

The Index of Linguistic Diversity: An Overview of a New Measure of Trends in the World's Languages*

The Index of Linguistic Diversity (ILD) is the first-ever quantitative measure of global trends in linguistic diversity. It measures changes in the number of mother-tongue speakers of a globally representative sample of the world's languages. Its objective is to provide solid data that show whether the world's languages (particularly indigenous languages) are losing speakers, and if so at what pace.

The ILD tracks trends in language demographics over the period 1970–2005. The key findings are:

- Globally, linguistic diversity declined 20%.
- The diversity of the world's indigenous languages declined 21%.
- Regionally, indigenous linguistic diversity declined over 60% in the Americas, 30% in the Pacific (including Australia), and almost 20% in Africa.
- The top 16 languages spoken worldwide increased their share of the global population from 45% to 55%.

1. INTRODUCTION

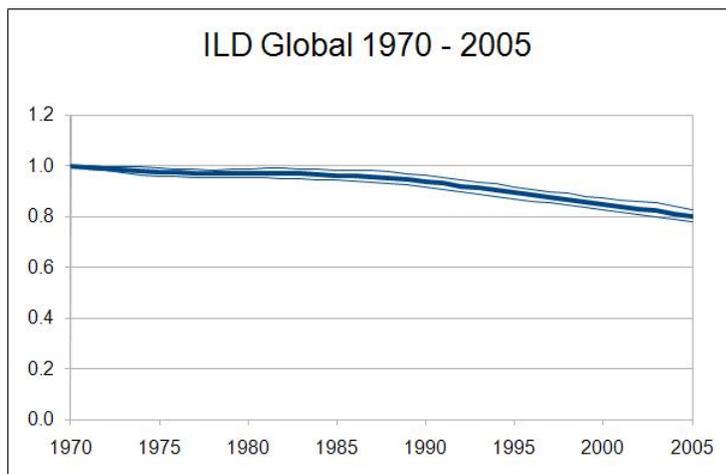


Figure 1: Declining Trend of Global Linguistic Diversity, 1970-2005

Concern about the future of the world's languages has been building for the better part of two decades. A large amount of qualitative evidence points to an impending mass extinction of languages. The quality of this evidence ranges from merely anecdotal to very accurate narrative

accounts based on firsthand knowledge of the language demographics of individual speech communities. It is a highly valuable body of evidence, leaving no room to doubt that the entirety of the world's languages—not just their number, but also the linguistic and cultural diversity they represent—is being severely diminished.

For a host of complex reasons, people are abandoning their mother tongues and switching to other languages, almost always ones with larger numbers of speakers; thereby, more and more people are being concentrated into fewer and fewer languages.

However, there is much less *quantitative* evidence of a global linguistic diversity crisis. To help fill this gap we have created the Index of Linguistic Diversity (ILD), which we believe to be the first-ever quantitative index of trends in linguistic diversity based on time-series data on numbers of mother-tongue speakers. The ILD assesses trends in linguistic diversity by comparing changes in the relative distribution of mother-tongue speakers against a benchmark of the situation prevailing in 1970, the earliest year we could set the index based on the data available. The index does this by measuring changes in the number of mother-tongue speakers from a globally representative sample of 1,500 languages over the period 1970–2005. The ILD can be calculated at different geographic scales and for different groupings of languages; each of these versions of the index uses the same methods.

