journal of anthropological archaeology **17,** 354–374 (1998)
article no. AA980328

# Explaining Global Patterns of Language Diversity

Daniel Nettle

*Merton College, Oxford OX1 4JD, United Kingdom*

Received July 2, 1997; revision received March 23, 1998; accepted April 8, 1998

The six and a half thousand languages spoken by humankind are very unevenly distributed across the globe. Language diversity generally increases as one moves from the poles toward the equator and is very low in arid environments. Two belts of extremely high language diversity can be identified. One runs through West and Central Africa, while the other covers South and South-East Asia and the Pacific. Most of the world's languages are found in these two areas. This paper attempts to explain aspects of the global distribution of language diversity. It is proposed that a key factor influencing it has been climatic variability. Where the climate allows continuous food production throughout the year, small groups of people can be reliably self-sufficient and so populations fragment into many small languages. Where the variability of the climate is greater, the size of social network necessary for reliable subsistence is larger, and so languages tend to be more widespread. A regression analysis relating the number of languages spoken in the major tropical countries to the variability of their climates is performed and the results support the hypothesis. The geographical patterning of languages has, however, begun to be destroyed by the spread of Eurasian diseases, Eurasian people, and the world economy. 

## INTRODUCTION

The diversity of human language is one of its most intriguing features, and over the last decade scholars from several disciplines have become interested in what patterns of linguistic diversity might tell us about the human past (Cavalli-Sforza et al 1988; Nichols 1990, 1992; Renfrew 1991; Mace and Pagel 1995; Dixon 1997). At least three senses of the term linguistic diversity need to be distinguished. Firstly, there are regions of the world, such as Cameroon or Papua New Guinea, where there are very large numbers of different languages, and others where there are very few. Such diversity is the topic of this paper, and I will refer to regions with many languages as being high in *language diversity,* just as biologists refer to regions with many species as high in species diversity.

Language diversity needs to be clearly distinguished from *phylogenetic diversity* of languages, which is a matter of how many different language families or branches of language families are present.[1,*] Sometimes the two types of diversity go together, as in Papua New Guinea and the Pacific. However, there is no necessary connection between them, and there are many discrepancies. Central Africa, for example, has hundreds of different languages and so is high in language diversity. However, almost all of those languages are very closely related, belonging to the Bantu family, and so the region is low in phylogenetic diversity. I will argue in this paper that the language diversity of Latin America is quite low relative to comparable parts of Africa and Asia. However, it is generally agreed (the claims of Greenberg [1987] notwithstanding) that the languages of Latin America belong to dozens of different families, and so the phyloge-

* See Notes section at end of paper for all footnotes.



0278-4165/98 $25.00

netic diversity of the continent is high, something which has implications for our understanding of the Latin American past (Nichols 1990).

The third sense of linguistic diversity which must be distinguished is that of *structural diversity* on some linguistic parameter. For example, the languages of the world order their basic constituents in a number of different ways. There are SOV languages (Subject-Verb-Object), SVO languages (Subject-Object-Verb), VSO languages and so on. A set of languages exhibits high structural diversity in word order if many different orders are represented among them. Structural diversity will tend to be correlated with phylogenetic diversity, since where there are many different families there will often be many different structural types of language. However, the connection is not a necessary one, since there can be several structural types within one family, or, obviously, several families with the same structural type. Nonetheless Nichols (1992:250) found that high structural diversity on the basic syntactic and morphological parameters she looked at was indeed coincident with high phylogenetic diversity, both being typical of the Pacific and New World.

This paper, then, is an investigation of language diversity, which is not closely related to the other two types of linguistic diversity. Structural diversity and particularly phylogenetic diversity give us information about the evolution of human populations along the axis of time. For example, it is the lack of phylogenetic diversity amongst the languages of Central and Southern Africa which tells us that the spread of Bantu-speaking farmers to their present locations was comparatively recent.

I will argue that language diversity also carries information of interest to the anthropologist, archaeologist or historian. However, this information is not primarily about events in deep time but about the organisation of people in space. That is, where a region is dotted with many small languages, we can infer that the population has arrayed itself, socially and economically, in many small (though not necessarily isolated) communities. Where a single language is spread over the whole region, we can infer that powerful mechanisms of regional integration have been at work, and we may wish to ask what those are. This paper, then, asks the question of, what, in general, determines the size of language communities found in a human population. In Section 1, I discuss the main geographical patterns in the distribution of languages across the globe. In Section 2, I discuss the principal vectors which spread languages, and suggest that the theory developed in Nettle (1996) to explain patterns of language diversity in West Africa might explain patterns in other parts of the world too. That theory linked the size of language groups, via subsistence patterns, to climate. In Sections 3 and 4, I test the theory for all the major tropical countries by means of a regression analysis relating the number of languages spoken in each country to climatic variables. The data presented strongly support the theory.

Before turning to Section 1 and the global distribution of languages, however, a brief discussion of how languages are identified is in order. Ascertaining the size of community speaking a particular language presupposes being able to delineate it, and throughout this paper I will refer to counts of languages in particular countries. I have been able to do this because linguists routinely list and refer to entities called languages. However, the concept is not unproblematic.

There is variation in speech norms both within and between all known communities. The problem for an analysis of language diversity is distinguishing language boundaries from dialectal or idiolectal variation. The criterion usually proposed to do this is that of mutual intelligibility, and formal techniques for measuring this have been devised (Casad 1974). However, not only are such techniques rarely used in the field, but their very basis is problematic. First, intelligibility is a gradient phenomenon, and there can be various degrees of partial understanding. One may find chains of dialects in which adjacent varieties are perfectly intelligible, but those at opposite ends of the chain are definitely not. It is unclear how best to draw boundaries in such cases. Second, mutual intelligibility can be asymmetrical. Third, intelligibility varies a great deal according to the context and the particular speakers involved, and fourth, intelligibility depends upon much the parties involved want to understand each other, which is a product not of linguistic structure but of social factors (Wolff 1959; Hudson 1996:35-36).

Such arguments have lead many authors to stress the problematic nature of the term "language" (Hudson 1996:36; Romaine 1994: Chapter 1). Whilst accepting its difficulties, I would argue that we need not disregard as worthless data from field linguists on the number of languages spoken in different areas. To see why, the parallel with the problem of individuating biological species can be considered.

The science of ecology makes and tests quantitative predictions about the number of species which will be found in different environments (see, e.g., Huston 1994), yet the concept of species is scarcely less problematic than that of language. For sexual organisms, the criterion cited is that of (in)ability to interbreed, yet this criterion has all the difficulties of the (un)intelligibility criterion for languages. There are chain species, examples of partial ability to interbreed, and even asymmetrical interfertility (Mayr 1963; Sokal and Crovello 1970; Abruzzi 1982). The justification for holding on to the concept of species is that, although there are problem cases, it is a convenient way of capturing real facts about biological diversity. I would make the same case for linguistic diversity and thus justify holding on to the language concept in the following ways.

First, though there are cases of blurred boundaries, there are many others where the communicative discontinuities are quite obvious, and indeed highly salient to the people concerned. Indeed, Dixon (1997:8) argues that the clear cases are typical and the blurred ones rather rare. Second, the linguistic arbitrariness of the language/dialect distinction is relatively unimportant, if linguists employ it approximately self-consistently. As long as this is the case, then counts of what field linguists have deemed to be languages will give a reasonable relative measure of language diversity. For example, Romaine (1994: Chapter 1) points out the arbitrariness of the language boundaries usually identified in Melanesia, which do not capture the range of communicative practices of this highly multilingual population. However, she is quite clear that there is more diversity *of something to do with varieties of language* in Melanesia than in other areas, and the conventional count of languages does capture this fact, even if it fails to capture other aspects of the linguistic geography of the area. Linguists seem to agree often enough in practice about what languages are for the assumption to be made that their results from different continents are roughly comparable. Third, although error is introduced into the data by the indeterminacy of language boundaries, this error is effectively random. It is thus more likely to obscure real trends than create apparent ones, and any clear patterns which are still evident are likely to represent genuine effects. Finally, the magnitude of variation in language diversity between different regions is very large. If there was a difference in the average number of square kilometers per language in, say, Papua New Guinea and Namibia, of two- or threefold, it would be tempting to conclude that there was nothing more going on than noise in the data. However, the real difference is 80-fold (about 540 $km^2$ for Papua New Guinea, and about 40,000 $km^2$ for Namibia), and so it seems reasonable to look for underlying causes.

