factor
1	Intergenerational language transmission
2	Absolute number of speakers
3	Proportion of speakers within the total population
4	Trends in existing language domains
5	Response to new domains and media
6	Materials for language education and literacy
7	Governmental and institutional language attitudes and policies, including official status and use
8	Community members' attitudes toward their own language
9	Amount and quality of documentation

TABLE 2. Krauss' (2007) framework for classifying language according to viability.

	Terminology	Designator	Description
	Safe	a+	One million speakers; officially supported regional or state language
Endangered	Stable	a	All speak, children and up
	Instable; eroded	a-	Some children speak, all children speak in some places
	Definitively endangered	b	Spoken only by parental generation and up
	Severely endangered	c	Spoken only by grandparental generation and up
	Critically endangered	d	Spoken only by very few, of great-grandparental generation
	Extinct	e	No speakers

ninety-five percent of languages are probably in this category. As with the other approaches, the concept of language shift is fundamental as the younger generations begin to adopt other languages for communication.

More recently, in response to these other methods of assessing language vitality, mostly GIDS and the UNESCO's framework, Lewis & Simons (2010) devised the EGIDS, which is currently used to estimate the overall development versus endangerment of any language in *Ethnologue* (Lewis, Simons, & Fennig 2013). Instead of the eight levels that GIDS has, EGIDS has thirteen levels, each level corresponding to a particular language label—*international, national, regional, trade, educational, written, vigorous, threatened, shifting, moribund, nearly extinct, dormant, and extinct*. The lower the number the safer the language is, and vice versa.

THE LANGUAGE ENDANGERMENT INDEX

TABLE 3. EGIDS levels, labels, and descriptions (adapted from Lewis & Simons 2010).

Level	Label	Description
0	International	The language is widely used between nations in trade, knowledge exchange, and international policy.
1	National	The language is used in education, work, mass media, and government at the national level.
2	Provincial	The language is used in education, work, mass media, and government within major administrative subdivisions of a nation.
3	Wider communication	The language is used in work and mass media without official status to transcend language differences across a region.
4	Educational	The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education.
5	Developing	The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable.
6a	Vigorous	The language is used for face-to-face communication by all generations and the situation is sustainable.
6b	Threatened	The language is used for face-to-face communication within all generations, but it is losing users.
7	Shifting	The child-bearing generation can use the language among themselves, but it is not being transmitted to children.
8a	Moribund	The only remaining active users of the language are members of the grandparent generation and older.
8b	Nearly extinct	The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language.
9	Dormant	The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency.
10	Extinct	The language is no longer used and no one retains a sense of ethnic identity associated with the language.

Short descriptions and longer ones are provided for each label. The labels and descriptions have been revised and updated since their inception, and Table 3 lists the different EGIDS levels and their short descriptions according to how they appear on the *Ethnologue* website (Lewis et al. 2013) at the time of writing.¹ According to Lewis & Simons (2010:110) then, a ‘threatened’ language would be one that is ‘used for face-to-face communication within all generations, but it is losing users’, while a ‘nearly extinct’ language would be one that is spoken only by ‘members of the grandparent generation or older who have little opportunity to use the language’. Besides assessing the downward trend of language shift, Lewis & Simons (2010) also provides a different set of descriptions for the categories representing levels of revitalization. This makes the EGIDS more comprehensive at both safe and endangered ends of the scale.

All four methods of assessing linguistic vitality are valuable for different reasons. While GIDS is the first framework of its kind, enabling researchers to

approach the issue of language vitality for the purposes of comparison, UNESCO's framework is broader, introducing aspects such as type and quality of available documentation, and it allows users to consider factors individually. Krauss' framework, with more endangered categories, highlights the immensity of the problem at hand, that many languages cannot be considered safe, and EGIDS, as Dwyer (2011) points out, helps yield fast results in assessing language vitality. These different frameworks, however, can still be enhanced.

GIDS has been criticized for not adequately describing all possible statuses of a language, and for being least elaborated at the lowest end of the scale, where disruption is supposedly the greatest (Lewis & Simons 2010). Consequently, EGIDS was designed to address this, but EGIDS has its own limitations. Using EGIDS to assess Wutun (ISO 639-3: wuh),² a Chinese-Tibetan-Mongolic contact language, Dwyer (2011) states that a language can be clearly endangered even if intergenerational transmission remains strong. The language would be labeled 'vigorous' within the EGIDS framework as it corresponds to Lewis & Simon's (2010) description of being used orally by all generations, and being learned at home by all children as their first language, even though there are low speaker numbers, lack of formal support, and ethnic misclassification according to Dwyer (2011). The implication is that any method of assessment should not be solely based on one factor.

On a related note, it is often challenging to put a particular language into a single category that describes multiple factors. For example, Central Okinawan (ISO 639-3: ryu) is rated as level 7 (shifting) in *Ethnologue* (Lewis et al. 2013), which is taken to mean that children are not learning the language, but all older generations are using it. But Read (2011) gives a more nuanced picture that shows that not only are children generally not learning Central Okinawan, but the child-bearing generation are mostly only passive speakers and do not use the language among themselves; this situation would be better described as level 8a (moribund). In addition, the history of government policies discouraging use of the language has led to shrinking domains and, despite having the most speakers of any Ryukyuan language, it has experienced a sharp decline in number of speakers, but these factors are not accounted for in the EGIDS descriptions. Furthermore, there have been some attempts at standardization and some resources are available, including a medical dictionary (Inafuku 1992), and there are revitalization attempts underway such as language classes (Hara 2005), for which the wording in level 5 (developing) would seem to apply. Clearly, a vitality rating system that takes these individual factors into account, rather than lumping them all into one, would be more useful in determining the overall viability of Central Okinawan, as well as that of many other languages.

