Endangered language families

D. H. Whalen, Gary F. Simons

Language, Volume 88, Number 1, March 2012, pp. 155-173 (Article)

Published by Linguistic Society of America
DOI: 10.1353/lan.2012.0012


LANGUAGE
A JOURNAL OF THE LINGUISTIC SOCIETY OF AMERICA

SHORT REPORT

Endangered language families

D. H. WHALEN
City University of New York,
Haskins Laboratories, and
Endangered Language Fund

GARY F. SIMONS
SIL International and
Graduate Institute of
Applied Linguistics

Linguists have increased their documentation efforts in response to the sharp decline in the number of languages. Greater awareness and new sources of funding have led to an upsurge in language documentation. While individual languages make unique contributions to the world's linguistic heritage, language families, by virtue of their shared heritage, have related contributions to make. The endangerment of entire families, while implied by the loss of language, has not been explored to date. Here, we examine estimates of how many linguistic stocks (the largest subgroups of related languages that are reconstructable) consist entirely of endangered languages and thus are endangered themselves. Our data set includes 372 stocks that had at least one living language in 1950. Our finding is that since that time, 15% of the world's linguistic stocks have become extinct and another 27% are now moribund in that direct estimates of endangerment indicate that no member languages are being learned by children. For cases where direct estimates were not found, we used population as a proxy for endangerment. If, as many predict, a further 50% of the world's languages become extinct or moribund in this century, then an additional 25% of the linguistic stocks would be at risk of being lost. If 90% share this fate, only 11% of stocks would have at least one presumably 'safe' language. A comparison of the vitality of linguistic stocks by world areas yields the following ranking from worst to best: Americas, Pacific, Asia, Africa, Europe. Finally, some aspects of language that are unique to disappearing language stocks are outlined. Renewed efforts at documenting members of such stocks seem justified.*

Keywords: endangered languages, language families, language stocks, language diversity, typology

1. INTRODUCTION. Language diversity is at risk, a circumstance that received prominent attention in the pages of this journal twenty years ago (Hale et al. 1992). Individual languages are falling silent at a rate that is likely to be unprecedented in human history. Languages share many common features and seem to be learnable by any typically developing child, but each language, by definition, has features that are unique to it. Related languages are likely to share some features, through inheritance, that are unusual in other languages. Therefore, language families can be seen as a further locus of language diversity (Hammarström 2010, Whalen 2004:331). With the greater time depth that separates families as compared with the separation of languages, the differences are likely to be deeper as well. Our aim in this article is to explore the degree to which this deeper level of diversity is also endangered.

Languages are presumed to have arisen and disappeared throughout history. Most linguists assume that all modern languages derive from a single human language (e.g. Dessalle 2007, Swadesh 1972:214), but it is possible that many features of modern languages entered an already varied base of protolanguages (Freedman & Wang 1996, Nichols 1992). In either case, diversification has probably occurred consistently, and it has certainly done so in the times for which we have evidence. Our two sources of evidence of ancient languages, writing and reconstruction, are insufficient to give us anything like a complete picture of the range of languages that have existed, and the

* We thank Johanna Nichols and Balthasar Bickel for providing their database of stocks from the Autotyp project. We thank Anahit Minasyan for facilitating access to the UNESCO data. We thank Johanna Nichols, Patricia Shaw, Jürgen Bohnemeyer, and an anonymous referee for helpful comments. This work was supported in part by NSF grant 0966411 to Haskins Laboratories.

resultant time span is extremely limited as well (e.g. Hock & Joseph 1996). Barring some new way of seeing into the past, these limitations will continue to hold.

Although languages have come and gone throughout history, the rate of disappearance we see today appears unexpectedly rapid. Various authors have pointed out the greatly increased rate of loss of languages taking place in the modern world (Crystal 2000, Hale et al. 1992, though see Whaley 2003). These extinction rates are unlikely to have been found in previous eras and are larger than the loss of bird and mammal species (Sutherland 2003). The languages lost carry great social significance for the peoples involved, and the resulting gaps in our knowledge of human language limit our attempts to understand language in its ultimate range of expression.

The wide variety of language patterns seen in the different language families argues for an extensive time depth for language diversification in general. It seems unlikely that we will have the ability to gauge accurately the time depth at which the various languages within a family diverged (Evans et al. 2006), though computational techniques hold some promise (Holman et al. 2011). The comparative method is not, by definition, applicable to comparisons between families, given the high rates of random similarity that can mimic divergence from a common source (Ringe 1999). Indeed, we would not define the families as distinct if it were possible to provide a convincing reconstruction of earlier forms from all of the languages. We do know, however, that the differences between families are larger than those within families, so that it is likely that larger amounts of time separate the (currently unreconstructable) divergences of the families than of the languages within families. Given that the diversification of the Indo-European languages took several thousand years, many more thousands of years of language development are likely to have been necessary for the rise of diversity in today's language families.

In this article, we examine how much of the world's linguistic diversity at the level of language families is at risk. There are many caveats to this enterprise. The first of these is the definition of 'family' itself, since there is not universal agreement among linguists as to the level in classification that it should refer to. For our analysis we equate it with the level of the linguistic stock—that is, the largest grouping of languages for which relatedness can be demonstrated and for which a plausible protolanguage can be reconstructed. Another limitation of the data available is that the state of endangerment is not directly known for most of the languages. When such estimates are available, we used them; otherwise, we used population estimates as our best indicator of endangerment. With these caveats, we find that a large proportion of the world's language families are at risk.

2. METHODS.

2.1. IDENTIFYING STOCKS. The base data set for our study was built from the comprehensive genealogical classification of all of the world's languages reported in the *Ethnologue* (Lewis 2009). The total inventory of 7,413 languages includes every human language that is known to have been in use in 1950, which is the year the *Ethnologue* first began tracking languages. The statistical summaries in tables 4 and 5 of Lewis 2009:26–32 list 116 language families. However, these 'families' are not based entirely on our ability to reconstruct plausible parent languages but often include geographical inferences as well. For example, the 'Australian' family includes approximately seventeen subgroups that could plausibly be assumed to have a common ancestor (Bowern & Koch 2004, Evans 2003). There are similarities that are evident across the continent, though some of those may be due to diffusion. It is not possible, however, to make definitive links among all of the languages using the comparative method. Even the seventeen subgroups of Australian are subject to further debate. Similarly, the languages of Afro-Asiatic, Niger-Congo, and Altaic seem to have some kind of historical connection within each of those groups, but it is not clear that it is due to descent from a common language, since there is no consensus on the proof of relatedness or the reconstruction of a protolanguage.

Not only is the term ‘language family’ commonly used to refer to groupings of languages that are larger than what can be clearly proven to have a common ancestor, but it is also commonly used for a smaller grouping that shares the same protolanguage. Thus, Germanic, Romance, and Indic are sometimes referred to as language families that are branches of a larger Indo-European language family.

In order to operationalize the definition of language family for our statistical analysis of language diversity at risk, we set it at the level of linguistic stock, understood to be a grouping of ‘all languages that are known to be genetically related’ (Sapir 1921:163). We have followed Nichols (1992) in interpreting stock as the largest grouping whose membership is provable by means of the comparative method. We used data from the Autotyp database graciously shared by Balthasar Bickel and Johanna Nichols to identify the nodes in the classification trees of the *Ethnologue* that correspond to the stocks in Autotyp. A total of 250 such stocks were identified.

Also classed as stocks are single languages that have not yet been recognized as members of a known stock. This includes languages identified in Lewis 2009 as ‘language isolates’. These are languages that are claimed to have no demonstrable relatives, either living or extinct. Most linguists would assume that these were related to other languages in the far past, but the current evidence is insufficient to reconstruct a common ancestor with any other attested human languages. Also classed as stocks for the purposes of this study are the unclassified languages. These are languages that have had classifications posited for them that are either too contentious or too unconvincing to be able to place them in another stock at the present time. A total of 122 such stocks were identified for this study.

Table 1 gives a statistical summary of the 372 stocks identified for this study, showing their distribution by the five major world areas (Africa, Americas, Asia, Europe, Pacific). Each stock is counted only once, so a stock that spans multiple areas is counted only with the area of its origin. For instance, Austronesian is counted for Asia and Indo-European is counted for Europe. Table 1 also indicates the numbers of languages (by area) that were included in the data set for this study. A total of 7,093 languages are included. This does not include all languages reported in Lewis 2009. Specifically excluded are deaf sign languages and contact languages (creoles, mixed languages, and pidgins) that have originated in a relatively short time frame and do not fit into the scheme of stocks that have been diverging for millennia. Also excluded are special-purpose languages like secret languages and ancient liturgical languages.