In what follows, then, I shall proceed to use counts of languages without further comment, though it should of course be borne in mind that these carry a large margin of error around them.

## 1. GLOBAL PATTERNS OF LANGUAGE DIVERSITY

There are many more languages spoken in the world than is commonly realised. The best available catalogue is the *Ethnologue,* compiled by the Summer Institute of Linguists, of which Grimes (1993) is a recent edition. Grimes (1993) lists 6528 living languages, as well as a number of extinct ones.

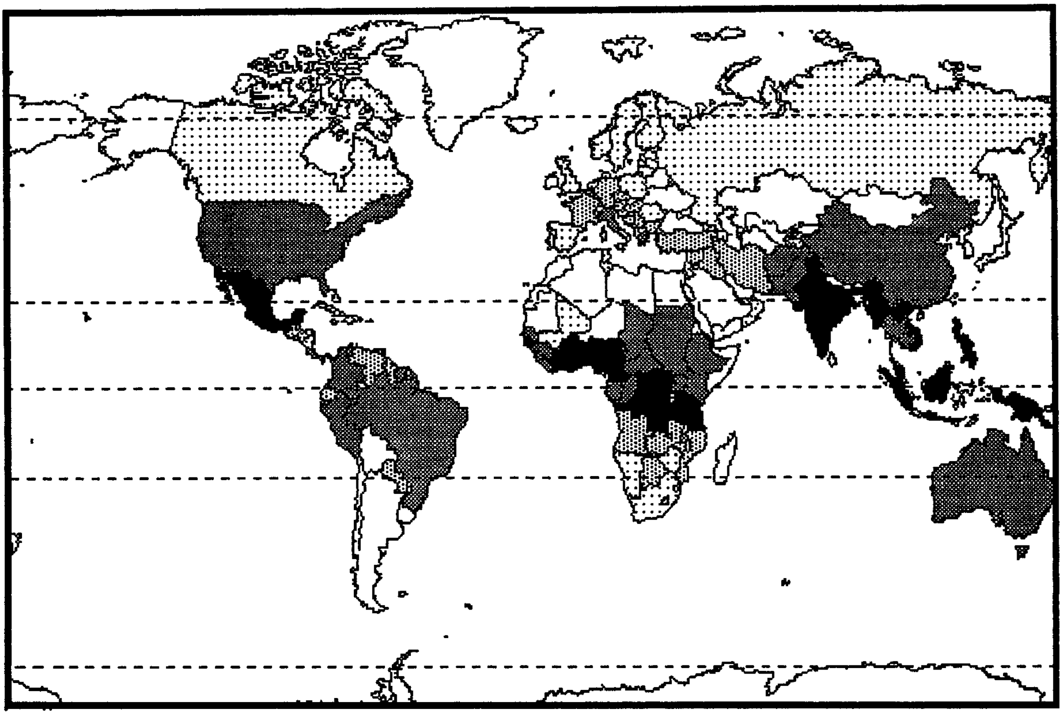
What is even more striking than the sheer number of languages is the unevenness of their distribution. Grimes (1993) lists languages by country, so, although the country is not an ideal unit of comparison, I will use it here. Eight hundred and sixty-two languages are spoken in Papua New Guinea. This is 13.2% of the global total, yet Papua New Guinea represents only 0.4% of the world's land area, and only 0.1% of the world's population live there.[2] At the other extreme, Russia has 1.5% of the languages and 15.3% of the land area. We can gain an idea of the global distribution by plotting the relative language diversity of each country, as in Map 1.

The simplest way to do this would be to divide the area of countries by the number of languages spoken and shade them according to this value. However, doing so artificially inflates the apparent diversity of smaller countries, for the following reason. Many languages cross national boundaries, and the smaller the country, the more of its languages tend to do so. For example, 21 languages are spoken in the Gambia, which is a very narrow strip of land, but all of them are also spoken in Senegal or elsewhere. Dividing the area of the Gambia by 21 would thus give an extremely misleading figure for the average area occupied by a Gambian language. I have taken two measures to minimize the small country effect. First, countries smaller than 50,000 $km^2$ have been excluded from the analysis. Second, rather than dividing the number of languages by the area of the country, I have performed a regression of the logarithm of the number of languages on the logarithm of the area of the country.[3] The number of languages does increase with increasing area ($\ln[LANGS] = 0.49 \ln[AREA] + 0.53$; $r = .45$, $df = 124$, $p < .001$), but with considerable scatter and hence a relatively low $r$ value. The relative language diversity of each country has thus been determined by the value of the standardized residual from this regression line.

The relative language diversity of all the countries of the world larger than 50,000 $km^2$ is shown on Map 1. The darker the shading, the more relatively diverse the country. Note that the resolution of the map is only at the level of the country—countries like India and Mexico which have very diverse and less diverse regions are shaded uniformly using the mean.

Several clear patterns can be observed in Map 1. First, language diversity tends to be greatest near the equator and decrease as one moves North or South away from it, as several authors have noted (Breton 1991; Nichols 1990, 1992; Mace and Pagel 1995). Species diversity decreases in a similar way as one moves away from the equator (Stevens 1989; Mace and Pagel 1995).

Second, there are more specific associ-



**MAP 1.** Map of the world showing the relative language diversity of the major countries. This is calculated by regressing the logarithm of the number of languages spoken in the country (Source: Grimes 1993) against the logarithm of the area of the country, and shading each country according to the standardised residual. The shading sheme is as follows: White (least diversity), $zres < -0.5$; Light dooting, $-0.5 < zres < 0$; Heavy hathcing: $0 < zres < 0.75$: Black (most diversity): $zres > 0.75$.

ations between language and biological diversity. In the Old World, there are two great belts of very high language diversity, which correspond almost exactly to the two great belts of equatorial forest which harbour so many of the Old World's species. One runs from Ivory Coast across West Africa into Zaire. The other runs from South India and peninsular Southeast Asia down through the Indonesian islands into New Guinea and the islands of the Western Pacific. Totalling up the languages listed in Grimes (1993) for the 17 main countries in these two belts[4] gives 3929, which is 60% of all those in the world. As well as being plentiful where species are numerous, languages are few where species are few. Large areas of white on Map 1 correspond to the Sahara and Arabian deserts, which though in the tropics are arid and poor in species.

Third, the New World distribution is slightly different from that of the Old World. There is a latitudinal pattern, but the highest diversity is not in Amazonia, as we might expect, but in Mexico. Furthermore, the overall level of diversity is lower than in the Old World. I will discuss possible reasons for this in Section 4 below.

None of these three patterns is a simple product of there being more people in the more diverse areas, as I shall show in Section 4. The orders of magnitude of the differences are also very large. We have, therefore, some strong trends to explain: language diversity is inversely related to latitude, is low in deserts and arid places, is high in the environment that produces equatorial forest, and is relatively low and anomalously distributed in the New World. In the next section, I briefly discuss the vectors by which languages are spread and outline the ecological theory from Nettle (1996) which, I will argue, may account for the observed patterns.

## 2. VECTORS OF LANGUAGE SPREAD

As we have seen, the vast majority of all languages are spoken by small communities in the tropics. Most of these communities are primarily dependent on non-industrial ways of making their living, or have been until very recently. Few languages are routinely written down, still less used in television or radio. Very few are the official vehicle of any nation-state or other large-scale political élite. In this section, then, I will suggest and briefly justify a number of generalizations which can be drawn about the "typical" case of how a language develops. These are, first, that it is transmitted orally; second, that it passes through primary social bonds; third, that state policies are not usually relevant; and fourth, that it is encapsulated in the wider system of economic life. I will now explain each of these generalisations in turn.