Krauss' model is a good basic schema to begin with, but it is less comprehensive with regard to individual factors that cause endangerment. Krauss (2007:2) recognizes that 'it remains a major study... to consider factors detracting from language "SAFETY"' (emphasis in the original). As discussed in the following sections, it is

precisely the aim of ELCat to consider these factors when addressing language viability. Evans (2010) also states that a difficulty in applying Krauss' model is that speakers mix words from dominant languages into their own, and although this may appear as language shift, it is not clear if speakers are doing this out of choice or if they do not have sufficient command of their own language. This criticism is basically applicable to all of the models discussed here, since ongoing language shift is always assumed. Thus, in addition to observing which generations are speaking a language, it is important also to examine the domains in which speakers use certain languages.

The UNESCO framework is careful to treat the different factors separately; it is not without flaws, however. First, it is not designed in such a way that one can determine the overall vitality of a language by combining the factors used in UNESCO's framework. Second, the UNESCO framework addresses the type and quality of documentation as being a key factor in language vitality. While an assessment of the type and quality of documentation is doubtlessly important because it helps indicate the potential for revitalization and the urgency of further research, it is not clear that the type and quality of documentation directly affects the vitality of a language. For example, a language such as Latin, which is extremely well documented, nevertheless no longer has native speakers.

None of these four frameworks is suited for use on ELCat, for a shared reason. Many languages and their social conditions are simply not well-documented. Lehmann (1999) suggests that while there is no reliable estimate of the number of languages that have received linguistic description, it is probable that nothing is known about half of the world's languages aside from their names. This may be an overstatement but, by any measure, there is precious little information and insufficient scholarship about a large portion of the world's languages. With these many gaps in information, one would not be able to apply confidently a GIDS, EGIDS, Krauss, or UNESCO descriptor to a particular language. In addition, while the UNESCO framework is broad and its factors comprehensive, it does not give an overall vitality score to the language being assessed, making it difficult to compare accurately across different languages. Although many different factors are featured in the UNESCO framework, its *Atlas of the world's languages in danger* appears mainly to use intergenerational language transmission as the principal indicator of degree of endangerment (Moseley 2010). Table 4 is a replication of a language assessment table that appears on the *Atlas of the world's languages in danger* website at the time of writing.³ Effectively, only the first factor in UNESCO's (2003) list of nine factors appears to be utilized when comparing levels of endangerment across different languages in the online atlas.

Learning from the successes of these previous methods, as well as from their shortcomings, ELCat researchers designed the LEI for the specific needs of the Catalogue and to improve assessment of the world's endangered languages now and in the future. The LEI assesses a language based on four separate factors, through which it is possible to present an overall vitality assessment for a given language,

TABLE 4. *UNESCO's degrees of endangerment based on intergenerational language transmission (adapted from Moseley 2010).*

Degree of endangerment	Intergenerational language transmission
Safe	The language is spoken by all generations; intergenerational transmission is uninterrupted.
Vulnerable	Most children speak the language, but it may be restricted to certain domains (e.g. home).
Definitely endangered	Children no longer learn the language as the mother tongue in the home.
Severely endangered	The language is spoken by grandparents and older generations; while the parent generation may understand it, they do not speak it to children or among themselves.
Critically endangered	The youngest speakers are grandparents and older, and they speak the language partially and infrequently.
Extinct	There are no speakers left.

and it is designed such that an overall score is attainable even in the absence of particular information for certain factors. Accordingly, the LEI also automatically generates the researchers' level of certainty, alongside each language's vitality, based on how many factors were used to determine the assessment. In addition, ELCat treats the amount and quality of documentation as an issue that researchers and the language community should be concerned about, but does not regard it as a component that contributes directly to language vitality. ELCat's companion platform, ELP is a resource for sharing and learning about endangered languages intended for many interested parties; it describes the scope of documentation available for a given language, and draws attention to languages that are underdocumented.

FACTORS

The LEI itself assesses the level of endangerment based on four factors: intergenerational transmission, absolute number of speakers, speaker number trends (whether increasing or decreasing), and domains of use. Each of these four factors is measured on a scale of zero to five, where each number is associated with a particular description. In general, the bigger the number, the more likely that the language being assessed is endangered. Researchers evaluate multiple sources that include information on a language's vitality. For each of these sources, the assessor chooses the best description for the language from those available on each scale. All descriptions are written in clear and straightforward terms, so that these scales can be used by anyone: community members, linguists, or the general public.

ELCat researchers contend that understanding a language's overall viability is only possible through a careful examination of the individual factors responsible

for that language's vitality, and that the four factors identified are knowable and comparable across languages. Some may prefer a more nuanced examination of a language's vitality, with the view that the factors responsible for a language's endangerment are too complex to be compared across languages. Researchers of this view would rally against quantitative measures, stating that quantitative measures can hardly be accurate. Fishman (1991:45) cautions that these researchers may assert that even the counting of individual speakers is complicated. He uses the term *resident* as an example, and questions if this should include seasonal workers, refugees, armed forces, and tourists, demonstrating that different researchers might treat different items on a quantitative list differently (Fishman 1991:46). Note that the LEI attempts to resolve this by providing prior definitions (as Fishman 1991 suggests) for terms used in vitality assessment. For example, the term *speakers*, when used in the LEI, includes native speakers as well as semi-speakers and heritage speakers. Others (such as Krauss 2007) believe that the factors relevant to language endangerment are perhaps so many that they require in-depth, major study. ELCat researchers, while sympathetic to these points of view, maintain that without understanding and investigating fundamental common factors responsible for language endangerment, very little progress will be made in assessing language vitality and, consequently, less can be done to help communities preserve their languages. ELCat strikes a balance between these different perspectives.