AREA	LANGUAGES	STOCKS	GROUPINGS	ISOLATES
Africa	2,107	56	39	17
Americas	1,115	149	78	71
Asia	2,313	78	52	26
Europe	195	6	5	1
Pacific	1,363	83	76	7
Totals	*7,093*	*372*	*250*	*122*

TABLE 1. Languages and stocks by world area. The last two columns show the breakdown of stocks into those that are groupings of related languages versus those that are single languages (isolates) that have never been proven to be related to another. The assignment of languages to five world areas follows Lewis 2009. The dividing line between Europe and Asia is the Ural Mountains. Asia includes insular Southeast Asia; the dividing line between Asia and the Pacific is the political boundary between Indonesia and Papua New Guinea.

2.2. Assessing endangerment. Two sources exist for relatively large numbers of direct assessments of language endangerment, UNESCO's *Atlas of the world's languages in danger* (Moseley 2010) and the *Ethnologue* (Lewis 2009). Because these both use the three-letter codes of the ISO 639-3 standard to uniquely identify languages (International Organization for Standardization 2007), it was relatively straightforward to compare across the two sources. Our aim was to classify each of the 7,093 languages in our data set as belonging to one of three categories: extinct, moribund, or viable. Extinct signifies that there are no known speakers; there may still be an identifiable ethnic population but nothing more than symbolic use of the language remains—it is no longer used for everyday communication. In defining moribund we follow Krauss who, in his landmark study of language endangerment, used this category to signify that the language is 'no longer being learned as mother-tongue by children' (Krauss 1992:4). All other languages are classed as being still viable. That in no way implies that they are not under serious threat, but only that our sources do not indicate that intergenerational transmission of the language has been broken.

The language endangerment assessments aggregated by UNESCO (Moseley 2010) come from a wide range of experts throughout the world. The levels of endangerment are coded (from highest to lowest) as: extinct, critically endangered, severely endangered, definitely endangered, and vulnerable. The definition of the vulnerable category states that most children still use the language, while definitely endangered is defined as 'children no longer learn the language as mother tongue in the home'. Thus we have mapped their three levels of 'endangered' to the category moribund, and mapped vulnerable to viable. There are cases where multiple ratings are given to one ISO 639 language code, either because different dialects have different statuses or because the same language has different statuses in different countries. In these cases we have taken the strongest reported level of vitality as the status for the language as a whole, seeing it as the best indicator of the long-term prospects for the language.

The *Ethnologue* (Lewis 2009) has categories for extinct and nearly extinct; these we mapped to extinct and moribund, respectively. Beyond that, *Ethnologue* does not assess language endangerment directly. However, it does make statements about intergenerational transmission in the 'Language use' section of entries. For our purposes, if the language was described as being spoken by 'older adults' or 'adults only' or 'mainly adults' or 'few children', it was classified as moribund. (Note that in making these assessments we have sometimes made use of information that is to be published in the next edition of *Ethnologue*.) The vast majority of entries make no statement about language use. In the absence of such information, we made the conservative judgment that the language is viable.

Table 2 gives a summary comparison of endangerment estimates in the UNESCO *Atlas* versus the *Ethnologue*. An obvious difference between the sources shows up in the 'Missing' column, which represents the 4,905 languages in the data set that do not have an entry in the UNESCO *Atlas*. While *Ethnologue* has the editorial policy to describe all languages, the UNESCO *Atlas* focuses on endangered languages. Thus the general presumption when a language is missing from the UNESCO *Atlas* is that it is safe. However, the UNESCO *Atlas* does not aim to be comprehensive with respect to extinct languages; thus, the *Ethnologue* documents 292 extinct languages that are missing from the UNESCO *Atlas*. Another striking result is that the two sources agree in classifying 432 languages as endangered (or moribund), but UNESCO identifies a further 964 as endangered that *Ethnologue* does not, while Ethnologue identifies 177 as moribund that are missing from the UNESCO *Atlas*. Possible explanations are missing data in *Ethnologue*, assessment of some languages in the UNESCO *Atlas* as 'definitely endangered' before intergenerational transmission has been broken, and failure to assign an ISO 639 code in the UNESCO data.

Ethnologue	*UNESCO estimates*				
estimates	EXTINCT	ENDANGERED	VULNERABLE	MISSING	TOTALS
EXTINCT	122	34	1	290	447
MORIBUND	23	432	18	177	650
VIABLE	4	964	590	4,438	5,996
Totals	*149*	*1,430*	*609*	*4,905*	*7,093*

TABLE 2. Correspondence of endangerment estimates between UNESCO's *Atlas* (Moseley 2010) and *Ethnologue* (Lewis 2009). Combinations treated in this study as 'extinct' (total 412) are indicated by a diagonal fill pattern and those treated as 'viable' (total 5,028) by a horizontal fill; those treated as 'moribund' (total 1,653) have no fill.

We did not attempt to rectify the discrepancies since the needed information is not available. Rather, we built our data set to combine the information in a way that suited our purposes. We classified a language as extinct only if both sources agreed in calling it extinct or if *Ethnologue* has it as extinct while it is missing from the UNESCO data. Similarly, we classified a language as viable only if *Ethnologue* has it as viable and it is not listed as endangered in the UNESCO data (that is, it is either vulnerable or missing and thus presumed safe). All other combinations were classified as moribund, since at least one of the sources gives evidence that the language is moribund. Table 3 gives a statistical summary of the vitality status of the 7,093 languages included in this study. The summary table shows the distribution of languages by vitality within each of the five major world areas. The count for each status is also given as a percentage of the total number of languages in the row so that one can easily compare between areas. For instance, the table shows that the rates of extinction and moribundity are highest in the Americas (16% and 43%, respectively) and lowest in Africa (2% and 12%).

AREA	EXTINCT	MORIBUND	VIABLE	TOTALS
Africa	46 (2%)	247 (12%)	1,814 (86%)	2,107
Americas	183 (16%)	476 (43%)	456 (41%)	1,115
Asia	47 (2%)	552 (24%)	1,714 (74%)	2,313
Europe	8 (4%)	95 (48%)	92 (48%)	195
Pacific	128 (9%)	278 (20%)	957 (70%)	1,363
Totals	*412 (6%)*	*1,648 (23%)*	*5,033 (71%)*	*7,093*

TABLE 3. Vitality status of languages by world area.

3. RESULTS.

3.1. EXTINCT STOCKS. Our data set shows that fifty linguistic stocks have gone extinct since 1950; that is, the last surviving language within the stock is no longer known to have any speakers. Table 4 gives the complete list. When a stock is a single language, it is parenthetically identified as being an isolate or unclassified. When the stock name has only one part, it is the name of a top-level family in the *Ethnologue*'s genealogical classification. When the name has two parts separated by a comma, the first part is the name of the top-level node and the second part is the name of the subgroup node that corresponds to the stock. It is not necessarily shown as an immediate subgroup in the *Ethnologue* tree. The groupings can be looked up online in the '*Ethnologue* language family index' (http://www.ethnologue.com/family_index.asp).

STOCK	LANGS	SAMPLE LANGUAGE	COUNTRY
Agavotaguerra (unclassified)	1	Agavotaguerra [avo]	Brazil
Aguano (unclassified)	1	Aguano [aga]	Peru
Australian, Giimbiyu	3	Erre [err]	Australia
Australian, Laragiyan	1	Laragia [lrg]	Australia
Australian, Limilngan-Wulna	2	Limilngan [lmc]	Australia
Australian, Umbugarla-Ngumbur	2	Ngurmbur [nrx]	Australia
Cagua (unclassified)	1	Cagua [cbh]	Colombia
Canichana (isolate)	1	Canichana [caz]	Bolivia
Chimakuan	1	Quileute [qui]	United States
Chimariko (unclassified)	1	Chimariko [cid]	United States
Chipiajes (unclassified)	1	Chipiajes [cbe]	Colombia
Chumash	7	Chumash [chs]	United States
Coahuiltecan	1	Tonkawa [tqw]	United States
Coxima (unclassified)	1	Coxima [kox]	Colombia
Gulf	4	Natchez [ncz]	United States
Hibito-Cholon	2	Hibito [hib]	Peru
Hokan, Esselen	1	Esselen [esq]	United States
Iapama (unclassified)	1	Iapama [iap]	Brazil
Kaimbé (unclassified)	1	Kaimbé [xai]	Brazil
Kamba (unclassified)	1	Kamba [xba]	Brazil
Kambiwá (unclassified)	1	Kambiwá [xbw]	Brazil
Kapinawá (unclassified)	1	Kapinawá [xpn]	Brazil
Karirí-Xocó (unclassified)	1	Karirí-Xocó [kzw]	Brazil
Kunza (unclassified)	1	Kunza [kuz]	Chile
Lenca (unclassified)	1	Lenca [len]	Honduras
Mato Grosso Arára (unclassified)	1	Arára, Mato Grosso [axg]	Brazil
Mawa (unclassified)	1	Mawa [wma]	Nigeria
Molale (unclassified)	1	Molale [mbe]	United States
Monimbo (unclassified)	1	Monimbo [mom]	Nicaragua
Natagaimas (unclassified)	1	Natagaimas [nts]	Colombia
Pankararé (unclassified)	1	Pankararé [pax]	Brazil
Pankararú (isolate)	1	Pankararú [paz]	Brazil
Pataxó Hã-Ha-Hãe (unclassified)	1	Pataxó Hã-Ha-Hãe [pth]	Brazil
Penutian, Oregon	5	Coos [csz]	United States
Pijao (unclassified)	1	Pijao [pij]	Colombia
Rer Bare (unclassified)	1	Rer Bare [rer]	Ethiopia
Salinan (isolate)	1	Salinan [sln]	United States
Subtiaba-Tlapanec	1	Subtiaba [sut]	Nicaragua
Tapeba (unclassified)	1	Tapeba [tbb]	Brazil
Tingui-Boto (unclassified)	1	Tingui-Boto [tgv]	Brazil
Tremembé (unclassified)	1	Tremembé [tme]	Brazil
Truká (unclassified)	1	Truká [tka]	Brazil
Tuxá (isolate)	1	Tuxá [tud]	Brazil
Uamué (unclassified)	1	Uamué [uam]	Brazil
Wakoná (unclassified)	1	Wakoná [waf]	Brazil
Wasu (unclassified)	1	Wasu [wsu]	Brazil
Weyto (unclassified)	1	Weyto [woy]	Ethiopia
Xinca (unclassified)	1	Xinca [xin]	Guatemala
Xukurú (unclassified)	1	Xukurú [xoo]	Brazil
Yeni (unclassified)	1	Yeni [yei]	Cameroon