First, language transmission is basically oral and informal. This means that linguistic norms pass between individuals who interact face-to-face. It seems to be part of the nature of linguistic transmission that, in the absence of positive interaction, people's languages diverge. It follows that people who maintain the same speech norms must be connected by social or economic ties, and that lack of face-to-face interaction between groups will tend to mean that their languages diverge.

Second, however, the mere presence of interaction between people is no guarantee of their linguistic homogeneity. Sociolinguistic differences can persist despite routine interaction between the people involved. Trade between groups can take place via *lingua francas,* or pidgins, or bilingual individuals, while the underlying differences between their first languages are undiminished.

What determines whether groups will converge or remain uniform linguistically is a matter of *social identification* (LePage 1968). Where individuals are closely involved enough to want to be identified with each other, they will mutually accommodate their speech behavior. However, they may also interact while actively maintaining ethnolinguistic distance, and the question of what determines whether they do so or not is an important one.

The answer seems to lie in the strength and nature of the social bonds involved (Milroy 1980). It is convenient in this context to distinguish between *primary* and *secondary* social bonds. Primary social bonds are relatively enduring, are often formed early in life, and are multivalent. This means that they are not formed for any one specific purpose. Rather they are generalized social linkages (Sahlins 1972) which may bring the actors together in many different contexts and at different times. Social bonds within ethnolinguistic groups in non-industrial societies typically appear to have this character; they form a dense web of relationships, which are often reinforced by biological and cultural kinship, and which may form the basis for common ritual activities and celebrations, common defence, common farm work or hunting, gifts, and food sharing, as well as trade narrowly defined. In most societies, these primary bonds have also been those on which people have depended for their basic livelihood.

Secondary social bonds, by contrast, are based on specific functions, such as a trade in a specialized good like salt or metal, or a specialized service. Purely commercial trade creates only a secondary bond, as reciprocation tends to be immediate, and no future moral responsibility for the other party is entailed (this is Sahlins' [1972] "balanced reciprocity"). Secondary bonds are associated with greater social distance than primary ones and are more typical of the relationships between ethnolinguistic groups than those within them. I would argue that secondary bonds are not themselves sufficient for sociolinguistic identification. Thus, whether social interaction will lead to the adoption of a common mother tongue is determined whether the social bonds created are primary or secondary.

Like almost all dichotomies, the distinction between primary and secondary bonds simplifies an underlying continuum. Pairings of farming and hunting or herding peoples give a good example of this. The pygmies of central Africa, for example, are specialized hunters, who nonetheless consume a great deal of their diet in cereals grown by neighboring farmers. It is unlikely that they could survive reliably by hunting and gathering alone (Bailey et al. 1989), and so this relationship is of vital importance to them, although it is in origin a single-purpose rather a multivalent one. Each pygmy group is paired with a particular group of farmers. The interdependence is so great that, in each case, the pygmies no have accommodated to the farmers and no longer have a distinctive language.

Further south in Africa, in the Kalahari, San hunter-gatherers also depend upon exchange with neighboring farmers, yet have remained linguistically distinct (Headland and Reid 1989). In the West African savanna, one finds widespread symbiosis between Fulani pastoralists and cereal farmers, particularly the Hausa. Even the most purely pastoral Fulani consume a high proportion of their calories as cereals which they have traded for milk and other products (Swift 1986:180). However, the distinction between Hausa and Fulani remains, and a relatively small proportion of those Fulani who are still herders have adopted Hausa as a mother tongue (though most speak it as a lingua franca). The difference in outcome from the pygmy case may reflect a difference in the extent of investment in the particular exchange relationship in the different cases. Fulani pastoral networks are extremely extensive, and go well beyond Hausaland, and the San seem to have both a wide choice of trading partners and great flexibility in their own subsistence arrangements. Neither San nor Fulani households are, then, as irreversibly attached to their particular agriculturalist partners than the pygmies, and so they have retained their distinct mother tongues.

Accepting that the distinction is somewhat simplistic, then, we can nonetheless assert that, in general, people come to have the same first language as those to whom they are linked by primary bonds in the sense I have described.

The third generalization about language transmission is that the influence of specialized political and governmental structures is of limited importance, at least with regard to mother tongues. There are obviously some cases of the domain of a language being determined by the extent of a particular political formation, such as the spread of vernacular Latin through Europe as part of Roman expansion. However, there are many more cases where great empires have controlled areas for long periods of time without ordinary country people coming to speak the language of the élite at all, or only learning it as a secondary tongue for use in dealing with government. The overwhelming majority of languages even today are unofficial, minority tongues, and most persist despite the incorporation of their speakers in wider state, religious and regional systems. Even in densely populated, early industrialized Europe, "the virtual identify of language, state and nation was approximated only in the nineteenth century" (Coulmas 1992:33), and in many European countries minority languages are still spoken. Overall, it is reasonable to conclude that the determinants of which first language one learns are usually based in very quotidian and local social situations, rather than in courts or parliaments.

The final generalization about language transmission is that the spread of a language is rooted in an economic system. It might appear that the choice of primary social associates is a purely social or cultural matter. However, in non-industrial societies where relatively little of the circulation of food and labor passes through the wages and money, there is no real distinction between economy and society. An individual's social associates also tend to be those with whom he labors in the fields; or from whom he borrows land, livestock, or seeds; or with whom he combines to appropriate and defend land. Bonds which are social in character are cemented by real exchanges of food and services, and it is to social associates that people turn in time of shortage (Colson 1979; for ethnographic examples, see e.g. Watts 1983; Cashdan 1985; Legge 1989). It seems, in general, that languages are rooted in networks of social bonds which have a real economic importance. This is a key assumption of the ecological theory which I will advance below.

We have now discussed in general terms the vectors by which languages are spread. Clearly, there are a small number of languages which present a very different profile to that outlined above. For example, the last 500 years has seen a small number of Eurasian languages spread widely, first in Eurasia, and then across the globe, by élite political, technological, and demographic dominance. I would argue that these events are rather untypical of language evolution as a whole. This is not because there have not been local language spreads at other times and other places; there always have, as polities waxed and waned. The difference is the scale of the European spreads, which was quite unprecedented. These recent spreads have until very recently affected only a fairly small proportion of the world's languages, since in most areas, indigenous languages have persisted alongside the colonizers. The Americas are an exception, which we discuss further in Sections 4 and 5. Perhaps a few dozen of the six and a half thousand languages were spread to their present range by recent élite dominance, with a few hundred already obliterated by their expansion. For the great bulk of the remainder, my earlier assertion that more local factors should be considered seems justified.

Thus we can return to our basic task of explaining the global distribution shown in Map 1. It is clear from the foregoing discussion that the formation of any particular ethnolinguistic group will be a complex interplay of many locally specific factors; formation of social bonds will depend upon precise topographical, military, epidemiological, demographic, and cultural situations, as well as more nebulous contingencies such as the rise and fall of local prestige and influence. However, I believe general explanations are appropriate for the global trends. The main patterns—the latitudinal gradient, the parallel with species diversity—pattern strongly with biological or ecological phenomena, which suggests that the appropriate theory will be one that links human agents to their ecological setting.

Individuals in any non-industrial society have to simultaneously solve many different ecological problems, from coping with disease, to providing fuel, fertilizer, and drinking water, to optimizing intergroup relations and population pressure. All of these may influence the evolution of their social systems. In my West

African study (Nettle 1996), I made out a case that one environmental factor could be identified which had a preeminent influence on many aspects of life, especially the formation of linguistic groups. I called that factor *ecological risk*.

Ecological risk is the amount of variation which people face in their food supply over time. Variation can occur both seasonally (within a year) and inter-annually, with periodic years of shortage. Households cannot operate for more than a few days without food and so must develop mechanisms to deal with temporary shortfalls in its supply.

There is a large literature on the mechanisms by which people cope with risk (e.g., Cashdan 1985; Minnis 1985; Huss-Ashmore, Curry and Hitchcock 1988; DeGarine and Harrison 1988; Halstead and O'Shea 1989; Cashdan 1990). Four key strategies may be identified ([Halstead and O'Shea 1989:3]; Colson [1979:21] divides them into five rather than four but her schema is basically the same). These are diversification of productive and income-generating activities, storage of food, mobility, and exchange.