The four basic factors that the LEI is composed of are largely universal and comparable, though ELCat also provides more nuanced information not covered by these factors. This may include items such as what other languages are spoken by the speakers of the endangered language, and actual speaker demographics if available. For example, in addition to noting that Taa, a Khoisan language spoken in Botswana and Namibia, has 2,600 speakers, ELCat also cites other relevant information on vitality provided by Brenzinger (2011): a significant number of Taa speakers in Botswana are shifting to Kgalagadi, while those who live on farms or in towns are shifting toward Afrikaans and Nama.⁴ While this additional information does not affect the language's endangerment rating, it is valuable for researchers who are concerned about what the dominant languages in Taa's environment are. Additional information such as what has been provided for Taa creates a more realistic understanding of the language's vitality.

Also in relation to providing a more nuanced picture, information can differ when different sources are used. As a case in point, based on personal communication by Bradley Rentz (2014), ELCat shows that there are fifteen to thirty speakers of Ngatik Men's Creole, which is spoken in Sapuafik, Ponape, and includes information such as the fact that women of the atoll have a passive understanding of the language. ELCat also shows that the numbers differ if one uses information from Wurm (2007), which states that Ngatik Men's Creole has 700 speakers.⁵ Accordingly, one may arrive at different levels of endangerment and certainty, using separate sources, and ELCat includes these distinct sources and the different levels of endangerment and certainty they project if there are indeed differences.

Taking the above issues into consideration, the following subsections provide further information about the use of the individual scales of intergenerational transmission, absolute number of speakers, speaker number trends, and domains of use.

Intergenerational transmission

Just as other frameworks, such as the GIDS and EGIDS, have recognized intergenerational transmission to be the most critical factor in assessing vitality, ELCat also views intergenerational transmission as essential to ensuring language vitality, and this factor carries twice the weight of each of the other three factors in the LEI. Doubling the weight of the intergenerational transmission score reflects the importance of language acquisition for the future survival of a language, because without transmission to younger generations, a language will cease to exist naturally regardless of other factors. There are a couple of arguments, however, against excessive weight being given to this criterion. Natural disasters, pestilence, and genocide could cause a language to become severely endangered, even if it previously had strong transmission, particularly languages with low speaker numbers. For example, Akuntsú (ISO 639-3: aqz), a Tupían language in Brazil, was severely affected, as was its community, by a massacre (genocide) perpetrated by colonizers and settlers in 1996 that left only six monolingual survivors (Aragon 2013). In such instances, extinction of the language, along with its people, may be inevitable. In addition, as Dwyer (2011) mentions, a language with strong intergenerational transfer may be clearly endangered, as in the case of Wutun, because the language is used in limited domains such as in the household and in the village, and has no official status. There is also no orthography or media in the language.

Nevertheless, the importance of intergenerational transmission is irrefutable, for it is certain that a language ultimately faces extinction if younger generations have no knowledge of it. ELCat identifies intergenerational transmission as the most critical factor in assessing level of endangerment. The scale of intergenerational transmission is presented in Table 5. At the two ends are two extreme scenarios. For example, Dusner (ISO 639-3: dsn), an Austronesian language of Indonesia, has very few speakers, all elderly, and it is not learned by the younger generations (Wurm 2007). It can be considered ‘critically endangered’ on the basis of the scale of intergenerational transmission. A ‘safe’ language, by contrast, is spoken by all community members, including children.

Absolute number of speakers

The next factor to consider is a language’s absolute number of speakers. Absolute numbers may include native speakers as well as semi-speakers and heritage speakers, where these numbers are available. For this factor’s scale, a language that is considered to be ‘critically endangered’ would have between one and nine speakers (see Table 6). An example of a ‘critically endangered’ language is Akuntsú, which has only five speakers at the time of research (Aragon 2013). Any language with a

THE LANGUAGE ENDANGERMENT INDEX

TABLE 5. *Scale of intergenerational transmission.*

5	4	3	2	1	0
CRITICALLY ENDANGERED	SEVERELY ENDANGERED	ENDANGERED	THREATENED	VULNERABLE	SAFE
There are only a few elderly speakers.	Many of the grandparent generation speak the language, but the younger people generally do not.	Some adults in the community are speakers, but the language is not spoken by children.	Most adults in the community are speakers, but children generally are not.	Most adults and some children are speakers.	All members of the community, including children, speak the language.

TABLE 6. *Scale of absolute number of speakers.*

5	4	3	2	1	0
CRITICALLY ENDANGERED	SEVERELY ENDANGERED	ENDANGERED	THREATENED	VULNERABLE	SAFE
1–9 speakers	10–99 speakers	100–999 speakers	1000–9999 speakers	10,000–99,999 speakers	≥ 100,000 speakers

small number of speakers is inherently more at risk to the physical factors of language endangerment (e.g. natural disasters, war, disease, etc.) than languages with large speaker numbers. A ‘safe’ language on the LEI scale by this criterion would have at least 100,000 speakers.

One possible criticism is that the top end of this scale is high, making the vulnerable category too wide, as it includes any language that has between 10,000 and 99,999 speakers—historically, many languages have been viable with this many speakers. ELCat researchers are of the opinion, however, that it is better to err on the side of caution when assessing the vitality of a language, since the threat of language extinction is undeniable and the rate of language death is alarmingly fast today, even at the less catastrophic rate of one every three months (Campbell et al. 2013), as opposed to the worst case scenario of ninety-five percent of all the world’s languages becoming extinct or doomed within 100 years (Krauss 2007). In fact, Krauss (2007) states that a very large proportion of what he deems to be ‘safe’ languages are spoken by a million or more, including children, or are languages that have official state or regional support, such as Icelandic or Faroese. As mentioned earlier, Krauss (2007) goes on to state that in all probability, no language with fewer than 10,000 speakers could be classified as ‘safe’. Given this wide net, ELCat deems it reasonable to define a ‘safe’ language on this criterion

as one that has at least 100,000 speakers, thus providing a narrow safe category, and a wider vulnerable category.