TABLE 4. Linguistic stocks that have become extinct since 1950. The LANGS column is the number of languages that have gone extinct since 1950, not the total number of languages known to have been a member of the stock. ISO 639 codes are included for sample languages.

3.2. MORIBUND STOCKS. Our data set shows that 102 linguistic stocks are currently moribund; that is, intergenerational transmission is reported to be broken in every surviving language within the stock. Table 5 gives the complete list. The stocks are named as explained above for Table 4. The stocks are listed in ascending order of the size of the largest surviving language.

STOCK	LANGS	LARGEST LANGUAGE	SPEAKERS	COUNTRY	*Ethnologue*	UNESCO
Luo (unclassified)	1	Luo [luw]	1	Cameroon	nearly extinct	
Taushiro (isolate)	1	Taushiro [trr]	1	Peru	nearly extinct	critical
Yámana (isolate)	1	Yámana [yag]	1	Chile	nearly extinct	critical
Abishira (unclassified)	1	Abishira [ash]	2	Peru	extinct	critical
Cayubaba (isolate)	1	Cayubaba [cyb]	2	Bolivia	extinct	critical
Tinigua (isolate)	1	Tinigua [tit]	2	Colombia	nearly extinct	critical
Bung (unclassified)	1	Bung [bqd]	3	Cameroon	nearly extinct	critical
Muniche (isolate)	1	Muniche [myr]	3	Peru	nearly extinct	critical
Yuki	1	Wappo [wao]	3	United States	nearly extinct	extinct
Chon	1	Tehuelche [teh]	4	Argentina	extinct	critical
Penutian, Maiduan	2	Maidu, Northwest [mjd]	4	United States	nearly extinct	critical
Australian, Marriammu	1	Marriammu [xru]	5	Australia	nearly extinct	
Kanoé (unclassified)	1	Kanoé [kxo]	5	Brazil	nearly extinct	critical
Penutian, California	1	Wintu [wit]	5	United States	nearly extinct	critical
Puelche (isolate)	1	Puelche [pue]	5	Argentina	nearly extinct	extinct
Kusunda (isolate)	1	Kusunda [kgg]	7	Nepal	nearly extinct	
Ongota (unclassified)	1	Ongota [bxe]	8	Ethiopia	nearly extinct	critical
Andamanese, Great	2	A-Pucikwar [apq]	10	India	nearly extinct	
Australian, Malagmalag Proper	2	Tyaraity [woa]	10	Australia	nearly extinct	
Hokan, Washo	1	Washo [was]	10	United States	nearly extinct	severe
Itonama (isolate)	1	Itonama [ito]	10	Bolivia	nearly extinct	critical
Alacalufan	1	Qawasqar [alc]	12	Chile	nearly extinct	critical
Australian, West Barkly	2	Wambaya [wmb]	12	Australia	nearly extinct	critical
Kutenai (isolate)	1	Kutenai [kut]	12	Canada	nearly extinct	severe
Penutian, Utian	10	Miwok, Central Sierra [csm]	12	United States	nearly extinct	critical
Ainu (isolate)	1	Ainu [ain]	15	Japan	nearly extinct	critical
Australian, Daly Proper	1	Madngele [zml]	15	Australia	nearly extinct	critical
Hokan, Karok-Shasta	2	Achumawi [acv]	16	United States	nearly extinct	critical
Kembra (unclassified)	1	Kembra [xkw]	20	Indonesia	nearly extinct	critical
Leco (isolate)	1	Leco [lec]	20	Bolivia	nearly extinct	critical
Lule-Vilela	1	Vilela [vil]	20	Argentina	nearly extinct	extinct
Caddoan	4	Caddo [cad]	25	United States	nearly extinct	critical
Kehu (unclassified)	1	Kehu [khh]	25	Indonesia	nearly extinct	
Kwaza (unclassified)	1	Kwaza [xwa]	25	Brazil	nearly extinct	critical
Massep (isolate)	1	Massep [mvs]	25	Indonesia	nearly extinct	critical
Penutian, Yokuts	1	Yokuts [yok]	25	United States	nearly extinct	severe
Trans-New Guinea, Mor	1	Mor [moq]	25	Indonesia	nearly extinct	severe
Australian, Bringen-Wagaydy	11	Maringarr [zmt]	30	Australia	nearly extinct	critical
Australian, Djamindjungan	1	Djamindjung [djd]	30	Australia	nearly extinct	severe
Namla (unclassified)	1	Namla [naa]	30	Indonesia	nearly extinct	
Himarimã (unclassified)	1	Himarimã [hir]	40	Brazil	nearly extinct	
Karahawyana (unclassified)	1	Karahawyana [xkh]	40	Brazil	nearly extinct	
Arutani-Sape	2	Arutani [atx]	42	Brazil	nearly extinct	critical
Hokan, Pomo	5	Kashaya [kju]	45	United States	nearly extinct	critical
Trans-New Guinea, Moraori	1	Morori [mok]	50	Indonesia	nearly extinct	severe
Yuchi (isolate)	1	Yuchi [yuc]	50	United States	nearly extinct	critical
Zaparoan	5	Arabela [arl]	50	Peru	nearly extinct	severe
Penutian, Chinookan	2	Wasco-Wishram [wac]	69	United States	nearly extinct	
Taiap (isolate)	1	Taiap [gpn]	80	Papua New Guinea		severe
Australian, Wororan	4	Ngarinyin [ung]	82	Australia		definite
Yukaghir	2	Yukaghir, Northern [ykg]	90	Russian Federation	nearly extinct	critical
Trumai (isolate)	1	Trumai [tpy]	100	Brazil	nearly extinct	critical

(TABLE 5. *Continues*)