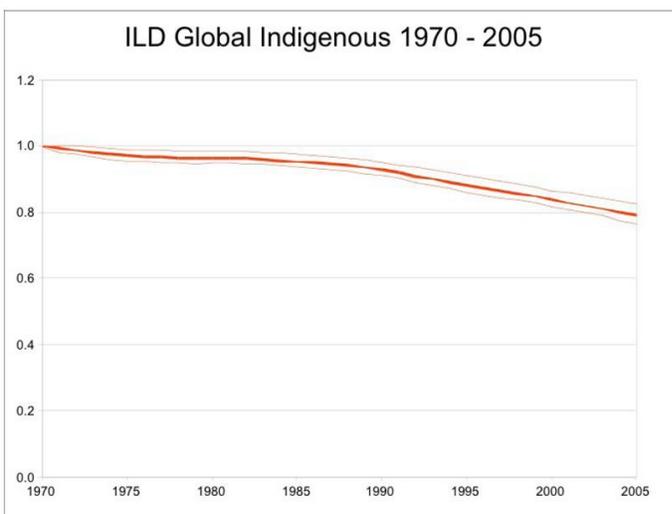


Figure 2: Declining Trend of Global Indigenous Linguistic Diversity, 1970-2005

The main finding of this research is that linguistic diversity has seriously declined since 1970. The overall linguistic diversity of the world, as measured by ILD Global, declined by 20% over the 35-year period (see Figure 1). We also assessed the diversity of the world's indigenous languages—which make up 80–85% of the total number—on both global and regional levels. We did this because the status of the world's indigenous languages is important to global initiatives such as the Convention on Biological Diversity, as well as to indigenous communities themselves. ILD Global Indigenous, which measures the diversity of the world's indigenous

languages, declined by 21% (Figure 2). The diversity of indigenous languages declined in all regions as well.

2. WHAT IS LINGUISTIC DIVERSITY?

Linguistic diversity is often viewed from three related perspectives: *language richness*, or the number of different languages spoken in a given geographical area; *phylogenetic diversity*, or the number of different lineages of languages found in an area; and *structural diversity*, or the variation found among structures within languages.

For the purposes of developing a quantitative measure such as the ILD, we departed slightly from these standard definitions of linguistic diversity, and borrow some related concepts from the field of ecology. Language richness can be thought of as being analogous to species richness, the number of species found in a given area. In addition to richness, a second component in species diversity is evenness, or the distribution of individual organisms among species. In the case of linguistic diversity, evenness is the distribution of individual speakers among languages. For example, two regions in both of which ten languages are spoken each have the same richness, but the region in which each language is spoken by 10% of the population has greater evenness, and therefore higher linguistic diversity, than one in which 91% of the population speaks one language and only 1% of the population speaks each of the other nine.

We think that this concept is critical in measuring changes in linguistic diversity over comparatively short time scales. Relatively few of the world's languages have become extinct as mother tongues in the last few decades, so language richness in most areas of the world has declined only slightly. And yet, we would argue, diversity has declined much more than this because the distribution of mother-tongue speakers among extant languages has become more uneven: more speakers are becoming concentrated in fewer languages. While phylogenetic and structural diversity are important, these concepts are not currently incorporated into the index. In summary, for the purposes of the ILD, **we define linguistic diversity as the number of languages and the evenness of distribution of mother-tongue speakers among languages in a given area.**

3. THE NEED FOR A LINGUISTIC DIVERSITY INDEX

If there are already projections of the future magnitude of language extinctions, why is there a need for an index like the ILD? First, published estimates of the percentage of languages likely to die out during this century are, to date, little more than informed conjecture. Categorical statements of the rate of extinction—"X number of languages are dying every year"—are widely quoted but almost never referenced to a rigorous estimate.

Second, even if better estimates were available, merely tracking when particular languages go extinct does not account for the loss of linguistic diversity occurring during the course of pre-extinction language shift. **A great deal of linguistic diversity is lost well before a declining language finally goes extinct, as speakers shift to other (usually larger) languages, intergenerational transmission declines, and usage becomes restricted to fewer speakers,**

domains, and functions. Quantifying changing distributions of mother-tongue speakers prior to extinction is therefore important.

Moreover, focusing on language extinction rates places undue emphasis on what is perceived to be the terminal state of linguistic diversity decline. If “language extinction” is to have any useful meaning, it must be specified that the term actually refers to the condition of a language no longer being spoken as a mother tongue.