Another possible objection to this scale is that the cut-off points for each of the six levels are arbitrary. In fact, cutting off at any number would be arbitrary but, ultimately, necessary. Although there is little difference between ninety-nine and one hundred speakers, or between nine and ten speakers, there is a big difference between five and fifty-five speakers (the medians of the first two levels in the scale, respectively) when considering language vitality. Numerical values are also necessary because a quantitative measure of endangerment based on fuzzy descriptors, such as ‘few speakers remaining’ or ‘many speakers’ would not work as these are relative—‘a few speakers’ in Papua New Guinea would likely not mean the same thing as ‘a few speakers’ in China. The ranges of LEI’s speaker number scale are chosen because they are comparable across levels within this factor, as well as the other LEI factors, and reasonable in their approximate influence on a language’s vitality.

While the absolute number of speakers is not the best (or the only) indicator of language endangerment, in many cases an estimate of speaker numbers is the only kind of vitality information available. For example, the information available at the time of writing for Vishavan (ISO 639-3: vis), a Dravidian language spoken in India, includes only speaker numbers (150) and approximate locations where the language is spoken (Shashi 1994). For this language and many others like it, a method of assessment that does not account for absolute numbers of speakers would be unusable and, therefore, such a system is not suitable for evaluating the vitality of all the world’s languages. GIDS and EGIDS do not measure speaker numbers. While UNESCO’s framework does involve looking at the absolute number of speakers, stating that a small population may be more susceptible to decimation or that it may easily merge with a neighboring group, UNESCO (2003) does not mention what number of speakers should be considered small.

In addition, as earlier mentioned, ELCat differs from other assessment tools by incorporating a scale that measures the level of certainty for all endangerment ratings based on how many factors were used to determine the assessment. If the absolute number of speakers is the only indicator used to determine the level of endangerment for a given language, the assessor and anyone who reads this assessment will know that there is a lower level of certainty about whether the language in question is endangered, or the extent of endangerment when compared to another language whose vitality assessment is based on more than one factor of endangerment. This is further addressed in a later section of this article.

Speaker number trends

When additional information is available, it is inadequate to consider solely the absolute number of speakers in assessing the vitality of a language. With regard to speaker numbers, it is also important to take into consideration speaker number

THE LANGUAGE ENDANGERMENT INDEX

TABLE 7. *Scale of speaker number trends.*

5	4	3	2	1	0
CRITICALLY ENDANGERED	SEVERELY ENDANGERED	ENDANGERED	THREATENED	VULNERABLE	SAFE
A small percentage of the community speaks the language, and speaker numbers are decreasing very rapidly.	Less than half of the community speaks the language, and speaker numbers are decreasing at an accelerated pace.	Only about half of community members speak the language. Speaker numbers are decreasing steadily, but not at an accelerated pace.	A majority of community members speak the language. Speaker numbers are gradually decreasing.	Most members of the community speak the language. Speaker numbers may be decreasing, but very slowly.	Almost all community members speak the language, and speaker numbers are stable or increasing.

trends. The scale of speaker number trends as depicted by Table 7 attempts to capture the direction and the rate of language shift. A language that is ‘critically endangered’ on this scale has a decreasing number of speakers, with many members of the speech community having shifted from the endangered language. Thao (ISO 639-3: ssf), an Austronesian language spoken in Taiwan, is an example of such a language (Blust 2003). There are only fifteen speakers of Thao ‘out of a considerably larger population which claims Thao ancestry’ (Blust 2003:1). The rate of shift away from Thao toward Taiwanese appears to have occurred at a rapid pace. On the opposite end, a language that is considered ‘safe’ on this scale would be one that is spoken by almost all community members. Its numbers are stable, and if there is any language shift, the number of speakers is increasing, with more people shifting towards the language in question, rather than away from it. One of the criticisms of GIDS was that its levels of endangerment are static, and provide no indication of whether people are shifting towards or away from the language, or if numbers are stable (Lewis & Simons 2010). The LEI scale of speaker number trends provides this information missing from other scales, recognizing that a safe language is one in which the change in speaker numbers over time is nonnegative. In other words, people are shifting towards the language in question (i.e. increasing speaker numbers), or at least people are not shifting from it (i.e. stable number of speakers).

Domains of use

The last factor of the LEI is a language’s domains of use. All tools that assess language vitality recognize the importance of taking into account changing trends within the different domains of language use. Domains of language use, a term

popularized by Fishman (1965, 1991) refers to ‘interactions that are rather unambiguously related (topically and situationally) to one or another of the major institutions of society’, and examples of these institutions include government, family, and religion, among others (Fishman 1991:44). While implicit, a cline can be observed across domains when language shift occurs. At a more advanced stage of language shift, when the language is only used in informal domains such as the family, it is assumed that the language is not used in formal domains such as higher education. At the other end of the cline, an asymmetrical trend is also implied on traditional scales of language vitality, such as Fishman’s GIDS (1991) and Lewis & Simon’s EGIDS (2010). If the language is used in higher education, it is usually inferred that the language is also used in informal domains such as the family.