STOCK	LANGS	LARGEST LANGUAGE	SPEAKERS	COUNTRY	*Ethnologue*	UNESCO
Australian, Garawan	1	Garawa [gbc]	110	Australia		severe
Usku (unclassified)	1	Usku [ulf]	110	Indonesia	older adults	critical
Carabayo (unclassified)	1	Carabayo [cby]	150	Colombia		extinct
Australian, Bunaban	2	Bunaba [bck]	160	Australia	older adults	severe
Aikanã (unclassified)	1	Aikanã [tba]	180	Brazil		definite
Yeniseian	1	Ket [ket]	190	Russian Federation	nearly extinct	severe
Centúúm (isolate)	1	Centúúm [cet]	200	Nigeria	older adults	
Wakashan	5	Heiltsuk [hei]	300	Canada	adults only	critical
Tol (isolate)	1	Tol [jic]	350	Honduras		critical
Shenenawa (unclassified)	1	Shenenawa [swo]	360	Brazil	older adults	
Andoque (isolate)	1	Andoque [ano]	370	Colombia	older adults	severe
Australian, Nyulnyulan	3	Bardi [bcj]	380	Australia	older adults	severe
Australian, Djeragan	3	Kitja [gia]	410	Australia		severe
Shabo (unclassified)	1	Shabo [sbf]	450	Ethiopia		severe
Harakmbet	2	Amarakaeri [amr]	500	Peru		definite
Imeraguen (unclassified)	1	Imeraguen [ime]	530	Mauritania	mainly adults	definite
Gilyak (isolate)	1	Gilyak [niv]	690	Russian Federation		severe
Salishan	19	Thompson [thp]	720	Canada	older adults	severe
Laal (unclassified)	1	Laal [gdm]	750	Chad		critical
Kujarge (unclassified)	1	Kujarge [vkj]	1,000	Chad		definite
Uru-Chipaya	2	Chipaya [cap]	1,200	Bolivia		definite
Penutian, Tsimshian	3	Gitxsan [git]	1,330	Canada		severe
Movima (isolate)	1	Movima [mzp]	1,450	Bolivia	older adults	severe
Chapacura-Wanham	3	Pakaásnovos [pav]	1,930	Brazil		definite
Nihali (isolate)	1	Nihali [nll]	2,000	India		critical
Yuracare (isolate)	1	Yuracare [yuz]	2,680	Bolivia	mainly adults	definite
Puinave (isolate)	1	Puinave [pui]	2,880	Colombia		definite
Candoshi-Shapra (isolate)	1	Candoshi-Shapra [cbu]	3,000	Peru		definite
Penutian, Plateau	5	Yakima [yak]	3,000	United States	older adults	
Zamucoan	2	Ayoreo [ayo]	3,070	Paraguay		definite
Lufu (unclassified)	1	Lufu [ldq]	3,200	Nigeria	older adults	
Tequistlatecan	2	Chontal, Highland Oaxaca [chd]	3,600	Mexico		severe
Khoisan, Southern	3	!Xóõ [nmn]	4,200	Botswana		definite
Tacanan	6	Reyesano [rey]	4,600	Bolivia	older adults	critical
Camsá (isolate)	1	Camsá [kbh]	4,770	Colombia		definite
Peba-Yaguan	1	Yagua [yad]	5,690	Peru		definite
Pumé (unclassified)	1	Pumé [yae]	5,840	Venezuela		definite
Chukotko-Kamchatkan	4	Chukchi [ckt]	7,740	Russian Federation		severe
Nilo-Saharan, Kuliak	2	Ik [ikx]	10,000	Uganda		severe
Cahuapanan	2	Chayahuita [cbt]	11,400	Peru		definite
Nilo-Saharan, Temein	2	Temein [teq]	13,000	Sudan		severe
Mascoian	4	Lengua [leg]	15,000	Paraguay		severe
Hatam (isolate)	1	Hatam [had]	16,000	Indonesia		critical
Iroquoian	8	Cherokee [chr]	16,400	United States		definite
Barbacoan	5	Guambiano [gum]	23,500	Colombia		definite
Warao (isolate)	1	Warao [wba]	28,100	Venezuela		definite
Altaic, Tungusic	11	Xibe [sjo]	30,000	China		severe
Ticuna (isolate)	1	Ticuna [tca]	48,600	Brazil		definite
Páez (isolate)	1	Páez [pbb]	77,400	Colombia		definite
Araucanian	2	Mapudungun [arn]	300,000	Chile		definite

TABLE 5. Linguistic stocks that are presently moribund. The LANGS column is the number of surviving languages within the stock. The remaining columns document the largest surviving language in the stock, giving its name and ISO 639 code, its estimated number of speakers (as reported in Lewis 2009), the main country where spoken, and its vitality assessments from Lewis 2009 and Moseley 2010.

3.3. ENDANGERED STOCKS. The term ENDANGERED does not have a consistent meaning across the linguistic literature. The meaning adopted by the UNESCO *Atlas* is more narrow, using the term only for situations in which the language is not normally being passed on to children. This is what Krauss (1992) referred to as moribund and that is the terminology we have followed. For Krauss, 'endangered' was a broader category. When he famously observed that 'at the rate things are going, the coming [twenty-first] century will see either the death or the doom of 90% of mankind's languages' (1992:7), he was not predicting that 90% of the world's languages would be extinct by 2100, but warning that if trends continue, 90% of the languages spoken in 1990 could be either extinct or moribund by the end of this century (and thus doomed to extinction in the twenty-second century). Thus his notion of endangerment extends to languages that are still viable at present but that are in danger of becoming moribund by the end of this century. Linguists have not universally adopted the scenario of language loss as extreme as 90%, but many warn of loss along the lines of 50% (e.g. Crystal 2000:19).

The data do not exist that can tell us which currently viable languages will no longer be viable three or four generations from now. One single data point, however, stands out as correlating well with endangerment, and that is size of the population speaking a language. Table 6 shows the relationship as demonstrated in our data set. The table groups languages by speaker population and shows what percentage of languages in each size range are currently assessed by at least one of our two sources (the *Ethnologue* and the UNESCO *Atlas*) as being moribund. The data show that 25% of all languages spoken today have been classified as moribund. The breakdown by population groups shows that virtually all languages with fewer than 100 speakers are moribund, over one third with a population in the hundreds are moribund, and nearly one fourth with a population in the single-digit thousands. At each further order-of-magnitude increase to size of population, the fraction of moribund languages continues to decrease from one eighth for populations in the tens of thousands, to one twelfth for populations in the hundreds of thousands, and to one fortieth for populations in the millions. Beyond that, no language with ten million or more speakers is assessed as being moribund. This accords well with the intuition that the larger a language is, the less likely it is to succumb to language shift. Perhaps surprising in Table 6 is the fact that seven languages with more than one million speakers are assessed by the UNESCO *Atlas* as being moribund. This could reflect a tendency to interpret 'definitely endangered' more in terms of the danger of doom discussed by Krauss (as indicated by an increasing trend toward children not learning the language), as opposed to a literal interpretation of this category's definition in the UNESCO scale as 'children no longer learn the language'.

POPULATION RANGE	TOTAL LANGUAGES	MORIBUND LANGUAGES	% MORIBUND
1 to 9	123	123	100.0%
10 to 99	335	320	95.5%
100 to 999	1,032	380	36.8%
1,000 to 9,999	2,004	473	23.6%
10,000 to 99,999	1,794	230	12.8%
100,000 to 999,999	871	67	7.7%
1,000,000 to 9,999,999	298	7	2.3%
Over 10,000,000	85	0	0.0%
Unknown	139	48	34.5%
Totals	*6,681*	*1,648*	*24.7%*

TABLE 6. Language moribundity in relation to language size.

Tables 7 through 9 treat speaker population size as a proxy indicator for endangerment to help explore the impact of further endangerment on linguistic diversity. First, we ask the question: What linguistic stocks would be most likely to be lost if 50% of languages are lost? Table 7 gives a list of linguistic stocks that are the most in danger from the standpoint of language size. The median size of the languages in the world today is 7,500 speakers (Lewis 2009:19); that is, half are larger than that and half are smaller. Table 7 lists every stock that is still assessed as viable, but in which the largest living language is smaller than the median size. Thus, if the smallest 50% of languages were to be lost, the stocks listed in Table 7 would be lost. The stocks are listed in ascending order of their largest language.

STOCK	LANGS	LARGEST LANGUAGE	SPEAKERS	COUNTRY
Quinqui (unclassified)	1	Quinqui [quq]	?	Spain
Bayono-Awbono	2	Awbono [awh]	100	Indonesia
Mbre (unclassified)	1	Mbre [mka]	200	Côte d'Ivoire
Molof (unclassified)	1	Molof [msl]	200	Indonesia
North Brazil	1	Arára, Pará [aap]	200	Brazil
Australian, Yiwaidjan	7	Maung [mph]	240	Australia
Odiai (isolate)	1	Odiai [bhf]	240	Papua New Guinea
Andamanese, South Andamanese	3	Jarawa [anq]	250	India
East Bird's Head-Sentani, Burmeso	1	Burmeso [bzu]	250	Indonesia
Tofanma (unclassified)	1	Tofanma [tlg]	250	Indonesia
Left May	2	Nakwi [nax]	280	Papua New Guinea
Murkim (unclassified)	1	Murkim [rmh]	290	Indonesia
Abinomn (isolate)	1	Abinomn [bsa]	300	Indonesia
Amto-Musan	2	Amto [amt]	300	Papua New Guinea
Elseng (isolate)	1	Elseng [mrf]	300	Indonesia
Yuwana (unclassified)	1	Yuwana [yau]	300	Venezuela
Mura	1	Pirahã [myp]	360	Brazil
South-Central Papuan, Yelmek-Maklew	2	Yelmek [jel]	400	Indonesia
Kaure	4	Kaure [bpp]	450	Indonesia
Arai-Kwomtari, Arai (Left May)	4	Ama [amm]	480	Papua New Guinea
Trans-New Guinea, Tanah Merah	1	Tanahmerah [tcm]	500	Indonesia
Lepki (unclassified)	1	Lepki [lpe]	530	Indonesia
Lower Mamberamo	2	Warembori [wsa]	600	Indonesia
Yale (isolate)	1	Yale [nce]	600	Papua New Guinea
Doso (unclassified)	1	Doso [dol]	700	Papua New Guinea
Trans-New Guinea, Yareban	5	Yareba [yrb]	750	Papua New Guinea
Yarí (unclassified)	1	Yarí [yri]	760	Colombia
Khoisan, Hatsa	1	Hadza [hts]	800	Tanzania
Trans-New Guinea, Kamula	1	Kamula [xla]	800	Papua New Guinea
Seri (isolate)	1	Seri [sei]	900	Mexico
Lakes Plain	20	Edopi [dbf]	1,000	Indonesia
Trans-New Guinea, Dem	1	Dem [dem]	1,000	Indonesia
Trans-New Guinea, Turama-Kikorian	3	Rumu [klq]	1,000	Papua New Guinea
Trans-New Guinea, East Strickland	6	Gobasi [goi]	1,100	Papua New Guinea
Nambiquaran	7	Nambikuára, Southern [nab]	1,150	Brazil
Ná-Meo (unclassified)	1	Ná-Meo [neo]	1,200	Viet Nam
Trans-New Guinea, Mombum	2	Koneraw [kdw]	1,200	Indonesia
Trans-New Guinea, West Kutubu	1	Fasu [faa]	1,200	Papua New Guinea
Mongol-Langam	3	Yaul [yla]	1,210	Papua New Guinea