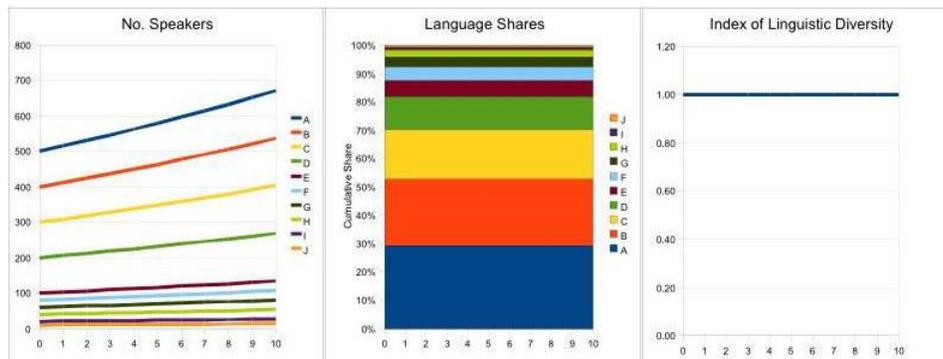
So, while obtaining accurate projections of mother-tongue language extinctions is important, they need to be augmented by a quantitative measure of current global trends in linguistic diversity. Clearly, the claims of those who tout the loss of linguistic diversity as a major problem for the world would be strengthened if there were quantitative evidence to support their arguments. **Government officials, other decision-makers, and the general public will likely take the decline of linguistic diversity more seriously if there is a readily understandable global metric that captures the current magnitude of the problem.** That is what the ILD is designed to provide.

4. WHAT THE ILD MEASURES

The ILD uses language evenness in conjunction with language richness as a proxy for linguistic diversity. Because the goal of the index is to measure trends in linguistic diversity, it must account for changes in evenness and richness: that is, changes in the relative distribution of mother-tongue speakers among discrete languages within the total population, as measured from the starting point of the index (currently 1970) to its ending point (currently 2005). The ILD indicates the rate of change in linguistic diversity by measuring how far, on average, the languages in a given grouping deviate from a hypothetical situation in which each language is neither increasing nor decreasing its share of the total population of that grouping.

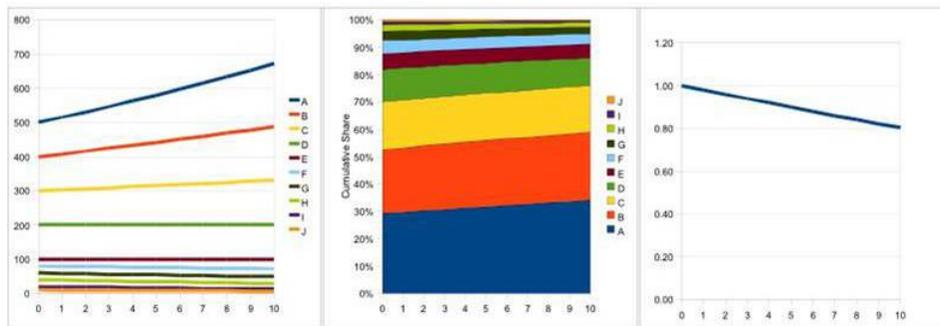
For example, ILD Global, an index of the world’s overall linguistic diversity, measures the average deviation of the world’s languages from a hypothetical situation in which each language is neither increasing nor decreasing its share of the global population.

How the ILD is calculated:



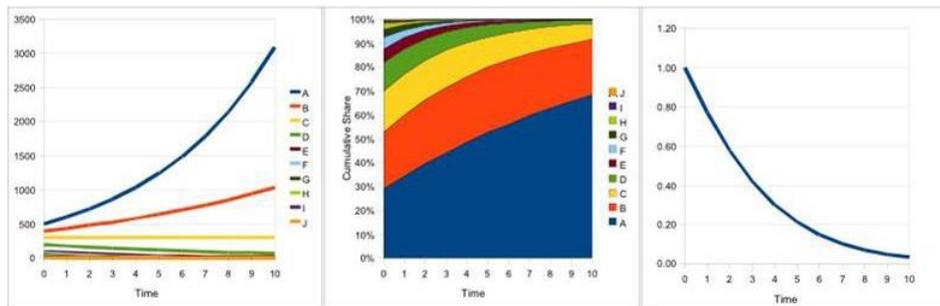
Scenario 1: Stable Equilibrium

The *ILD* indicates the rate of change in linguistic diversity by measuring how far, on average, the languages in a given grouping deviate from a hypothetical situation in which each language is neither increasing nor decreasing its share of the total population of that grouping. **Scenario 1** shows that hypothetical situation. Imagine a world with just 10 languages, here marked A through J. The largest of these languages, Language A (in dark blue) has 500 speakers at the beginning of our imaginary survey; this is shown on the left-side graph. It, along with the other 9 languages, gets bigger over 10 years (time is the X axis, at the bottom of all the graphs). But each languages gets bigger at exactly the same rate, so that each one's share of the total population — shown in the middle graph — is flat across the 10-year span. This is the condition of hypothetical stability against which the *ILD* measures change. As you can see in the rightmost graph, the *ILD* remains unchanged under this stable equilibrium scenario.