In the current context, I will discuss only the mechanism most relevant to language diversity, that of exchange. In Nettle (1996), I argued that patterns of social exchange, which determined language spread, were related to patterns of rainfall. Where rainfall is continuous through the year, households can reliably produce food all year round and so are little dependent on exchange outside the immediate vicinity for subsistence purposes. Where small groups are basically self-sufficient, there is no incentive for them to form networks of primary bonds outside the immediate vicinity, and so many small languages evolve. Where there is a marked dry season or probability of a drought year, on the other hand, households form social ties over a wide area, to gain access to resources elsewhere in time of local shortage. Here, the languages are much more widespread, as the extensive social networks spread linguistic norms over a wider area. In the extreme case of the northern savanna of West Africa, pastoralists cope with between the brief and erratic rainy seasons by moving, but also by exchange of stock and services through "wide networks of kinsmen throughout West Africa" (Berg 1976:24; cited in Legge 1989). Correspondingly, languages like Fulfulde are spread over millions of square kilometers.

The theory which emerged from the discussion in Nettle (1996) was, then, the following: the greater the ecological risk, the larger the social networks people must form to ensure a reliable supply of basic subsistence products. Since linguistic norms are spread through those same social networks, it follows that the average size of a language group should also increase as ecological risk increases. This prediction was tested using a regression analysis and was found to hold true both when the size of a language was estimated in terms of area occupied and when it was estimated in terms of number of speakers.

In this paper, I test whether the predictions of the ecological risk theory hold for other parts of the world as well as West Africa. I use the country as the basic unit of analysis. This is convenient because the best available dataset on global language diversity (Grimes 1993) is organized by country. However, in many ways, countries are troublesome units of comparison. First, they are not all the same size. Second, they are not ecologically homogeneous, and, third, languages overlap their boundaries. All of these problems have to be taken into consideration in the statistical analysis, as I will describe in Section 3 below.

In Nettle (1996), the magnitude of ecological risk for each location was measured by calculating the number of months in the year in which enough rain falls for useful plant growth to occur using the Growing Season formula (LeHouérou 1989), which is described in Section 3 below. The formula captures the extent to which biological production is variable within a year. It may also be a reasonable proxy for how much it varies between years, since locations with shorter rainy seasons tend to have more erratic rainfall than those with longer ones. The formula is also used here, and an average value of the Growing Season (the Mean Growing Season or MGS) is obtained for each country using climatological records. The MGS can vary from 0 (no months in which there is adequate rainfall for food production, hence no possibility of self-sufficiency and extremely high risk) to 12 months (continuous rainfall and hence food production all year round, very low risk). The Growing Season formula takes account of rainfall and temperature. The assumption implicit in its use is thus that these are the main limiting factors on food production. This is certainly a simplification. It may, however, be reasonably accurate at least for the tropics. In the temperate latitudes, variation in day length and the presence of frost become more important, and for this reason as well as for some others given in Section 3, only tropical countries are considered in the regression analysis.

Extensive social networks are effective in mitigating risk for two slightly different reasons, and these give rise to two different hypotheses to be tested. First, increasing the spatial extent of a social network gives individuals in it access to more different microenvironments and types of land, which may have food products available at different times. This *spatial averaging* of ecological risk leads to the prediction that the spatial extent of language groups should increase as the degree of ecological risk they are exposed to increases. Converting this into the number of languages per country gives the first hypothesis to be tested:

Hypothesis 1. The longer the Mean Growing Season, the more languages there will be spoken in a country of a given area.

Second, increasing the number of productive households in a social network decreases the variation experienced by all of them as a simple consequence of the law of large numbers. This is true as long as there is some degree of statistical independence between them. This *numerical averaging* leads to the prediction that the number of individuals in a language group should increase as the degree of ecological risk increases. Converting this into the relationship between Mean Growing Season and languages per country gives the second hypothesis.

Hypothesis 2. The longer the Mean Growing Season, the more languages there will be spoken in a country of a given population.

The two hypotheses would only be equivalent in a world where population density was the same everywhere. I will therefore test them separately. In Section 3, I give the details of the methods used to do this. In Section 4, I give the results.

## 3. TESTING THE THEORY: METHODS

Ecological and linguistic data were collected for all the countries of the world over 50,000 km$^2$. Smaller countries were excluded because the problem of languages crossing national boundaries inflates their apparent diversity, as described in Section 1 above. Exceptions were made for the Solomon Islands and Vanuatu, since, though both these countries are small, they are rich in languages none of which crosses their borders as they are islands.

The analysis was restricted to countries falling wholly or mainly within the tropical zone. This is strictly defined as the area between the tropics of Capricorn and Cancer, but for the present purposes, the broader definition of the area between 30°N and 30°S is preferable, as it captures more of the area of the climatic regimes we know as "equatorial" and "tropical." The analysis has not been extended to the temperate latitudes, for several reasons. First, the available measure of ecological risk (the Mean Growing Season) takes account only of rainfall and average temperature, and thus cease to realistically reflect the possibilities of food production as one moves out of the tropical zone. Second, the economies of the tropical countries are more rural and more dominated by subsistence activities than those of most of the temperate countries. The linguistic map of, say, European countries, reflects to a considerable extent the rise of standard national languages since the industrial revolution, rather than longer term ecological processes.

The Growing Season formula (Le-Houérou 1989) was employed as a measure of ecological risk. A month is included in the growing season if the average daily temperature is more than 6°C and the total precipitation in millimeters is more than twice the average temperature in centigrade. The Growing Season is an inverse measure of ecological risk: the more growing months there are, the less the ecological risk.

The Mean Growing Season (henceforth MGS) was found for each country using meteorological records in Wernstadt (1972). The number of weather stations per country varies from 100 to over 200, and the time period for which information is available varies from a few years to 50 or more. The stations are designed to be distributed evenly around each country. Since countries are not ecological units, and are often large enough to span many ecological regimes, there is in some cases a danger of producing a meaningless "average" climate which does not correspond to that experienced by any of its communities. As a safeguard against this, countries where the standard deviation of the growing seasons from the different weather stations was greater than 2 months were excluded. This affected 19 countries, which are marked with an asterisk in the Appendix.

The number of living languages spoken by a resident community in each country (LANGS) was obtained from the *Ethnologue* 12th edition (Grimes 1993). For the reasons given in Section 1, simply dividing the area or population of each country by the number of languages exaggerates the diversity of smaller countries. I have instead taken the area of the country (for Hypothesis 1) or its population (for Hypothesis 2) as one independent variable in the multiple regression analysis, the Mean Growing Season being the other.

A number of countries were included in the sample used in Nettle (1996). As this study is intended to independently test the applicability of the theory to other areas, those countries were also excluded from most of the analyses. The 13 affected countries are marked with a dagger (†) in the Appendix.

There are two different linguistic variables of interest, corresponding to the two different hypotheses. One is the number of languages spoken relative to the country's area, the other the number of languages relative to the number of inhabitants. To calculate these variables, figures for both the area (AREA) and the population (POP) of each country were obtained from the UN Demographic Yearbook. The population figures are mid-year estimates for 1991.

To reduce skewness and kurtosis. logarithms were taken of the LANGS, AREA and POP variables. The skewness and kurtosis of the MGS did not differ significantly from 0, and so it was left untransformed.

Obviously, both the ecological and linguistic variables are highly approximate, the former because of the simplicity of the formula and the unevenness of the data, the latter because of the inherent difficulties involved in counting languages. We should thus expect at best an approximate relationship.

## 4. TESTING THE THEORY: RESULTS

The appendix displays the number of living languages (LANGS), population (POP), area (AREA), number of complete weather station records, and Mean Growing Season (MGS) for 74 countries, including, for the sake of completeness, the West African countries and those with variable climates, which are excluded from the regressions unless otherwise stated.

> Hypothesis 1. The longer the Mean Growing Season, the more languages there will be spoken in a country of a given area.