In relation to the more advanced stages of language shift on the cline of domains, Fishman (1987) explains what he calls the *folklorization* of a language where an indigenous language comes to be used primarily in only irrelevant or unimportant domains. Crystal (2000) notes that in some cases, people discover that they have fewer opportunities to use their language because its use in the community is marginalized and not found in official domains, such as in the local offices of civil services and banks. It is also possible for a language to lose all but one domain of use. For example, Ge‘ez (ISO 639-3: gez), also known as Ancient Ethiopic, is a language used only for liturgical purposes (Weninger 2004). It is important to note that the domain of religion is not regarded as an official one, considering that speakers of threatened languages may practise nondominant religions and traditions. This notion follows from Fishman (1991:99), who states that the remaining few speakers of a language would include ‘reciters’, ‘blessers’, ‘cursers’, and ‘prayers’. In other cases, it is common for an endangered language to come to be used in fewer and fewer public domains, until it is only heard in familial, home settings. For example, Takia (ISO 639-3: tbc), an Austronesian language spoken in the Madang Province of Papua New Guinea is used only at home, while its use in other domains such as at the hospital, wharf, and plantation, has been supplanted by Tok Pisin (Tyron 2007). When speakers no longer use the language in many domains, the maintenance of the language becomes increasingly undervalued over time.

As shrinking domains appear to be both a symptom and a cause of language decline, a language rated as ‘safe’ on the scale of domains of use is one that is used in most domains, such as at home, in education, for government and official transactions, and in mass media (see Table 8). It is often the case that speakers value such a language and may promote it, and that education and literacy in the language are valued by most community members. Many large languages, such as English or Mandarin, fit this description. At the other end of the scale are languages that are used in very specific domains, such as in ceremonies, songs, prayers, proverbs, or certain limited domestic activities. An example of a ‘critically endangered’ language on this scale would be Ge‘ez.

THE LANGUAGE ENDANGERMENT INDEX

TABLE 8. *Scale of domains of use.*

5	4	3	2	1	0
CRITICALLY ENDANGERED	SEVERELY ENDANGERED	ENDANGERED	THREATENED	VULNERABLE	SAFE
Used only in a few very specific domains, such as in ceremonies, songs, prayer, proverbs, or certain limited domestic activities.	Used mainly just in the home and/or with family, and may not be the primary language even in these domains for many community members.	Used mainly just in the home and/or with family, but remains the primary language of these domains for many community members.	Used in some nonofficial domains along with other languages, and remains the primary language used in the home for many community members.	Used in most domains except for official ones such as government, mass media, education, etc.	Used in most domains, including official ones such as government, mass media, education, etc.

CALCULATING LEVEL OF ENDANGERMENT AND LEVEL OF CERTAINTY

In addition to giving scores for the individual factors, ELCat aggregates these scores to arrive at an overall level of endangerment, which is useful for quick evaluations and comparisons across languages. First, the scores assigned for each factor are summed, with the score on intergenerational transmission weighted double. The total score is then converted to a percentage of the highest attainable score based on the number of factors used. When all four factors are used, the highest attainable score is 25. If only two factors are used (e.g. absolute number of speakers and domains of use), the highest attainable score is 10. If two factors are used, and one of these is intergenerational transmission (e.g. intergenerational transmission and speaker number trends), the highest attainable score is 15 (intergenerational transmission times two equals 10 points possible, plus 5 points possible for speaker number trends equals 15). The total score (sum of the scores in each factor) is then divided by the highest attainable score (based on the factors used) and converted to a percentage by multiplying by 100. The formula for establishing this aggregate score as a percentage is as follows.

$$\text{Level of endangerment} = \{[(\text{intergenerational transmission score} \times 2) + \text{absolute number of speakers score} + \text{speaker number trends score} + \text{domains of use score}] / \text{total possible score based on number of factors used}\} \times 100$$

The percentage derived from this formula is interpreted using the scale on the left in [Table 9](#). The output of the formula gives two critical pieces of information: (i) the overall endangerment rating derived from the individual factors, and (ii) the level of

TABLE 9. *Language Endangerment Index and levels of certainty.*

LANGUAGE ENDANGERMENT INDEX	LEVEL OF CERTAINTY
100–81% = Critically endangered	25 points possible = 100% certain, based on the evidence available
80–61% = Severely endangered	20 points possible = 80% certain, based on the evidence available
60–41% = Endangered	15 points possible = 60% certain, based on the evidence available
40–21% = Threatened	10 points possible = 40% certain, based on the evidence available
20–1% = Vulnerable	5 points possible = 20% certain, based on the evidence available
0% = Safe	

certainty based on the number of factors known and used in the rating. These two pieces of information are not combined, but considered together by the user when interpreting an overall level of endangerment assigned to a language in ELCat. In Table 9, overall percentages (results of the formula above) in the left column are assigned six discrete levels of endangerment (*critically endangered*, *severely endangered*, *endangered*, *threatened*, *vulnerable*, and *safe*).

The column on the right in Table 9 shows the certainty scale. This represents how certain the assessor can be about an overall endangerment rating based on what factors are used in the assessment. It is a comparison of the number of factors used in the assessment with the total number of factors in the LEI. For example, if a language is assessed on only one factor (e.g. absolute number of speakers), then the points possible would be 5, which would equal only 20% certainty. If that hypothetical language has approximately 25,000 speakers (which would score a 4 out of 5 on absolute number of speakers), it would be rated ‘vulnerable’ overall, but the certainty of that rating (only 20%) is low because it is important to consider the other factors of endangerment as well. In addition, if a separate hypothetical language is assessed based only on intergenerational transmission, the points possible would be 10, as intergenerational transmission is doubly weighted, hence the assessor can be 40% certain about the level of endangerment assigned to the language on the LEI. The certainty rating is important because it allows ELCat users not to conflate uncertainty with endangerment, but to consider them separately.