Scenario 2: Steady Erosion

In **Scenario 2**, we see something very similar to what is happening in the real world. Here, the 3 largest languages are increasing in population while the rest remain flat (leftmost graph). This is shown in the middle graph, where the amount of available space is being taken up more and more by the three largest languages (dark blue, orange, and yellow). The *ILD* in this scenario declines by 20%.



Scenario 3: Serial Extinction

In **Scenario 3** of our illustration of our simplified world we look at what would happen if the 3 largest languages really increase their share of the world's population at the expense of smaller languages that go extinct one

after another. The blue, orange, and yellow bands in the middle graph bulge dramatically, literally helping to squeeze out small languages as more and more of the world's people become concentrated in fewer and fewer languages, The ILD in the rightmost graph plunges dramatically.

The ILD can be said to measure the concentration or distribution of mother-tongue speakers among the world's languages. **What does it mean to say that ILD Global declined 20% over the period 1970–2005? It means that, for all languages spoken worldwide in 1970, their average share of the world's population declined by 20% over 35 years.**

It is worth noting again that the ILD is not a measure of language extinction: a 20% decline in the index does not mean that 20% of languages went extinct over the period being measured. For example, it is possible to imagine that most of the world's languages could decline until only a few speakers of each are left, while a few languages become dominant with many millions of speakers: the ILD would show a marked decline and yet the total number of extant languages would remain constant. In that case the number of extinctions would remain zero, yet the ILD would indicate that almost all linguistic diversity had been lost.

5. THE ILD DATABASE

SO06 K08-1
kum Main language name as given in E05
 Old E-code: K08K Yes
 Kumyk
 Russia (Europe) possible head anomaly? Yes

Ethnologue 2005 382,554 <input type="checkbox"/> 2005 E05	Ethnologue 2000 382,500 <input type="checkbox"/> 1993 L93	Ethnologue 1996 <input type="checkbox"/>	Ethnologue 1992 <input checked="" type="checkbox"/> 188,000 1970 census
Ethnologue 1988 <input checked="" type="checkbox"/> 188,000 1970 census	Ethnologue 1984 <input type="checkbox"/>	Ethnologue 1978 <input type="checkbox"/> 188,000 1970 census	Gunnmark 1985 <input type="checkbox"/>
Gunnmark 1983 288,000 <input type="checkbox"/> 1983	Mellet/Cohen 1983 <input type="checkbox"/>	Mellet/Cohen 1980 83,400 <input type="checkbox"/> 1997	Trankle 1979 <input type="checkbox"/> 112,000

Example of ILD Database Entry Form

The ILD database of time-series data on language demographics, which we believe to be the world's largest to date, contains information from nine editions of *Ethnologue*, the most comprehensive compendium of the world's languages, as well as five other compendia of speaker numbers.

The ILD is based on a sample of 1,500 languages selected at random from the 7,299 languages listed in the 15th edition of *Ethnologue* (2005). (The 16th edition, 2009, appeared too late for us

to include in this study.) This sample size—representing just over 20% of the world’s languages—is higher than is needed to constitute a statistically representative global sample. Having a sample size much larger than required for global analysis allows statistically valid analysis of subglobal samples.

Our long-term aim is to base the ILD on a variety of data sources, not just *Ethnologue*. However, we decided to restrict the first version of the ILD to *Ethnologue* data to minimize potential inconsistencies in language-status assessment that could come from incorporating multiple sources of data into a single time series.

The ILD database and methodology are described in the appendixes to the published version of the [ILD](#).

6. RESULTS

Global Linguistic Diversity. ILD Global (see Figure 1 above), which covers all the languages in the sample, both indigenous and non-indigenous, shows a slow decline from 1.0 to 0.95 between 1970 and 1988, but a steeper decline from 0.95 to 0.80 between 1988 and 2005. The upper and lower confidence limits show the boundaries of the 95% confidence interval, and are depicted in this and the other graphs as small lines above and below the main trendline.