(TABLE 7. *Continues*)

STOCK	LANGS	LARGEST LANGUAGE	SPEAKERS	COUNTRY
Arafundi	3	Nanubae [afk]	1,270	Papua New Guinea
Trans-New Guinea, Kwalean	3	Uare [ksj]	1,300	Papua New Guinea
Trans-New Guinea, Tirio	5	Makayam [aup]	1,300	Papua New Guinea
Maku	6	Hupdë [jup]	1,360	Brazil
Australian, Gunwingguan	24	Gunwinggu [gup]	1,410	Australia
Australian, Murrinh-Patha	2	Murrinh-Patha [mwf]	1,430	Australia
Trans-New Guinea, Manubaran	2	Doromu-Koki [kqc]	1,500	Papua New Guinea
Kiowa Tanoan	6	Tiwa, Southern [tix]	1,630	United States
Katukinan	3	Kanamarí [knm]	1,650	Brazil
Waorani (isolate)	1	Waorani [auc]	1,650	Ecuador
Senagi	2	Dera [kbv]	1,690	Indonesia
Trans-New Guinea, Inland Gulf	7	Mubami [tsx]	1,730	Papua New Guinea
Australian, Tiwian	1	Tiwi [tiw]	1,830	Australia
Bangi Me (isolate)	1	Bangi Me [dba]	2,000	Mali
Pauwasi	5	Emem [enr]	2,000	Indonesia
Piawi	2	Haruai [tmd]	2,000	Papua New Guinea
Somahai	2	Momuna [mqf]	2,000	Indonesia
Yele-West New Britain, West New Britain	2	Pele-Ata [ata]	2,000	Papua New Guinea
South-Central Papuan, Morehead-Upper Maro	17	Yei [jei]	2,390	Indonesia
Kuot (isolate)	1	Kuot [kto]	2,400	Papua New Guinea
South-Central Papuan, Pahoturi	2	Agob [kit]	2,440	Papua New Guinea
Arai-Kwomtari, Kwomtari	6	Fas [fqs]	2,500	Papua New Guinea
East Geelvink Bay	11	Barapasi [brp]	2,500	Indonesia
Nimboran	5	Gresi [grs]	2,500	Indonesia
Sulka (isolate)	1	Sulka [sua]	2,500	Papua New Guinea
Tor-Kwerba	24	Kwerba [kwe]	2,500	Indonesia
Sko	7	Vanimo [vam]	2,670	Papua New Guinea
Hokan, Yuman	9	Havasupai-Walapai-Yavapai [yuf]	2,690	United States
Trans-New Guinea, East Kutubu	2	Foi [foi]	2,800	Papua New Guinea
Malakhel (unclassified)	1	Malakhel [mld]	2,860	Afghanistan
Arauan	5	Kulina [cul]	2,940	Brazil
Abun (isolate)	1	Abun [kgr]	3,000	Indonesia
South-Central Papuan, Waia	1	Tabo [knv]	3,000	Papua New Guinea
Trans-New Guinea, Kolopom	3	Kimaghima [kig]	3,000	Indonesia
Urarina (isolate)	1	Urarina [ura]	3,000	Peru
Yuat	6	Biwat [bwm]	3,040	Papua New Guinea
Mairasi	3	Mairasi [zrs]	3,300	Indonesia
Eastern Trans-Fly	4	Wipi [gdr]	3,500	Papua New Guinea
Trans-New Guinea, South Bird's Head	10	Kokoda [xod]	3,700	Indonesia
Yele-West New Britain, Yele	1	Yele [yle]	3,750	Papua New Guinea
Australian, Pama-Nyungan	178	Arrernte, Eastern [aer]	3,820	Australia
Kol (isolate)	1	Kol [kol]	4,000	Papua New Guinea
Trans-New Guinea, Pawaian	1	Pawaia [pwa]	4,000	Papua New Guinea
Trans-New Guinea, Bosavi	9	Beami [beo]	4,200	Papua New Guinea
North Bougainville	4	Rotokas [roo]	4,320	Papua New Guinea
Border	15	Amanab [amn]	4,420	Papua New Guinea
Keres	2	Keres, Eastern [kee]	4,580	United States
Kara (unclassified)	1	Kara [kah]	4,800	Central African Republic
Warduji (unclassified)	1	Warduji [wrd]	5,000	Afghanistan
Bhatola (unclassified)	1	Bhatola [btl]	5,050	India
Tsimané (isolate)	1	Tsimané [cas]	5,320	Bolivia

(TABLE 7. *Continues*)

STOCK	LANGS	LARGEST LANGUAGE	SPEAKERS	COUNTRY
Trans-New Guinea, West Bomberai	3	Iha [ihp]	5,500	Indonesia
East New Britain	7	Qaqet [byx]	6,350	Papua New Guinea
Tucanoan	25	Cubeo [cub]	6,800	Colombia
Mpur (isolate)	1	Mpur [akc]	7,000	Indonesia

TABLE 7. Linguistic stocks that are most in danger. All languages in the stock are smaller than the median language size of 7,500 speakers. The LANGS column is the total number of living languages in the stock.

Table 8 gives a list of linguistic stocks that would be in danger if language loss were as high as 90%. In our data set, the ninetieth percentile for language size is 340,000 speakers; that is, 90% of languages are smaller than that and 10% are larger. Table 8 lists every stock that is still assessed as viable, and in which the largest living language is above the median size but below the ninetieth percentile. Thus, if the smallest 90% of languages were to be lost, these stocks would be lost.

STOCK	LANGS	LARGEST LANGUAGE	SPEAKERS	COUNTRY
Witotoan	6	Huitoto, Murui [huu]	7,800	Peru
Trans-New Guinea, Awin-Pare	2	Aekyom [awi]	8,000	Papua New Guinea
Trans-New Guinea, Oksapmin	1	Oksapmin [opm]	8,000	Papua New Guinea
Ramu-Lower Sepik	32	Angoram [aog]	8,220	Papua New Guinea
Trans-New Guinea, Mailuan	6	Mailu [mgu]	8,500	Papua New Guinea
Central Solomons	4	Bilua [blb]	8,740	Solomon Islands
Trans-New Guinea, Asmat-Kamoro	11	Asmat, Casuarina Coast [asc]	9,000	Indonesia
Trans-New Guinea, Dagan	9	Daga [dgz]	9,000	Papua New Guinea
Trans-New Guinea, Awyu-Dumut	15	Awyu, South [aws]	9,340	Indonesia
Zuni (isolate)	1	Zuni [zun]	9,650	United States
Trans-New Guinea, Kiwaian	7	Kiwai, Southern [kjd]	9,700	Papua New Guinea
Trans-New Guinea, Kayagar	3	Kayagar [kyt]	10,000	Indonesia
Trans-New Guinea, Koiarian	7	Ese [mcq]	10,000	Papua New Guinea
Trans-New Guinea, Marind	6	Yaqay [jaq]	10,000	Indonesia
Trans-New Guinea, Teberan	2	Dadibi [mps]	10,000	Papua New Guinea
Trans-New Guinea, Duna-Bogaya	2	Duna [duc]	11,000	Papua New Guinea
Trans-New Guinea, Mek	7	Nalca [nlc]	11,100	Indonesia
Muskogean	6	Choctaw [cho]	11,400	United States
Trans-New Guinea, Finisterre	40	Rawa [rwo]	11,500	Papua New Guinea
Huavean	4	Huave, San Mateo del Mar [huv]	12,000	Mexico
Salivan	3	Piaroa [pid]	12,280	Venezuela
Trans-New Guinea, Damal	1	Damal [uhn]	14,000	Indonesia
Trans-New Guinea, Goilalan	5	Fuyug [fuy]	14,000	Papua New Guinea
East Bird's Head-Sentani, East Bird's Head	3	Meyah [mej]	14,800	Indonesia
Trans-New Guinea, Wiru	1	Wiru [wiu]	15,300	Papua New Guinea
Nilo-Saharan, Kresh	2	Gbaya [krs]	16,000	Sudan
Trans-New Guinea, Huon	21	Nabak [naf]	16,000	Papua New Guinea
Torricelli	56	Bukiyip [ape]	16,200	Papua New Guinea
Yanomam	4	Yanomamö [guu]	17,640	Venezuela
Trans-New Guinea, Ok	21	Ngalum [szb]	18,000	Indonesia
Macro-Ge	32	Kaingang [kgp]	18,500	Brazil
Siouan	17	Dakota [dak]	19,280	United States
Trans-New Guinea, Madang	106	Waskia [wsk]	20,000	Papua New Guinea
Maybrat	2	Mai Brat [ayz]	20,000	Indonesia
Mataco-Guaicuru, Guaicuruan	5	Toba [tob]	21,410	Argentina