Table 10 shows the assessments of three example languages using the LEI. These three languages are selected to represent different language groups and geographical areas. To illustrate how the scale works, languages were chosen that had different amounts of vitality information available at the time of research. The three languages assessed are Anmatyerre (ISO 639-3: amx), a Pama-Nyungan language spoken in Central Australia, Muskogee (ISO 639-3: mus), a Muskogean language of North America, and Sentinelese (ISO 639-3: std), an unclassified language of the Andaman Islands.

Anmatyerre, also known as Anmatjirra, Anmatjera, and Unmatjera, has 1,500 native speakers (Caffery 2010). Children learn it as their first language and it is

TABLE 10. Assessment of levels of endangerment of Anmatyerre, Muskogee, and Sentinelese using the LEI.

Language	Factors of endangerment				Level of endangerment (Language Endangerment Index)	Level of certainty
	Intergenerational transmission (weight x 2)	Absolute number of speakers	Speaker number trends	Domains of use		
Anmatyerre	0 SAFE All members of the community, including children, speak the language.	2 THREATENED 1000–9999 speakers	1 VULNERABLE Most members of the community speak the language. Speaker numbers may be decreasing, but very slowly.	0 SAFE Used in most domains, including official ones such as government, mass media, education, etc.	$3/25 \times 100 = 12\%$ VULNERABLE	100% certain (25/25 possible points, based on four factors)
Muskogee	4 SEVERELY ENDANGERED Many of the grandparent generation speak the language, but the younger people generally do not.	2 THREATENED 1000–9999 speakers	5 CRITICALLY ENDANGERED A small percentage of the community speaks the language, and speaker numbers are decreasing very rapidly.		$15/20 \times 100 = 75\%$ SEVERELY ENDANGERED	80% certain (20/25 possible points, based on three factors, including I.T., which is doubly weighted)
Sentinelese		4 SEVERELY ENDANGERED 10–99 speakers			$4/5 \times 100 = 80\%$ SEVERELY ENDANGERED	20% certain (5/25 possible points, based on one factor)

spoken across all generations. Because of their remote location, speakers have been able to retain their specific customs, practices, and language. The main threat to the language lies, however, in governmental policy and the relatively small total number of speakers. Bilingual education was disrupted in 2008 in Australia's Northern Territory; a policy was implemented that stipulates that the first four hours of education in schools are to be conducted fully in English. Caffery (2010) notes this means that children no longer can use their first language as a means to learn to read and write in English. Although this represents a potential threat to Anmatyerre's original domains of use, the language is still considered 'safe' in its domains of use; this would change should the language no longer be used in the official domain of education. Because information is available for all four factors, the language can be assessed using the LEI with the highest level of certainty. Details of how this language is scored on each scale, and the overall levels of endangerment and certainty assigned are presented in Table 10.

Muskogee, also known as Creek, is spoken by less than 6,200 speakers in the United States (Golla, Goddard, Campbell, Mithun, & Mixco 2008). There are between 4,000 to 6,000 people who speak a variety of this language in the former territory of the Muskogee (Creek Nation) and Seminole Nation in east-central Oklahoma, and less than 200 members of Seminole Tribe in Florida who speak another variety (Golla et al. 2008). In 2001, the youngest speaker in Oklahoma was recorded to be eighteen, while the youngest speaker in Florida was forty-four (Golla et al. 2008). On the scale of intergenerational transmission, the language is assessed to be 'severely endangered', since many of the grandparent generation speak the language, while young people generally do not. On the scale of speaker number trends, Muskogee is assessed to be 'critically endangered' because young people are increasingly not speaking the language and speaker numbers are decreasing very rapidly. While Golla and colleagues (2008:13) report that Muskogee is used at church services, for ceremonial speeches, and in hymns, there is little other information regarding in which other domains the language is used. Hence the level of endangerment of Muskogee can only be assessed based on three factors, including intergenerational transmission, absolute number of speakers, and speaker number trends. The overall assessment provided by LEI is 'severely endangered', with less than complete certainty, until more is reported about its domains of use. Details regarding how levels of endangerment and certainty have been calculated are presented in Table 10.

Last of these is Sentinelese (ISO 639-3: std), about which we know very little. It is spoken on North Sentinel Island, which is located among the Andaman Islands. The community of Sentinelese prefer to have very little contact with the outside world. In general, the Sentinelese have been hostile to outsiders who have tried to land on their island (Singh 1978). Because of their isolation, we know very little about the use of this language, so it is not yet possible to rate it on three of the four factors. In 2001, the Census of India surveyed the Sentinelese from a distance, and recorded that there were thirty-nine individuals (India 2001). This

number is lower than the one reported in the 1971 census carried out by the Indian government. It was estimated then that there were approximately 100 speakers/members of the ethnic group (Basu 1991). Because of the circumstances, it is not yet possible to obtain more information about the vitality of Sentinelese. A language that has relatively small speaker numbers, however, is inherently more at risk from natural disasters or epidemics and the encroachment of the larger regional languages than a language with greater numbers of speakers; therefore ELCat gives the rating of ‘severely endangered’ based on the reported absolute number of speakers, with the understanding that this is only 20% certain because information is lacking regarding three of the four factors, as shown in Table 10.

Except for level of certainty, a higher number on any LEI scale indicates a greater likelihood that the language is endangered; a lower number means the language is less endangered. The rationale for assigning a greater number to a situation where the language is more likely to be endangered is straightforward, as ELCat is concerned with highlighting the issue of language endangerment. Rather than being a direct measure of language vitality, ELCat measures level of endangerment, and a larger number would indicate the amount of attention the language requires, in proportion to its level of endangerment.

CONCLUDING REMARKS

This article describes a system for measuring and comparing levels of language endangerment. It was designed and implemented for ELCat, but this system can be applied by others to assess language endangerment in many contexts, and it is our hope that it will become a useful tool in the study of language endangerment. It is different from other methods of assessment, primarily because it assigns overall assessments to languages by combining four separate factors that affect vitality, and it can be used even when there are gaps in the information available, since it is able to give a level of certainty regarding how endangered a language is based on the number of factors used.