Multiple regression shows that this is indeed the case. The equation produced is:

$$\ln[\text{LANGS}] = 0.58 \ln[\text{AREA}] + 0.23\ \text{MGS} - 5.32$$

(43 countries: $r = .66$, $df = 40$, $p < .001$)

The relationship between Mean Growing Season and the number of languages once the area of the country has been controlled for is shown in Fig. 1. If the West African countries are included, it is obvious that they sit in the same distribution. Indeed, the regression equation including them is virtually the same:

$$\ln[\text{LANGS}] = 0.53 \ln[\text{AREA}] + 0.24\ \text{MGS} - 4.77$$

(55 countries: $r = .65$, $df = 52$, $p < .001$)

It is also immediately apparent that all the Latin American countries (including the Caribbean) fall well below the rest of the distribution. There are a number of possible reasons for this. One is simply sparsity of pre-colonial population. Even now only a small fraction of the land in tropical South America is populated (less than 5% according to Partridge 1989:5). It is therefore no surprise that there are fewer groups than might be expected. Mexico, where diversity is the highest in the Americas, was also more thoroughly peopled at the time of European expansion than areas further South. Perhaps a stronger reason for low diversity in the Americas is that so many indigenous languages have been lost. There are a number whose demise is known about, but many more who must have disappeared without record, principally through the infectious diseases which ravaged the continent in the decades following European contact (Crosby 1986). In Cuba, the most extreme outlier in Fig. 1, the indigenous population was decimated within a very few years of European contact (Niddrie 1971:82), and the precolonial linguistic situation is largely unknown.

If the Latin American countries are excluded, the regression relationship is much improved (Fig. 2):

$$\ln[\text{LANGS}] = 0.42 \ln[\text{AREA}] + 0.30\ \text{MGS} - 3.48$$

(36 countries: $r = .82$, $df = 33$, $p < .001$)

The relationship is not just a product of comparing different continents. Significant relationships are also found within the Asia/Pacific and African groups. The data from Latin America are too few to obtain an independent multiple regression.

Asia/Pacific:

$$\ln[\text{LANGS}] = 0.49 \ln[\text{AREA}] + 0.32\ \text{MGS} - 4.22$$

(18 countries: $r = .87$, $df = 15$, $p < .001$)

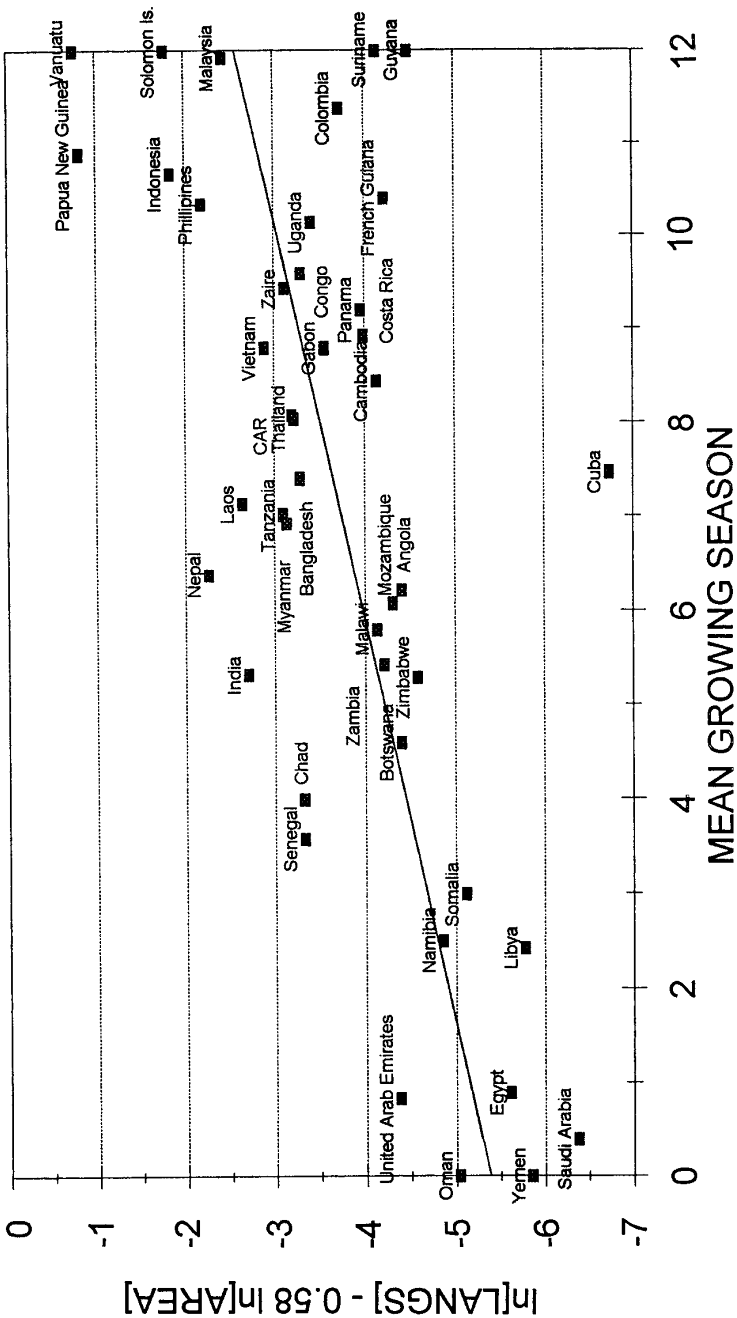



**FIG. 1.** The number of languages in each country regressed against the mean growing season in months once the effects of the country's size have been controlled for.