(TABLE 8. *Continues*)

STOCK	LANGS	LARGEST LANGUAGE	SPEAKERS	COUNTRY
Trans-New Guinea, Gogodala-Suki	4	Gogodala [ggw]	22,000	Papua New Guinea
Yauma (unclassified)	1	Yauma [yax]	22,200	Angola
Trans-New Guinea, Eleman	7	Toaripi [tqo]	23,000	Papua New Guinea
Mataco-Guaicuru, Mataco	7	Wichí Lhamtés Vejoz [wlv]	25,000	Argentina
Panoan	28	Shipibo-Conibo [shp]	26,000	Peru
South Bougainville	9	Terei [buo]	26,500	Papua New Guinea
Carib	31	Macushi [mbc]	29,100	Brazil
East Bird's Head-Sentani, Sentani	4	Sentani [set]	30,000	Indonesia
Mixe-Zoque	17	Popoluca, Highland [poi]	30,000	Mexico
Guahiban	5	Guahibo [guh]	34,200	Colombia
Majhwar (unclassified)	1	Majhwar [mmj]	34,300	India
Algic	44	Cree, Woods [cwd]	35,000	Canada
Trans-New Guinea, Binanderean	13	Orokaiva [okv]	35,000	Papua New Guinea
Khoisan, Sandawe	1	Sandawe [sad]	40,000	Tanzania
Sepik	56	Ambulas [abt]	44,000	Papua New Guinea
Trans-New Guinea, Angan	13	Hamtai [hmt]	45,000	Papua New Guinea
Jivaroan	4	Shuar [jiv]	46,700	Ecuador
Eskimo-Aleut	11	Inuktitut, Greenlandic [kal]	57,800	Greenland
Choco	12	Emberá, Northern [emp]	60,200	Panama
Khoisan, Northern	6	Vasekela Bushman [vaj]	61,300	Namibia
Trans-New Guinea, Kainantu-Goroka	29	Kamano [kbq]	63,200	Papua New Guinea
Nilo-Saharan, Eastern Je	4	Gaam [tbi]	67,200	Sudan
Nilo-Saharan, Tama	3	Tama [tma]	67,900	Chad
Niger-Congo, Kordofanian	24	Acheron [acz]	70,000	Sudan
Nilo-Saharan, Nyimang	2	Ama [nyi]	70,000	Sudan
Trans-New Guinea, East Timor	3	Makasae [mkz]	70,000	East Timor
Nilo-Saharan, Kadugli-Krongo	6	Katcha-Kadugli-Miri [xtc]	75,000	Sudan
West Papuan	23	Galela [gbi]	79,000	Indonesia
Nilo-Saharan, Daju	7	Daju, Dar Fur [daj]	80,000	Sudan
Nilo-Saharan, Surmic	10	Me'en [mym]	80,000	Ethiopia
Nilo-Saharan, Nara	1	Nara [nrb]	81,400	Eritrea
Burushaski (isolate)	1	Burushaski [bsk]	87,000	Pakistan
Trans-New Guinea, West Timor-Alor-Pantar	19	Bunak [bfn]	100,000	East Timor
Trans-New Guinea, Wissel Lakes	5	Ekari [ekg]	100,000	Indonesia
Totonacan	12	Totonac, Highland [tos]	120,000	Mexico
Trans-New Guinea, Chimbu-Wahgi	17	Melpa [med]	130,000	Papua New Guinea
Niger-Congo, Dogon	14	Dogon, Tomo Kan [dtm]	133,000	Mali
Chibchan	21	Ngäbere [gym]	133,090	Panama
Tarascan	2	Purepecha, Western Highland [pua]	135,000	Mexico
Nilo-Saharan, Berta	1	Berta [wti]	147,000	Ethiopia
Na-Dene	46	Navajo [nav]	149,000	United States
Nilo-Saharan, Komuz	6	Gumuz [guk]	160,000	Ethiopia
Trans-New Guinea, Engan	14	Enga [enq]	165,000	Papua New Guinea
Trans-New Guinea, Dani	13	Dani, Western [dnw]	180,000	Indonesia
Misumalpan	4	Mískito [miq]	183,000	Nicaragua
Nilo-Saharan, Kunama	1	Kunama [kun]	188,880	Eritrea
Khoisan, Central	13	Nama [naq]	251,100	Namibia
Waxianghua (unclassified)	1	Waxianghua [wxa]	300,000	China
Arawakan	59	Wayuu [guc]	305,000	Colombia

TABLE 8. Linguistic stocks that are potentially in danger. At least one language in the stock is larger than the median language size of 7,500 speakers, but none is among the largest 10% of languages with more than 340,000 speakers. The LANGS column is the total number of living languages in the stock, regardless of size.

Finally, Table 9 gives a list of linguistic stocks that are probably safe even in the case of 90% language loss. It lists every stock with at least one language that is among the top 10% of languages by population (that is, it has more than 340,000 speakers).

STOCK	LANGS	LARGEST LANGUAGE	SPEAKERS	COUNTRY
Oto-Manguean	176	Mazahua, Central [maz]	350,000	Mexico
Niger-Congo, Kru	39	Bassa [bsq]	408,730	Liberia
Uto-Aztecan	61	Nahuatl, Eastern Huasteca [nhe]	410,000	Mexico
Nilo-Saharan, Fur	2	Fur [fvr]	501,800	Sudan
Basque	1	Basque [eus]	658,960	Spain
Nilo-Saharan, Bongo-Bagirmi	41	Ngambay [sba]	896,000	Chad
Niger-Congo, Ijoid	10	Izon [ijc]	1,000,000	Nigeria
Nilo-Saharan, Nubian	11	Kenuzi-Dongola [kzh]	1,045,000	Sudan
Niger-Congo, Adamawa-Ubangi	158	Zande [zne]	1,142,000	Democratic Republic of the Congo
Afro-Asiatic, Omotic	29	Gamo-Gofa-Dawro [gmo]	1,240,000	Ethiopia
East Caucasian	29	Chechen [che]	1,341,000	Russian Federation
Hmong-Mien	38	Miao, Chuanqiandian Cluster [cqd]	1,400,000	China
West Caucasian	5	Kabardian [kbd]	1,632,500	Russian Federation
Nilo-Saharan, Central Sudanic, East	22	Lugbara [lgg]	1,637,000	Uganda
Mayan	69	K'iche', Central [quc]	1,900,000	Guatemala
Aymaran	3	Aymara, Central [ayr]	2,262,900	Bolivia
Nilo-Saharan, Songhai	8	Zarma [dje]	2,438,400	Niger
Niger-Congo, Mande	71	Bamanankan [bam]	2,772,340	Mali
Afro-Asiatic, Berber	25	Tamazight, Central Atlas [tzm]	3,150,000	Morocco
Nilo-Saharan, Saharan	9	Kanuri, Central [knc]	3,240,500	Nigeria
Altaic, Mongolic	14	Mongolian, Peripheral [mvf]	3,380,000	China
Quechuan	46	Quechua, South Bolivian [quh]	3,635,000	Bolivia
Niger-Congo, Atlantic	64	Wolof [wol]	3,976,500	Senegal
Kartvelian	5	Georgian [kat]	4,255,270	Georgia
Nilo-Saharan, Nilotic	63	Dholuo [luo]	4,410,000	Kenya
Tupi	76	Guaraní, Paraguayan [gug]	4,850,000	Paraguay
Niger-Congo, Gur	96	Mòoré [mos]	5,061,700	Burkina Faso
Niger-Congo, Kwa	79	Akan [aka]	8,300,000	Ghana
Uralic	37	Hungarian [hun]	12,501,270	Hungary
Afro-Asiatic, Cushitic	45	Somali [som]	13,871,700	Somalia
Niger-Congo, Benue-Congo	973	Yoruba [yor]	19,380,800	Nigeria
Tai-Kadai	92	Thai [tha]	20,362,390	Thailand
Afro-Asiatic, Chadic	195	Hausa [hau]	24,988,000	Nigeria
Altaic, Turkic	40	Turkish [tur]	50,750,120	Turkey
Afro-Asiatic, Semitic	78	Arabic, Egyptian Spoken [arz]	53,990,000	Egypt
Korean (isolate)	1	Korean [kor]	66,305,890	Korea, South
Austro-Asiatic	169	Vietnamese [vie]	68,634,000	Viet Nam
Dravidian	85	Telugu [tel]	69,758,890	India
Austronesian	1,257	Javanese [jav]	84,608,470	Indonesia
Japonic	12	Japanese [jpn]	122,080,100	Japan
Indo-European	437	Spanish [spa]	328,518,810	Spain
Sino-Tibetan	449	Chinese, Mandarin [cmn]	845,456,760	China