There also remains the question of how ELCat deals with languages that are thought to have recently lost their last speaker or those that have experienced a period of broken transmission, but are currently being revived. Because languages are often a vitally important piece of the culture for speakers, rememberers, and their descendants, the loss of a language is a sensitive subject, and a declaration that a language is ‘dead’ or ‘extinct’ can be another source of trauma for peoples who often have already experienced a great deal of upheaval and suffering. Therefore, ELCat utilizes the term *dormant* out of respect to community members, to learners and potential learners of the language, and to passive rememberers of the language in order to encourage revival efforts; labelling a language as *extinct* can discourage revitalization efforts. ELCat does not include languages that clearly and unambiguously have not been spoken since the distant past, for example, Ancient Egyptian and Alanic. It does recognize, however, that

occasionally some languages, thought to have become recently extinct, have had previously unknown speakers turn up—perhaps reminiscent of Mark Twain’s response that reports of his demise had been greatly exaggerated. ELCat thus uses the term *dormant* for any language that is thought to have lost its last remaining speaker recently (within approximately the last fifty years), or for any language where any doubt remains as to the possible existence of any speakers. Such is the case of Tuxá (ISO 639-3: tud), an isolate of Brazil, which was reported to have two speakers in 1961 (Crevels 2012). It is likely that the two have died and that all remaining members of this group have shifted to Brazilian Portuguese at the time of this writing, so it is classified as ‘dormant’.

Related to ‘dormant’ languages, ELCat uses an additional term, *awakening*, for languages once considered ‘dormant’, but where there exists within the community some form of targeted language revitalization undertaken and overseen by a coherent and organized group of interested parties, with the expressed goal of creating new speakers of the language. A good example is Wampanoag (ISO 639-3: wam), an Algonquian language formerly spoken mainly in Massachusetts. Despite being ‘dormant’ for over a century, earnest revitalization is under way and some members of the community are learning Wampanoag as a second language (Fenelon & Hall 2008; Hinton 2011). ELCat recognizes that communities are undertaking efforts to awaken their previously ‘dormant’ heritage languages, and encourages revival efforts of languages with ‘dormant’ and ‘awakening’ assessments.

On a separate note, the LEI is an instrument that might be taken as liberal with regard to what it considers to be an endangered language. Basically, if a language is to score even 1% on the endangerment scale, it is considered ‘vulnerable’. A language cannot be considered safe unless it has a score of zero for each factor and a score of 100% on the certainty level. Hence, a language may be ‘vulnerable’ even if it has more than 100,000 speakers; a language is considered potentially to be ‘at risk’ until information concerning all the remaining factors of endangerment is known. ELCat finds this to be a suitably cautious approach in facing the magnitude of the problems of language endangerment and the alarmingly fast rate at which the world’s linguistic diversity is being lost.

In conclusion, the LEI considers more carefully both the individual factors and how these factors combine to give an overall vitality status for an endangered language than any other approach, and with its level of certainty, does not hide what remains to be known. Therefore, it is hoped that it will prove a valuable tool not just for ELCat, but for the study of language vitality and endangered languages generally.

NOTES

*We would like to thank those who contributed to the development of the LEI, including Lyle Campbell, Eve Okura, Kaori Ueki, and Anna Belew, among others. We also thank the editor of *Language in*

Society, Jenny Cheshire, and the two anonymous reviewers for their valuable suggestions and comments, as well as Lenore Grenoble for her response to our article. Any errors that remain are solely ours.

¹www.ethnologue.com; accessed December 10, 2013.

²ISO 639 is a set of standards emanating from the International Organisation for Standardization that aims to identify all known human languages with a three-letter code. Here, *wuh* refers to Wutun. The registration authority is now SIL International (Summer Institute of Linguists International).

³www.unesco.org/culture/en/endangeredlanguages/atlas; accessed December 26, 2013.

⁴For more information, please refer to <http://www.endangeredlanguages.com/lang/593>.

⁵For more information, please refer to <http://www.endangeredlanguages.com/lang/3054>.