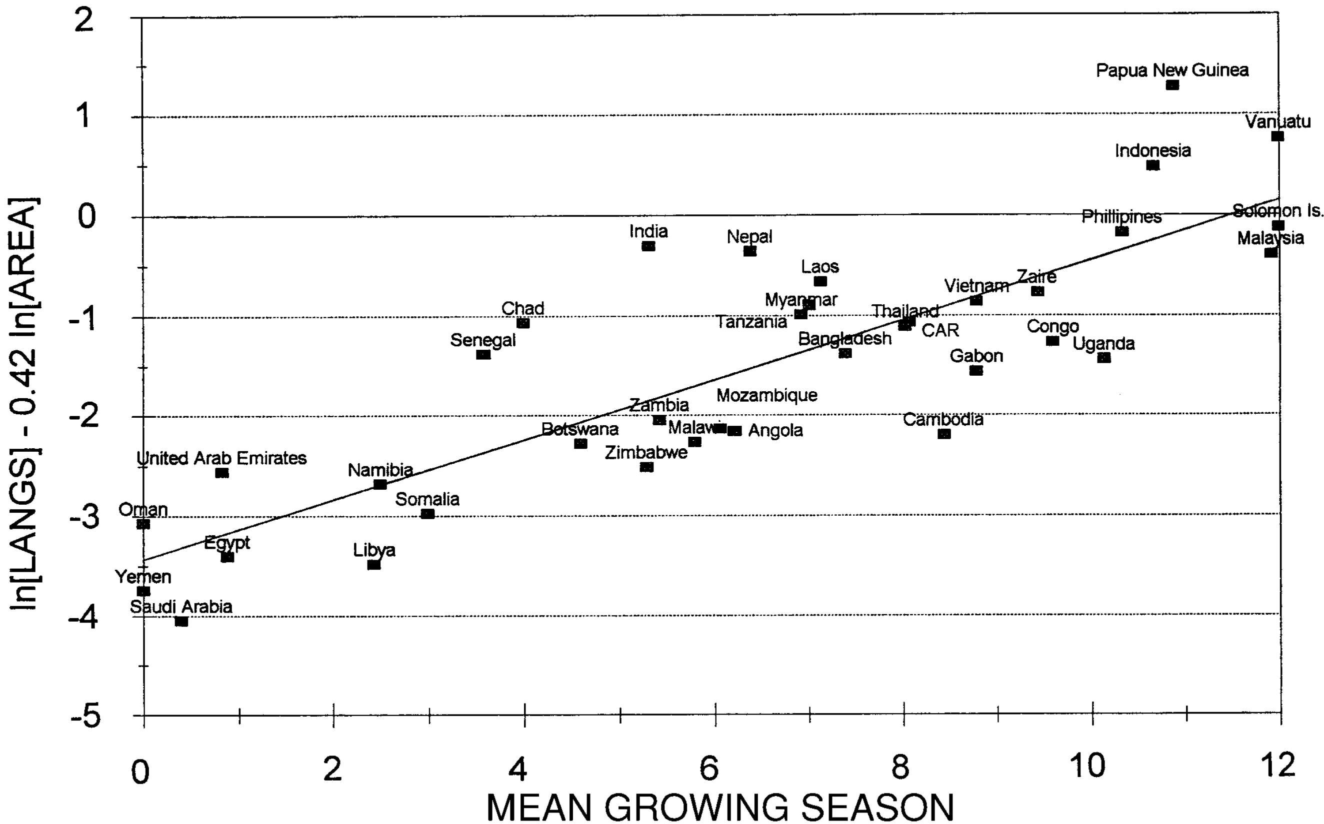



**FIG. 2.** The number of languages in each country regressed against the mean growing season in months once the effects of the country's size have been controlled for, with the Latin American countries excluded.

Africa: (West African countries excluded)

$$\ln[\text{LANGS}] = 0.50 \ln[\text{AREA}] + 0.23\ \text{MGS} - 4.37$$

(18 countries: $r = .77$, $df = 15$, $p < .01$)

The remarkable similarity of the independent equations from the two continental regions show what a powerful and universal determinant ecological risk is.

> Hypothesis 2. The longer the Mean Growing Season, the more languages there will be spoken in a country of a given population.

Multiple regression again confirms that this is the case. The equation is:

$$\ln[\text{LANGS}] = 0.33 \ln[\text{POP}] + 0.18\ \text{MGS} - 0.62$$

(43 countries: $r = .63$, $df = 40$, $p < .001$).

Again, the WestAfrican countries sit in the same distribution. Including them gives the equation:

$$\ln[\text{LANGS}] = 0.34 \ln[\text{POP}] + 0.18\ \text{MGS} - 0.65$$

(55 countries: $r = .64$, $df = 52$, $p < .001$)

Once again, excluding the Latin American countries improves the relationship.

$$\ln[\text{LANGS}] = 0.25 \ln[\text{POP}] + 0.22\text{MGS} - 0.32$$

(36 countries: $r = .78$, $df = 33$, $p < .001$)

Once again, the relationship holds independently in the two major continents:

Asia/Pacific:

$$\ln[\text{LANGS}] = 0.24 \ln[\text{POP}] + 0.26\ \text{MGS} + 0.02$$

(18 countries: $r = .82$, $df = 15$, $p < .001$)

Africa: (West African countries excluded)

$$\ln[\text{LANGS}] = 0.10 \ln[\text{POP}] + 0.20\ \text{MGS} + 1.55$$

(18 countries: $r = .64$, $df = 15$, $p < .05$)

The predictions of both ecological risk hypotheses are clearly met for the whole of the tropical world. If the previously excluded countries with very variable growing seasons are included, the significance of the relationships is not changed, the $r$ values being slightly lowered (Hypothesis 1: $r = 0.61$, $df = 58$, $p < .001$; Hypothesis 2: $r = .54$, $df = 58$, $p < .001$; Latin America included, West African countries excluded in both cases).

The results of this analysis show a consistently strong relationship between language diversity and climatic patterns. They suggest that ecological risk has been an extremely important factor—probably the most important single factor—in people's strategies of group formation and communication in the tropical world as a whole. The $r$ values using the two different hypotheses do not differ significantly, so it is not possible to conclude that either the spatial or the numerical process is more important. Usually, they go hand in hand.

## 5. CONCLUSIONS

It seems, then, that a key determinant of linguistic diversity is to be found in the basic facts of ecology and subsistence. There are many other factors, too, which we have not considered here and no doubt contribute greatly to the observed variance. A more sophisticated analysis would incorporate such factors as relief, epidemiology and population density. It would also consider the influence of different subsistence systems on the relationship of social network and ecology. Access to marine resources, for example, allows communities to diversify when faced with seasonal shortages. Livestock herding also functions as a form of protein storage, and the available species of stock varied widely between areas in traditional societies (Diamond 1997a: Chapter 9). Nonetheless, given the unavoidably approximate nature of the data used, the correlations are impressive and testify to a strong influence of ecology on the history of human communities.

The assumption of this paper has been that the current pattern of language diversity represents some kind of ecological equilibrium. This assumption seems largely justified by the correlations obtained. However, it should be remembered that the equilibrium observed is a product of a particular historical period, since almost all of the societies on which the data set is based are, or were until recently, subsistence farmers and herders. The pattern of language diversity was no doubt very different when the continents were covered with hunter-gatherers and will again be very different as the changes associated with the modern world economy become more widespread.

Before the transition from hunting and gathering to farming, the amount of language diversity may have been even greater than it is now, since hunter-gatherer communities tend to be smaller than farming communities living in similar habitats. In the Kalahari, San hunter-gatherer groups number a few thousand while their Tswana farmer neighbors are two million (Grimes 1993). In aboriginal Australia, communities never exceeded about 500 individuals, though their territory sizes varied according to the ecological productivity of the area (Birdsell 1953). Thus even in arid areas, very large communities did not evolve. This difference of outcome may reflect the different subsistence strategies of farmers and foragers; foragers respond to ecological variability primarily by moving to alternative resources. Farmers, with their heavy investment in land, cannot so easily move and so must use large social networks to bring the resources to them.

As agriculture spread out from its centers of development, it pushed waves of linguistic and demographic homogeneity before it (Renfrew 1991). This process was particularly clear in Africa, with the Bantu expansion, and Eurasia, with the Indo-European, Sino-Tibetan and other spreads (Diamond 1997a, 1997b). The languages of the spreading farmers eventually broke up into daughters whose number was determined by the ecological regime of the area, but it is likely that the number of new languages in many areas was less than the number of hunter-gatherer languages which had been subsumed or displaced. The number of languages in the world may thus have been lowered at the onset of the Neolithic, when the global pattern we have detected in this paper became established.

The global pattern is currently undergoing another transformation. This is caused by the economic, technological, and demographic take-off of Eurasian populations and the consequent spread of Eurasian peoples, crops, diseases, and languages to the other continents. The causes and dynamics of this expansion are well beyond the scope of this paper, but it is clear that their net effect will be to greatly reduce the world's language diversity; one recent projection suggested that 90% of living languages are threatened with extinction in the next century (Pinker 1994:259).

In some continents, such as Australia and the Americas, the Eurasian expansion has involved a demographic and linguistic replacement of indigenous peoples, who mainly succumbed to infectious disease, with Eurasian daughter populations (Crosby 1986). In South America, the indigenous diversity has already been lost, which is why South America fails to pattern with the other continents in Fig. 1. In North America and Australia, most of the languages had, as of 1993, a few elderly speakers, so they still appear in the data set of this paper, but they are not being transmitted to new generations and so will disappear rapidly over the next 20 years (Krauss 1992; Dixon 1997).

In other continents where Eurasians never settled in numbers, such as tropical Africa, language diversity is still being lost, though in this case it is due to linguistic replacement without corresponding demographic replacement. English and French, along with the larger and more dominant indigenous languages such as Swahili, are spreading at the expense of local languages as more people are sucked into the larger social networks associated with the modern world economy. Similar processes can be observed in the Pacific, with the spread of Tok Pisin, Tagalog, and Bahasa alongside English. What the ultimate results of these changes in human ecology on the world's language diversity will be it is as yet impossible to say.

## ACKNOWLEDGMENTS

This research was carried out in the department of Anthropology at University College London, and generously funded by the Medical Research Council. I am grateful to Leslie Aiello, Robin Dunbar, Ruth Mace, Daisy Williamson and many others for advice and assistance. All remaining errors are of course my own.