TABLE 9. Linguistic stocks that are probably safe. At least one language in the stock is among the largest 10% of languages with more than 340,000 speakers. The LANGS column is the total number of living languages in the stock, regardless of size.

3.4. SUMMARY. Taking the linguistic stock as a basic unit of linguistic diversity, the above data tables give us a way of judging the degree of linguistic diversity that is at risk in the current language endangerment crisis. Table 10 summarizes the above results by tabulating the distribution of linguistic stocks by linguistic vitality and by world area. Each stock is counted only once, so a stock that spans multiple areas is counted only with the area of its origin. For instance, as before, Austronesian is counted for Asia and Indo-European is counted for Europe. Each count is also converted to a percentage of the total for the area.

AREA	EXTINCT	MORIBUND	MOST IN DANGER	POTENTIALLY IN DANGER	PROBABLY SAFE	TOTAL
Africa	4 (7%)	12 (21%)	4 (7%)	17 (30%)	19 (34%)	56
Americas	42 (28%)	61 (41%)	16 (11%)	24 (16%)	6 (4%)	149
Asia	0	17 (22%)	31 (38%)	18 (23%)	12 (15%)	78
Europe	0	0	1 (17%)	0	5 (83%)	6
Pacific	4 (5%)	12 (14%)	42 (51%)	25 (30%)	0	83
Totals	*50 (15%)*	*102 (27%)*	*94 (25%)*	*84 (23%)*	*42 (11%)*	*372*

TABLE 10. Vitality status of linguistic stocks by world area. EXTINCT: stocks that have become extinct since 1950 (Table 4). MORIBUND: stocks in which all remaining languages are judged by at least one of our data sources as not being passed on to children (Table 5). MOST IN DANGER: stocks in which at least one language is thought to be currently viable, but all remaining languages are below the median size of 7,500 speakers (Table 7). POTENTIALLY IN DANGER: stocks that contain languages above the median size, but in which no language is in the top 10% with over 340,000 speakers (Table 8). PROBABLY SAFE: stocks that have at least one language in the largest 10% of languages (Table 9).

The global totals in Table 10 show that 40% of the linguistic diversity (as reflected by the number of linguistic stocks) present in 1950 has already suffered 'death or doom' (to echo the words of Krauss (1992:7)). If this century were to further see the doom of 50% of the world's languages, then an additional 25% of the linguistic stocks would be at risk of being lost.

Figure 1 offers a graphical summary of the relative distribution of vitality of linguistic stocks by world areas. The top bar in the chart shows the global proportions of stocks by the five categories in Table 10. The next five bars represent the five areas; they have been listed from top to bottom in order of our impressionistic judgment as worst to best. Americas is clearly the worst, with 28% of stocks extinct and another 41% currently moribund. In terms of extinct and moribund stocks, the Pacific has a total of 19%, which is not the next worst; however, we have ranked the Pacific in the second worst position because 51% of its stocks are in the vulnerable situation of having only small languages. In the next position is Asia, which has had no stocks become extinct since 1950, though 22% are currently moribund. Africa is listed next; though its current situation is worse (with 7% extinct and 21% moribund), the long-term prospects look better since 34% of its stocks are categorized as probably safe and another 30% have languages larger than the median size. Europe ranks the best in this analysis since five of the six stocks originating in its area are categorized as probably safe. Note that this is in contrast to the situation with individual languages, for which only about half of European languages are classified as safe.

4. DISCUSSION. The loss of languages reduces the range of phenomena that linguists can address and makes claims for universal features of language more and more tenuous. The adequacy of our descriptive framework is constantly challenged by work with

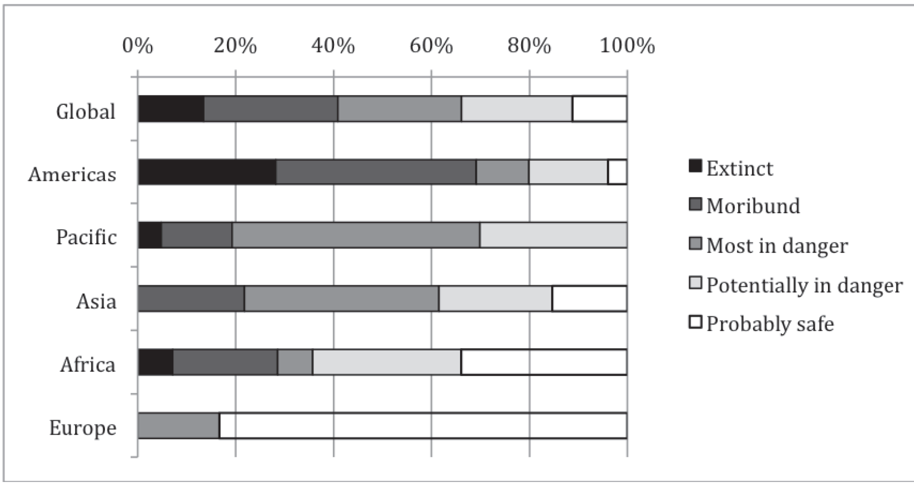

FIGURE 1. Relative distribution of linguistic stock vitality by world area.

previously undescribed languages, indicating a fundamental gap in our theorizing (Whalen 2004). The loss of an entire family can have even more extensive impact on linguistic theorizing.

The information in Tables 5 and 7 (in conjunction with the survey of available documentation in Hammarström 2010) could thus be used as a guide in setting priorities for new language documentation work. While the loss of any language without good documentation leaves a significant gap in the knowledge base of humankind, the loss of a whole linguistic stock without documentation leaves an even bigger gap.

The loss to speaker communities is felt to greater and lesser degrees. There are some languages whose speakers seem content to let them die (Ladefoged 1992). Many of today's indigenous groups, however, have come to regret decisions made by previous generations to drop their heritage language in favor of a majority language. Nonetheless, it is still true that maintaining a language as a means of everyday communication is a decision for the community to make, not outside linguists. At any rate, thorough documentation is likely to be of use to descendants as well as to future linguists, and it is still recommended for any endangered language.

The loss of a language family means the loss of important information that may shed light on the history and prehistory of a region. For instance, imagine that the entire Mayan family had disappeared before any of the writing system was deciphered. It was only through careful analysis of the workings of the descendant languages that we made any real progress on deciphering one of the richest orthographies ever invented (Coe 1992). Having none of the languages left to analyze would have left the interpretation open to the kinds of ambiguities that limit our interpretations of texts in Etruscan or Linear A (Coulmas 1991).

The loss of the last language in a family may also mean the loss of unique evidence about the range of phenomena that are possible in human language. This is particularly a concern for typologists and theorists. The story is told by Pullum (2008) of how new evidence from an endangered language falsified a universal claim he had previously made about syntax. After an extensive review of the literature in 1975, he began to teach that a basic word order of OVS never occurs in language, until one of his students demonstrated that Hixkaryana [hix] had exactly that word order (Derbyshire 1977).

Hixkaryana is an endangered Carib language of the Amazon region whose population was only about a hundred when Derbyshire began to work with them; fortunately, they have experienced an upsurge.

One intriguing class of sounds that is largely, though not exclusively, used in endangered language families is that of clicks (Ladefoged & Traill 1994). Clicks do occur in Zulu [zul], which is not endangered, and Dahalo [dal], which is an endangered language in a nonendangered family (Afro-Asiatic). Yet their extensive use and elaboration, such as found in Ju|'hoansi [ktz] (Miller 2007) or !Xóõ [nmn] (Traill 1977), occurs only in endangered families. Without these languages, it would be quite straightforward to claim that only nonlinguistic uses could be made of these sounds; such uses occur widely if not universally. The intriguing question of whether these are innovations and thus rare or stubborn remnants of an ancient sound could not be addressed without descriptive and experimental data collected from these smaller languages (Engstrand 1997, Traunmüller 2003).