Global Indigenous Linguistic Diversity. ILD Global Indigenous (see Figure 2 above), which covers only the indigenous languages in the sample, declined from 1.0 to 0.94 between 1970 and 1988, and from 0.94 to 0.79 between 1988 and 2005. It shows a marginally greater decline than the global ILD, but the two trends are largely similar as most of the languages in the global dataset are indigenous languages.

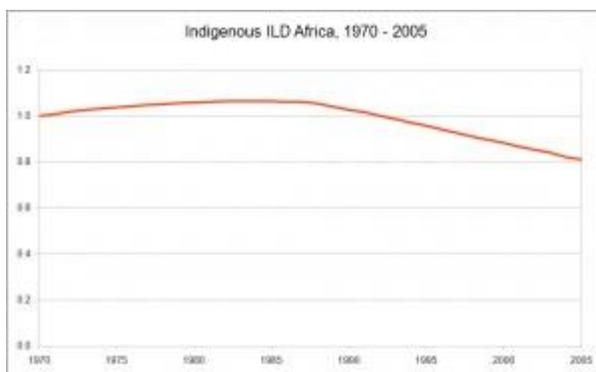


Figure 3: Trend of Indigenous Linguistic Diversity of Africa, 1970–2005

Regional Indigenous Linguistic Diversity. Changes in indigenous linguistic diversity differ among regions. ILD Africa Indigenous increased from 1.00 to 1.07 between 1970 and 1985, and then declined rapidly from 1.07 to 0.83 in 2005 (Figure 3). The increase in the 1970s and early 1980s suggests that African indigenous languages were becoming more equally distributed in

terms of speaker numbers during that period, but from the mid-1980s on the distribution became increasingly skewed, with many languages' share of the total African population declining.

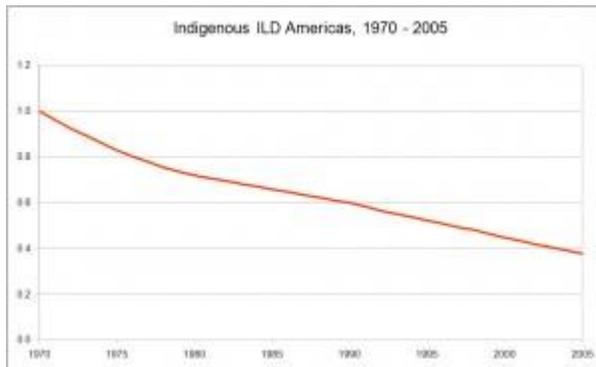


Figure 4: Trend of Indigenous Linguistic Diversity of the Americas, 1970–2005

ILD Americas Indigenous shows the steepest decline of any region, falling from 1.00 to 0.71 between 1970 and 1980, and from 0.71 to 0.36 between 1980 and 2005 (Figure 4).

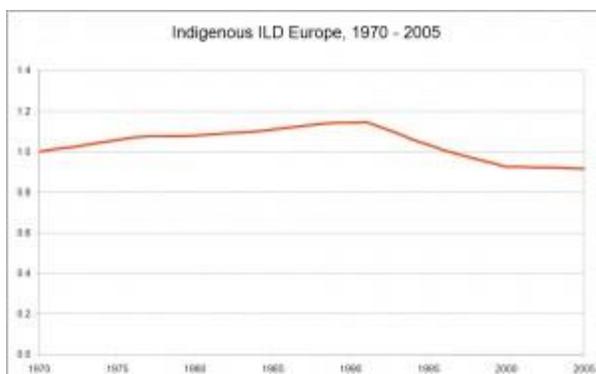


Figure 5. Trend of Indigenous Linguistic Diversity of Eurasia, 1970–2005

ILD Eurasia Indigenous, like its African counterpart, showed an initial increase from 1.00 to 1.10 between 1970 and 1981, suggesting that there was a slight gain in the proportion of the total population speaking an indigenous language. It flattened out for about a decade between 1981 and 1991, and then declined very slightly to 1.07 in 2005 (Figure 5). Overall the index shows little change in linguistic diversity in Eurasia.