REFERENCES

- Aragon, Carolina (2013). Phonetic realization of medial stop consonants in Akuntsú. *Working Papers in Linguistics* 44(1):1–14. Mānoa: University of Hawai‘i at Mānoa.
- Basu, Arabinda (1991). Small tribal populations: Observed patterns of fertility and mortality. In Baidyanath Saraswati (ed.), *Tribal thought and culture: Essays in honour of Surajit Chandra Sinha*, 169–82. New Delhi: Ashok Kumar Mittal Concept.
- Blust, Robert (2003). *Thao dictionary*. Taipei: Academica Sinica, Institute of Ethnology.
- Brenzinger, Matthias (2011). The twelve modern Khoisan languages. In Alena Witzlack-Makarevich & Martina Ernszt (eds.), *Khoisan languages and linguistics: Proceedings of the 3rd International Symposium*. July 6–10, 2008, Riezlern/Kleinwalsertal, 1–32. Cologne: Rüdiger Köppe.
- Caffery, Josephine (2010). Central Australian endangered languages: So what? *Dialogue (Academy of Social Sciences in Australia)* 29:78–86.
- Campbell, Lyle; Nala Huiying Lee; Eve Okura; Sean Simpson; Kaori Ueki; & John Van Way (2013). New knowledge: Findings from the Catalogue of Endangered Languages (ELCat). Paper presented at the 3rd International Conference on Language Documentation and Conservation. February 28–March 3, 2013, University of Hawai‘i at Mānoa.
- Crevels, Miy (2012). Language endangerment in South America: The clock is ticking. In Lyle Campbell & Veronica Grondona (eds.), *The indigenous languages of South America: A comprehensive guide*, 176–234. Berlin: Mouton de Gruyter.
- Crystal, David (2000). *Language death*. Cambridge: Cambridge University Press.
- Dwyer, Arianne M. (2011). Tools and techniques for endangered-language assessment and revitalization: Vitality and viability of minority languages. *Proceedings of the Trace Foundation Lecture Series*, October 23–24, 2009, New York. Online: https://kuscholarworks.ku.edu/dspace/bitstream/1808/7109/1/Dwyer2011_AssessRevitalize.pdf.
- Evans, Nicholas (2010). *Dying words: Endangered languages and what they have to tell us*. Malden, MA: Wiley-Blackwell.
- Fenelon, James V., & Thomas D. Hall (2008). Revitalization and indigenous resistance to globalization and neoliberalism. *American Behavioral Scientist* 51:1867–901.
- Fishman, Joshua (1965). Who speaks what language to whom and when? *La Linguistique* 1(2):67–88.
- (1987). Language spread and language policy for endangered languages. In Peter H. Lowenburg, *Proceedings of the Georgetown University Round Table on Language and Linguistics*, 1–15. Washington, DC: Georgetown University Press.
- (1991). *Reversing language shift*. Clevedon: Multilingual Matters.
- Golla, Victor; Ives Goddard; Lyle Campbell; Marianne Mithun; & Mauricio Mixco (2008). North America. In Ronald E. Asher & Christopher Moseley (eds.), *Atlas of the world’s languages*, 2nd edn, 7–41. London: Routledge.
- Gorenflo, Larry J.; Suzanne Romaine; Russell A. Mittermeier; & Kristen Walker-Painemilla (2012). Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas. *Proceedings of the National Academy of Sciences of the United States of America* 109(21):8032–37.

- Hale, Kenneth (1992). Language endangerment and the human value of linguistic diversity. *Language* 68:35–42.
- Hara, Kiyoshi (2005). Regional dialect and cultural development in Japan and Europe. *International Journal of the Sociology of Language* 175/176:193–211.
- Haspelmath, Martin (2012). Should linguistic diversity be conserved like biodiversity? Online: <http://dlc.hypotheses.org/195>; accessed December 2014.
- Hinton, Leanne (2011). Language revitalization and language pedagogy: New teaching and learning strategies. *Language and Education. Special Issue: 'Applied linguist needed': Cross-disciplinary networking for revitalization and education in endangered language contexts* 25:307–18.
- Inafuku, Seiki (ed.) (1992). *Igaku Okinawago Jiten [A medical dictionary of Okinawan]*. Ginowan: Roman Shobō Honten.
- India, Office of the Registrar General & Census Commissioner (2001). *Census of India*. New Delhi: Government of India, Ministry of Home Affairs.
- Krauss, Michael (2007). Classification and terminology for degrees of language endangerment. In Matthias Brenzinger (ed.), *Language Diversity Endangered*, 1–8. Berlin Mouton de Gruyter.
- Lehmann, Christian (1999). Documentation of endangered languages: A priority task for linguistics. *ASSIDUE 1: Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt*, 1–15. Erfurt: Universität Erfurt.
- Lewis, M. Paul, & Gary F. Simons (2010). Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique* 55(2):103–20.
- Lewis, M. Paul; Gary F. Simons; & Charles D. Fennig (eds.) (2013). *Ethnologue: Languages of the world*. 17th edn. Dallas, TX: SIL International. Online: <http://www.ethnologue.com>.
- Maffi, Luisa (2005). Linguistic, cultural, and biological diversity. *Annual Review of Anthropology* 34 (1):599–617.
- Moseley, Christopher (ed.) (2010). *Atlas of the world's languages in danger*. Paris: UNESCO Publishing. Online: <http://www.unesco.org/culture/languages-atlas/en/atlasmap.html>.
- Read, Zachary (2011). *Number of speakers of Central Okinawan: A guide to Okinawan, Jlect*. Online: <http://www.jlect.com/downloads/Number-of-Central-Okinawan-Speakers.pdf>.
- Shashi, Shyam Singh (1994). *Encyclopedia of Indian tribes*. New Delhi: Anmol.
- Singh, N. Iqbal (1978). *The Andaman story*. New Delhi: Vikas.
- Sutherland, William J. (2003). Parallel extinction risk and global distribution of languages and species. *Nature* 423:276–79.
- Tsunoda, Tasaku (2005). *Language endangerment and language revitalization: An introduction*. Berlin: Mouton de Gruyter.
- Tyron, Darrell (2007). The languages of the Pacific region: The Austronesian languages of Oceania. In Matthias Brenzinger (ed.), *Language diversity endangered*, 391–409. Berlin: Mouton de Gruyter.
- UNESCO ad hoc expert group on endangered languages (Matthias Brenzinger, Arianne M. Dwyer, Tjeerd de Graaf, Collette Grinevald, Michael Krauss, Osahito Miyaoka, Nicholas Ostler, Osamu Sakiyama, María E. Villalón, Akira Y. Yamamoto, Ofelia Zapeda) (2003). Language vitality and endangerment. Document submitted to the International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages, Paris, 10–12 March 2003. Online: <http://www.unesco.org/culture/ich/doc/src/00120-EN.pdf>.
- Weninger, Stefan (2004). *Ge'ez (Classical Ethiopic)*. Munich: LINCOM Europa.
- Wurm, Stephen A. (2007). Australia and the Pacific. In Christopher Moseley (ed.), *Encyclopedia of the world's endangered languages*, 424–557. London: Routledge.

(Received 16 February 2014; revision received 17 February 2015;
accepted 17 March 2015; final revision received 17 March 2015)