## APPENDIX

| Country | Languages | Area[a] | Population[b] | Stations[c] | MGS[d] | *SD*(GS)[e] |
|---|---|---|---|---|---|---|
| Algeria* | 18 | 2381741 | 25660 | 102 | 6.60 | 2.29 |
| Angola | 42 | 1246700 | 10303 | 50 | 6.22 | 1.87 |
| Australia* | 234 | 7713364 | 17336 | 134 | 6.00 | 4.17 |
| Bangladesh | 37 | 143998 | 118745 | 20 | 7.40 | 0.73 |
| Benin† | 52 | 112622 | 4889 | 7 | 7.14 | 0.99 |
| Bolivia* | 38 | 1098581 | 7612 | 48 | 6.92 | 2.50 |
| Botswana | 27 | 581730 | 1348 | 10 | 4.60 | 1.69 |
| Brazil* | 209 | 8511965 | 153322 | 245 | 9.71 | 5.87 |
| Burkina Faso† | 75 | 274000 | 9242 | 6 | 5.17 | 1.07 |
| C.A.R. | 94 | 622984 | 3127 | 13 | 8.08 | 1.21 |
| Cambodia | 18 | 181035 | 8442 | 9 | 8.44 | 0.50 |
| Cameroon† | 275 | 475422 | 12239 | 35 | 9.17 | 1.75 |
| Chad | 126 | 1284000 | 5819 | 11 | 4.00 | 1.81 |
| Colombia | 79 | 1138914 | 33613 | 35 | 11.37 | 1.37 |
| Congo | 60 | 342000 | 2346 | 10 | 9.60 | 1.69 |
| Costa Rica | 10 | 51100 | 3064 | 38 | 8.92 | 1.78 |
| Cote d'Ivoire† | 75 | 322463 | 12464 | 9 | 8.67 | 1.25 |
| Cuba | 1 | 110861 | 10736 | 13 | 7.46 | 1.55 |
| Ecuador* | 22 | 283561 | 10851 | 44 | 8.14 | 3.47 |
| Egypt | 11 | 1001449 | 54688 | 50 | 0.89 | 0.89 |
| Ethiopia* | 112 | 1221900 | 53383 | 36 | 7.28 | 3.10 |
| French Guiana | 11 | 90000 | 102 | 5 | 10.40 | 0.80 |
| Gabon | 40 | 267667 | 1212 | 14 | 8.79 | 0.77 |
| Ghana† | 73 | 238553 | 15509 | 28 | 8.79 | 1.68 |

**APPENDIX**—*Continued*

| Country | Languages | Area[a] | Population[b] | Stations[c] | MGS[d] | *SD*(GS)[e] |
|---|---|---|---|---|---|---|
| Guatemala* | 52 | 108889 | 9467 | 59 | 9.31 | 2.23 |
| Guinea† | 29 | 245857 | 5931 | 8 | 7.38 | 1.22 |
| Guyana | 14 | 214969 | 800 | 5 | 12.00 | 0.00 |
| Honduras* | 9 | 112088 | 5265 | 13 | 8.54 | 2.53 |
| India | 405 | 3287590 | 849638 | 218 | 5.32 | 1.92 |
| Indonesia | 701 | 1904569 | 187765 | 58 | 10.67 | 1.82 |
| Kenya* | 58 | 580367 | 25905 | 34 | 7.26 | 3.61 |
| Laos | 93 | 236800 | 4262 | 7 | 7.14 | 0.35 |
| Liberia† | 34 | 111369 | 2705 | 21 | 10.62 | 0.84 |
| Libya | 13 | 1759540 | 4712 | 54 | 2.43 | 1.60 |
| Madagascar* | 4 | 587041 | 11493 | 81 | 7.33 | 2.96 |
| Malawi | 14 | 118484 | 8556 | 20 | 5.80 | 1.50 |
| Malaysia | 140 | 329749 | 18333 | 63 | 11.92 | 0.37 |
| Mali† | 31 | 1240192 | 9507 | 17 | 3.59 | 1.97 |
| Mauritania† | 8 | 1025520 | 2036 | 8 | 0.75 | 0.83 |
| Mexico* | 243 | 1958201 | 87836 | 272 | 5.84 | 2.69 |
| Mozambique | 36 | 801590 | 16084 | 90 | 6.07 | 1.39 |
| Myanmar | 105 | 676578 | 42561 | 30 | 6.93 | 0.81 |
| Namibia | 21 | 824292 | 1837 | 6 | 2.50 | 1.89 |
| Nepal | 102 | 140797 | 19605 | 16 | 6.39 | 1.98 |
| Nicaragua* | 7 | 130000 | 3999 | 8 | 8.13 | 2.15 |
| Niger | 21 | 1267000 | 7984 | 10 | 2.40 | 1.28 |
| Nigeria*† | 427 | 923768 | 112163 | 24 | 7.00 | 2.16 |
| Oman | 8 | 212457 | 1559 | 2 | 0.00 | 0.00 |
| Panama | 13 | 75517 | 2466 | 5 | 9.20 | 0.75 |
| Papua N.G. | 862 | 462840 | 3772 | 8 | 10.88 | 1.96 |
| Paraguay* | 21 | 406752 | 4397 | 16 | 10.25 | 2.51 |
| Peru* | 91 | 1285216 | 21998 | 40 | 2.65 | 4.22 |
| Phillipines | 168 | 300000 | 62868 | 64 | 10.34 | 1.92 |
| Saudi Arabia | 8 | 2149690 | 14691 | 10 | 0.40 | 0.92 |
| Senegal | 42 | 196722 | 7533 | 12 | 3.58 | 1.11 |
| Sierra Leone† | 23 | 71740 | 4260 | 23 | 8.22 | 0.59 |
| Solomon Is. | 66 | 28896 | 3301 | 1 | 12.00 | 0.00 |
| Somalia | 14 | 637657 | 7691 | 28 | 3.00 | 1.69 |
| South Africa* | 32 | 1221037 | 36070 | 114 | 6.05 | 3.50 |
| Sri Lanka* | 7 | 65610 | 17240 | 17 | 9.59 | 2.59 |
| Sudan* | 134 | 2505813 | 25941 | 43 | 4.02 | 2.82 |
| Suriname | 17 | 163265 | 429 | 2 | 12.00 | 0.00 |
| Tanzania | 131 | 945087 | 28359 | 45 | 7.02 | 1.90 |
| Thailand | 82 | 513115 | 56293 | 54 | 8.04 | 1.57 |
| Togo† | 43 | 56785 | 3643 | 11 | 7.91 | 1.78 |
| UAE | 9 | 83600 | 1629 | 6 | 0.83 | 0.69 |
| Uganda | 43 | 235880 | 19517 | 21 | 10.14 | 1.17 |
| Vanuatu | 111 | 12189 | 163 | 4 | 12.00 | 0.00 |
| Venezuela* | 40 | 912050 | 20226 | 44 | 7.98 | 2.73 |
| Vietnam | 88 | 331689 | 68183 | 40 | 8.80 | 1.59 |
| Yemen | 6 | 527968 | 12302 | 2 | 0.00 | 0.00 |
| Zaire | 219 | 2344858 | 36672 | 16 | 9.44 | 1.90 |
| Zambia | 38 | 752618 | 8780 | 30 | 5.43 | 0.67 |
| Zimbabwe | 18 | 390759 | 10019 | 52 | 5.29 | 1.43 |
| Nicaragua* | 7 | 130000 | 3999 | 8 | 8.13 | 2.15 |
| Niger | 21 | 1267000 | 7984 | 10 | 2.40 | 1.28 |

**APPENDIX**—*Continued*

| Country | Languages | Area[a] | Population[b] | Stations[c] | MGS[d] | *SD*(GS)[e] |
|---|---|---|---|---|---|---|
| Nigeria*† | 427 | 923768 | 112163 | 24 | 7.00 | 2.16 |
| Oman | 8 | 212457 | 1559 | 2 | 0.00 | 0.00 |
| Panama | 13 | 75517 | 2466 | 5 | 9.20 | 0.75 |
| Papua N.G. | 862 | 462840 | 3772 | 8 | 10.88 | 1.96 |
| Paraguay* | 21 | 406752 | 4397 | 16 | 10.25 | 2.51 |
| Peru* | 91 | 1285216 | 21998 | 40 | 2.65 | 4.22 |
| Phillipines | 168 | 300000 | 62868 | 64 | 10.34 | 1.92 |

* Denotes a country with variable Growing Season, excluded from most of the analyses.
† Denotes a country used in Nettle (1996) and excluded from some of the analyses.
[a] $km^2$.
[b] Thousands.
[c] The number of weather stations used in calculating MGS.
[d] The Mean Growing Season (months).
[e] The standard deviation of the Growing Season values from the different weather stations in that country.