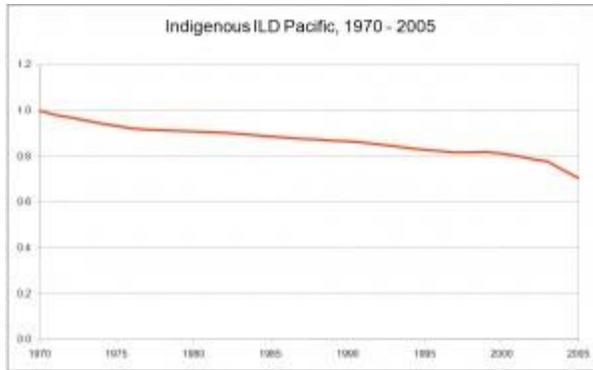


Figure 6: Trend of Indigenous Linguistic Diversity of the Pacific, 1970–2005

ILD Pacific Indigenous (which includes Australia) shows the second steepest decline after the Americas. The index fell steadily from 1.0 to 0.82 in 1999, then dropped steeply from 0.82 to 0.70 between 1999 and 2005 (Figure 6). The widening confidence intervals in the last few years of the index suggest a higher degree of uncertainty in the trend after 1999, which would be reduced with additional data.

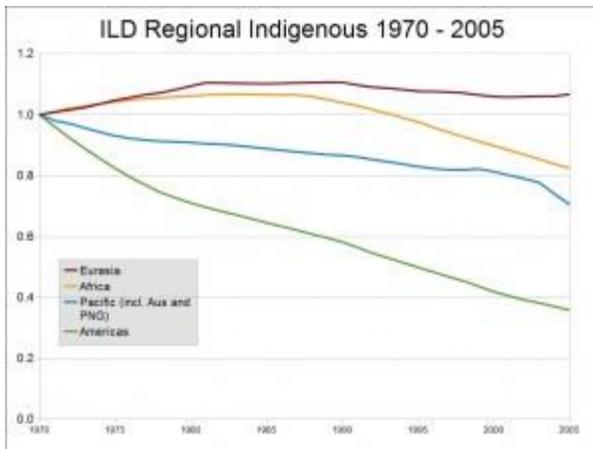


Figure 7: Comparison of Regional Trends of Indigenous Linguistic Diversity, 1970–2005

The four regional ILDs are compared in Figure 7.

7. DISCUSSION

Decline in Global Linguistic Diversity. Figure 1 above shows the global trendline for the ILD. ILD Global shows a slow decline from 1.0 to 0.95 between 1970 and 1988, but a steeper decline from 0.95 to 0.8010 between 1988 and 2005. The overall decline of 20% in the space of 35 years shows that linguistic diversity is being lost at a significant rate, but even more importantly, the rate of loss has increased from about -0.3% per year in the 1970s and 1980s to more than -1.0% per year in the 1990s and 2000s. This is a stark indication of the scale of the recent loss of global linguistic diversity. **The rapid disappearance of one-fifth of the linguistic diversity that existed in the world in 1970 is a quantitative depiction of the continuing widespread shift**

from smaller languages to larger languages. The more the ILD Global declines, the more the world's mother-tongue speakers are concentrated into fewer languages.

Decline in Global Indigenous Linguistic Diversity. Figure 2 above shows that the decline in the diversity of the world's indigenous languages has been similar, which is unsurprising in that **most of the languages in the world (by our estimate, 80–85%) are indigenous languages.** ILD Global Indigenous declined from 1.0 to 0.79 between 1970 and 2005—a 21% decrease. The average annual rate of decline in indigenous linguistic diversity was slightly faster than the global average in the 1970s and 1980s, but only by a fraction of a percent per year.

Indigenous communities themselves would certainly want to know the status of indigenous languages. Moreover, the Convention on Biological Diversity (CBD) identified stemming the rate of loss of linguistic diversity and in the number of speakers of indigenous languages as one of its indicators for assessing progress toward meeting its 2010 Biodiversity Target. The acceleration in the loss of linguistic diversity indicated by the ILD Global Indigenous implies that this particular CBD target could not be met. Prospects remain uncertain for the next CBD target (2020).