The process of metathesis is not very common in phonology (Blevins & Garrett 1998), but its use for grammatical purposes is rare enough that its very existence has been in debate (e.g. Stonham 1994). Just a few cases seem to be irreducible to other processes, in Klallam [clm] (Thompson & Thompson 1969) of the endangered Salishan family, and in the (already extinct) Costanoan branch of the Penutian family (Okrand 1979). The reasons for the rarity of this process are still being worked out, but it is clear that its existence and details depend on a small collection of documentation from families that are fast falling silent.

Smaller languages can make patterns apparent that are later found to exist in subtler forms in larger languages. One example is grammatical evidentials (Aikhenvald 2004, Chafe & Nichols 1986). Every language encodes some information about how a speaker came by the facts presented, but only a few have an elaborated, mandatory morphological system that is independent of tense, aspect, and mood. For instance, in Tuyuca [tue] (Barnes 1984), a Tucanoan language with approximately 800 speakers, there is an extensive evidential system, including a division of 'firsthand knowledge' into 'visual' and 'nonvisual'.

5. CONCLUSIONS. The present results, while using available assessments of language endangerment and population estimates as a proxy when direct assessment is missing, point to a potentially severe gap in our knowledge of language families. Although new languages continue to arise (Lightfoot 2006), language families take further millennia to develop. The family for which we have the best time-depth data, Indo-European, has shown how individual dialects tend to change so much in 500–1,000 years that it is difficult to call them the same language. There is a real sense in which Latin is not dead but, rather, one of the liveliest languages around, having been succeeded by Spanish, Portuguese, French, and other widely spoken languages. None of these modern languages, however, would allow an easy two-way conversation with an ancient Roman. Language change has similarly made various descendant languages—Hindi, English, German—relatively impenetrable to their antecedents—Sanskrit, Old English, Old High German. We can be fairly certain that new languages will evolve again, even though the conditions in the modern world may not be as favorable toward diversification as they have been in the past. However, the re-evolution of today's range of language families—not just individual languages—would take tens of thousands of years.

We are thus truly perched on the edge of a cataclysmic loss of linguistic diversity. Patterns that appear only within one family are falling silent; there are probably more such patterns than we currently know about, given the shallow state of description of most endangered languages. Incorporating such results has the potential to revolutionize the field of linguistics (Whalen 2004). The contribution of linguists to document these languages will, over the coming decades, determine the extent to which future theorists will be able to fully understand the range of humanly possible languages.

REFERENCES

AIKHENVALD, ALEXANDRA Y. 2004. *Evidentiality*. Oxford: Oxford University Press.

BARNES, JANET. 1984. Evidentials in the Tuyuca verb. *International Journal of American Linguistics* 50.255–71.

BLEVINS, JULIETTE, and ANDREW GARRETT. 1998. The origins of consonant-vowel metathesis. *Language* 74.508–56.

BOWERN, CLAIRE L., and HAROLD J. KOCH (eds.) 2004. *Australian languages: Classification and the comparative method*. Amsterdam: John Benjamins.

CHAFE, WALLACE, and JOHANNA NICHOLS (eds.) 1986. *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex.

COE, MICHAEL D. 1992. *Breaking the Maya code*. London: Thames and Hudson.

COULMAS, FLORIAN. 1991. *Writing systems of the world*. Oxford: Basil Blackwell.

CRYSTAL, DAVID. 2000. *Language death*. Cambridge: Cambridge University Press.

DERBYSHIRE, DESMOND C. 1977. Word order universals and the existence of OVS languages. *Linguistic Inquiry* 8.590–99.

DESSALLE, JEAN-LOUIS. 2007. *Why we talk: The evolutionary origins of language*. Oxford: Oxford University Press.

ENGSTRAND, OLLE. 1997. Why are clicks so exclusive? *Phonum* 4.191–95.

EVANS, NICHOLAS (ed.) 2003. *The non-Pama-Nyungan languages of northern Australia: Comparative studies of the continent's most linguistically complex region*. Canberra: Pacific Linguistics.

EVANS, STEVEN N.; DON RINGE; and TANDY WARNOW. 2006. A stochastic model of language evolution that incorporates homoplasy and borrowing. *Phylogenetic methods and the prehistory of languages*, ed. by Peter Forster and Colin Renfrew, 75–90. Cambridge: McDonald Institute for Archaeological Research.

FREEDMAN, DAVID A., and WILLIAM S.-Y. WANG. 1996. Language polygenesis: A probabilistic model. *Anthropological Science* 104.131–38.

HALE, KENNETH; MICHAEL KRAUSS; LUCILLE J. WATAHOMIGIE; AKIRA Y. YAMAMOTO; COLETTE CRAIG; L. MASAYESVA JEANNE; and NORA ENGLAND. 1992. Endangered languages. *Language* 68.1–42.

HAMMARSTRÖM, HARALD. 2010. The status of the least documented language families in the world. *Language Documentation and Conservation* 4.177–212.

HOCK, HANS-HENRICH, and BRIAN D. JOSEPH. 1996. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.

HOLMAN, ERIC W.; CECIL H. BROWN; SØREN WICHMANN; ANDRÉ MÜLLER; VIVEKA VELUPILLAI; HARALD HAMMARSTRÖM; SEBASTIAN SAUPPE; HAGEN JUNG; DIK BAKKER; PAMELA BROWN; OLEG BELYAEV; MATTHIAS URBAN; ROBERT MAILHAMMER; JOHANN-MATTIS LIST; and DMITRY EGOROV. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology* 52.841–75.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2007. ISO 639-3:2007: Codes for the representation of names of languages—Part 3: Alpha-3 code for comprehensive coverage of languages. Geneva: International Organization for Standardization. Online: http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39534. Registration authority: http://www.sil.org/iso639-3/.

KRAUSS, MICHAEL. 1992. The world's languages in crisis. *Language* 68.4–10.

LADEFOGED, PETER. 1992. Another view of endangered languages. *Language* 68.809–11.

LADEFOGED, PETER, and ANTHONY TRAILL. 1994. Clicks and their accompaniments. *Journal of Phonetics* 22.33–64.

LEWIS, M. PAUL (ed.) 2009. *Ethnologue: Languages of the world*. 16th edn. Dallas: SIL International. Online: http://www.ethnologue.com.

LIGHTFOOT, DAVID W. 2006. *How new languages emerge*. Cambridge: Cambridge University Press.

MILLER, AMANDA L. 2007. Guttural vowels and guttural co-articulation in Ju|'oansi. *Journal of Phonetics* 35.56–84.
MOSELEY, CHRISTOPHER (ed.) 2010. *Atlas of the world's languages in danger*. Paris: UNESCO. Online: http://www.unesco.org/culture/languages-atlas/.
NICHOLS, JOHANNA. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
OKRAND, MARC. 1979. Metathesis in Costanoan grammar. *International Journal of American Linguistics* 45.123–30.
PULLUM, GEOFFREY. 2008. Obituary: Desmond Derbyshire (1924–2007). *LINGUIST List* 19.1. Online: http://linguistlist.org/issues/19/19-1.html.
RINGE, DON. 1999. How hard is it to match CVC-roots? *Transactions of the Philological Society* 97.213–44.
SAPIR, EDWARD. 1921. *Language*. New York: Harcourt, Brace.
STONHAM, JOHN T. 1994. *Combinatorial morphology*. Amsterdam: John Benjamins.
SUTHERLAND, WILLIAM J. 2003. Parallel extinction risk and global distribution of languages and species. *Nature* 423.276–79.
SWADESH, MORRIS. 1972. *The origin and diversification of language*. London: Routledge and Keegan Paul.
THOMPSON, LAURENCE C., and M. TERRY THOMPSON. 1969. Metathesis as a grammatical device. *International Journal of American Linguistics* 35.213–19.
TRAILL, ANTHONY. 1977. The phonological status of !Xõo clicks. *Khoisan linguistic studies*, ed. by Anthony Traill, 107–31. Johannesburg: University of Witwatersrand African Studies Institute.
TRAUNMÜLLER, HARTMUT. 2003. Clicks and the idea of a human protolanguage. *Phonum* 9.1–4.
WHALEN, D. H. 2004. How the study of endangered languages will revolutionize linguistics. *Linguistics today: Facing a greater challenge*, ed. by Piet van Sterkenburg, 321–42. Amsterdam: John Benjamins.
WHALEY, LINDSAY J. 2003. The future of native languages. *Futures* 35.961–73.

[whalen@haskins.yale.edu]
[gary_simons@sil.org]

[Received 14 July 2011;
accepted 1 September 2011]