## REFERENCES CITED


Abruzzi, W. S.
1982 Ecological theory and ethnic differentiation among human populations. *Current Anthropology* **23:** 13–21.

Bailey, R. C., G. Head, M. Jenike, B. Owen, R. Rechtman, and E. Zechenter
1989 Hunting and gathering in a tropical rainforest: Is it possible? *American Anthropologist* **91:** 59–82.

Barbujani, G.
1991 What do languages tell us about human microevolution? *Trends in Ecology and Evolution* **6:** 151–155.

Barbujani, G., and R. R. Sokal
1990 Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences of the U.S.A.* **87:** 1816–1819.

Berg, E.
1976 *The economic impact of drought and inflation in the Sahel.* Research for Economic Development, Ann Arbor.

Birdsell, J. B.
1953 Some environmental and cultural factors influencing the structuring of Australian aboriginal populations. *American Naturalist* **87:** 171–207.

Breton, R. J-L.
1991 *Geolinguistics: Language dynamics and ethnolinguistic geography*. University of Ottawa Press, Ottawa.

Casad, E. H.
1974 *Dialect intelligibility testing.* Summer Institute of Linguistics, Norman, OK.

Cashdan, E.
1985 Coping with risk: Reciprocity amongst the Basarwa of Northern Botswana. *Man (n.s.)* **20:** 454–474.
1990 (Editor) *Risk and uncertainty in tribal and peasant economies.* Westview, Boulder.

Cavalli-Sforza, L. L., A. Piazza, L. Menozzi, and J. Mountain
1988 Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences of the U.S.A.* **85:** 6002–6006.

Colson, E.
1979 In good years and bad: Food strategies of self-reliant societies. *Journal of Anthropological Research* **35:** 18–29.

Coulmas, F.
1992 *Language and Economy*. Blackwell, Oxford.

Crosby, A. W.
1986 *Ecological imperialism: The biological expansion of Europe, 900–1900.* Cambridge University Press, Cambridge.

DeGarine, I., and G. A. Harrison (Editors)
1988 *Coping with uncertainty in the food supply.* Oxford University Press, Oxford.

Diamond, J.
1997a *Guns, germs and steel: The fates of human societies.* Jonathan Cape, London.
1997b The language steamrollers. *Nature* **389:** 544–546.

Dixon, R. M. W.
1997 *The rise and fall of languages.* Cambridge University Press, Cambridge.

Greenberg, J. H.
1987 *Language in the Americas.* Stanford University Press, Stanford.

Grimes, B. F.
1993 *Ethnologue: The world's languages. 12th edition.* Summer Institute of Linguistics, Dallas.

Halstead, P., and J. O'Shea
1989 Introduction: cultural responses to risk and uncertainty. In *Bad year economics: Cultural responses to risk and uncertainty,* edited by P. Halstead and J. O'Shea, pp. 1–7. Cambridge University Press, Cambridge.

Headland, T. N. and L. A. Reid
1989 Hunter-gatherers and their neighbours from prehistory to the present. *Current Anthropology* **30:** 43–66.

Hudson, R. A.
1996 *Sociolinguistics.* 2nd Edition. Cambridge University Press, Cambridge.

Huss-Ashmore, R., J. J. Curry, and R. K. Hitchcock (Editors)
1988 *Coping with seasonal constraints.* University Museum, University of Pennsylvania, Philadelphia.

Huston, M. A.
1994 *Biological diversity: The coexistence of species on changing landscapes.* Cambridge University Press, Cambridge.

Krauss, M.
1992 The world's languages in crisis. *Language* **68:** 4–10.

Legge, K.
1989 Changing responses to drought amongst the Wodaabe of Niger. In *Bad year economics: Cultural responses to risk and uncertainty,* edited by P. Halstead and J. O'Shea, pp. 81–86. Cambridge University Press, Cambridge.

LeHouérou, H. N.
1989 *The grazing land ecosystems of the African Sahel.* Ecological Studies 75, Springer, Berlin.

LePage, R.
1968 Problems of description in multilingual communities. *Transactions of the Philological Society* **1968:** 189–212.

Mace, R., and M. Pagel
1995 A latitudinal gradient in the density of human languages in North America. *Proceedings of the Royal Society of London, B* **261:** 117–121.

Mayr, E.
1963 *Animal species and evolution.* Belknap Press, Cambridge.

Milroy, L.
1980 *Language and social networks.* Blackwell, Oxford.

Minnis, P. E.
1985 *Social adaptation to food stress: A prehistoric south-western example.* University of Chicago Press, Chicago.

Nettle, D.
1996 Language diversity in West Africa: An ecological approach. *Journal of Anthropological Archaeology* **15:** 403–438.

Nichols, J.
1990 Linguistic diversity and the first settlement of the New World. *Language* **66:** 475–521.
1992 *Linguistic diversity in space and time.* Chicago University Press, Chicago.

Niddrie, D. L.
1971 The Carribean. In *Latin America: Geographical perspectives,* edited by H. Blakemore and C. T. Smith, pp. 73–120. Methuen, London.

Partridge, W. L.
1989 The human ecology of tropical land settlement in Latin America: Overview. In *The Human ecology of tropical land settlement in Latin America,* edited by D. A. Schumann and W. L. Partridge, pp. 3–19. Westview, Boulder.

Pinker, S.
1994 *The language instinct.* Penguin, Harmondsworth.

Renfrew, C.
1991 Before Babel: Speculations on the origins of linguistic diversity. *Cambridge Archeological Journal* **1:** 3–23.

Romaine, S.
1994 *Language in society: An introduction to Sociolinguistics.* Oxford University Press, Oxford.

Sokal, R. R., and T. J. Crovello
1970 The biological species concept: A critical evaluation. *American Naturalist* **104:** 127–153.

Stevens, G. C.
1989 The latitudinal gradient in species range: how so many species coexist in the tropics. *American Naturalist* **133:** 240–256.

Swift, J.
1986 The economics of production and exchange in West African pastoral societies. In *Pastoralists of the West African Savanna,* edited by M. J. Adamu and A. H. M. Kirk-Greene, pp.. Manchester University Press, Manchester.

Watts, M.
1983 *Silent violence: Food, famine and the peasantry in northern Nigeria. Berkeley: University of California Press.*

Wernstadt, F. L.
1972 *World climatic data.* Climatic Data Press, Lemont, PA.

Wolff, H.
1959 Intelligibility and inter-ethnic attitudes. *Anthropological Linguistics* **1:** 34–41.

## NOTES

[1] Nichols (1992:232–237) calls this kind of diversity *genetic diversity*. This designation is confusing given the recent interest in comparing the distribution of languages with that of human DNA (e.g., Cavalli-Sforza et al 1988; Barbujani and Sokal 1990; Barbujani 1991), and so I avoid it here.

[2] These percentages are calculated from the World bank's *World Development Report 1993*.

[3] The reason for taking logarithms of these variables can be found in Section 3.

[4] Ivory Coast, Ghana, Togo, Benin, Nigeria, Cameroon, Zaire, Tanzania, India, Vietnam, Laos, Phillipines, Malaysia, Indonesia, Papua New Guinea, Vanuatu, and the Solomon Islands. The number given will be very slightly inaccurate, as languages spoken in two neighboring countries will have been counted twice. This probably involves no more than 30 or 40 languages in a total of nearly 4000.