Declines in Regional Indigenous Linguistic Diversity. A comparison of the various regional indigenous ILDs (see Figure 7 above) shows some interesting results. Some regions are declining more rapidly than others, particularly the Americas, which declined by 64% over the period (Figure 4 above). The fact that the Americas showed the greatest overall decline should not necessarily be interpreted as meaning that linguistic diversity is, consequently, lower there than in other regions. It simply means that the Americas underwent the most rapid decline of all four regions between 1970 and 2005. It may well have been the case that the Americas were much more linguistically diverse in 1970 compared with other regions, such as Europe for example, in which the majority of linguistic diversity was lost prior to 1970.

The Pacific region (Figure 6 above) shows the second greatest rate of decline, 30% over 35 years, while ILD Africa Indigenous (Figure 3 above) declined by nearly 20%. This suggests that indigenous languages are in very rapid decline in comparison to total population growth in the region as a whole in the Americas, and in rapid decline in Africa and the Pacific.

Eurasia was the only region to show an increase in its indigenous ILD (Figure 5 above). There, indigenous languages are growing at the same rate as the overall population.

Some Caveats and Limitations. While we expect the ILD to prove a useful tool to communities, analysts and academics, policymakers, and the general public, any index is only as good as the underlying data available at the time. *Ethnologue* is the best single source for data on the numbers of speakers of languages around the world, and information from its various editions is an indispensable part of any analysis of recent trends in language demographics. Nonetheless, *Ethnologue* data come from a variety of primary and secondary sources and are, inevitably, uneven. We believe that *Ethnologue* time-series data are valid, but without question language demographic data in general can be improved. It should be borne in mind when using the initial version of the ILD that better data will, in the future, produce even more accurate trendlines.

It is also important to acknowledge that **global indices such as the ILD should be used to provide broad contextual background for policy frameworks, rather than as guidance for on-the-ground policy decisions.** No large-scale language index can hope to fully represent the complexities that must be accounted for in any policy affecting individual language communities. Nor can a global or regional index do more than outline the state of linguistic diversity at these levels; much more fine-grained analyses are required to get a complete picture.

Quantitative analyses such as the ILD must be supplemented by knowledge derived through other methods. This is especially relevant with respect to languages because **most linguistic diversity is tied to traditional knowledge systems of indigenous people.** These systems primarily rely on non-quantitative observational science and narrative, often transmitted orally rather than in writing. Therefore, any global numerical index, including the ILD, runs the risk of being irrelevant (or, worse, antithetical) to the needs of indigenous communities if it is not properly qualified—and, in addition, supplemented by other information that is generated by the communities themselves.

The ILD and similar global indices that deal with potentially controversial phenomena, such as language policy, must carefully be placed in context whenever they are used as an educational or policy-orientation tool, and should never be used as a sole source of information.

Future Development of the ILD. As part of future work, we plan to add data from the 16th edition of *Ethnologue* and **to expand the database to achieve complete coverage of all the world's languages.** We also intend to enter into the ILD database all available speaker-numbers data from other global compendia of language statistics, as well as information from UNESCO's *Atlas of the World's Languages in Danger* and other UNESCO-led data-gathering efforts. All of these will provide data with which to compare, or add to, those from *Ethnologue*.

But the full potential of the ILD methodology won't be realized until we are able to expand it to include other language demographic data in addition to counts of mother-tongue speakers. To fully understand the status of and trends in the world's linguistic diversity, we need to go beyond using language richness (the number of discrete languages) and language distribution as a proxy. For example, it may be possible to create versions of the ILD that address phylogenetic diversity by using data on language family affiliations that are already included in *Ethnologue*. The methodology could also be applied to certain special language categories, thus producing versions such as ILD Creoles or ILD Isolates. There may be scope for incorporating structural diversity into the ILD by drawing on data from the *World Atlas of Language Structures*. **Even better understanding will come when we are able to augment speaker-numbers data with deeper knowledge about all the factors that determine language demographics and drive trends in linguistic diversity.**

*The text on this page is an abridged and slightly modified version of the 2010 article by the same name, authored by David Harmon and Jonathan Loh and published in the journal *Language Documentation & Conservation*. Reproduced with permission.