

# 1 Introduction

The post-contact history of Martuthunira is one that has led to their almost complete extinction in little more than a hundred years. Their decline is part of a general pattern which has seen the people of the coastal Pilbara and Ashburton River districts almost completely wiped out while inland groups such as the Panyjima and Yindjibarndi continue to boast thriving communities. The demise of the coastal groups can be attributed both to introduced disease and, perhaps, to a general despair following the complete breakdown of social structure following European settlement. There were only three remaining speakers of Martuthunira [in 1981]...While this grammar of Martuthunira allows the interpretation of the literal meaning of narrative texts and, to a lesser extent, the texts of songs, a full understanding and evaluation will never be possible. Too much of the cultural context which gives them their deeper meaning has been lost. (Alan Dench, 1995)

About 97% of the world's people speak about 4% of the world's languages; and conversely, about 96% of the world's languages are spoken by about 3% of the world's people. ('Language Vitality and Endangerment', a report by the UNESCO Ad Hoc Expert Group on Endangered Languages, 2003, citing H. Russell Bernard, 1996:142)

The world's seven billion people speak about seven thousand languages today. No one knows how many languages have existed over the entire history of humankind, but the number is certainly far higher than seven thousand. For most of human history we have no way of knowing what languages were spoken where or when; direct evidence becomes available only with the invention of writing, or rather with the earliest preserved writings. These come from ancient Sumer, whose language – once the language of a major civilization and empire – has been dead for more than four thousand years. The ancient Egyptians were close behind the Sumerians in developing true writing, but their language too is long dead (although its last descendant, Coptic, lost its final native speakers only a few hundred years ago). Why did Sumerian and Egyptian die? In these two cases the primary reason was that their speakers' empires were taken over by alien (SEMITIC-speaking) cultures – Sumer by the Akkadians during the third millennium BCE, Egypt by Arabic speakers in the seventh century CE. Sumerians and Egyptians weren't massacred wholesale; instead, their speakers eventually found it expedient to shift to other languages, primarily Akkadian and Arabic.

It is certain that hundreds, perhaps thousands, of less famous languages have suffered the same fate over the past several thousand years. We have

information about a handful of these from ancient written documents, and we have information about many others from modern sources: language death is a worldwide phenomenon, a result of processes that continue at a greatly accelerated rate today. This book is designed to introduce the general topic of language endangerment – all too often a prelude to language death, or at least dormancy – with its potential cultural and scientific consequences, and to describe some methods designed to prevent endangerment from leading to the disappearance of a threatened language.

According to most experts' estimates, at least half of the world's seven thousand languages will vanish before the end of this century. The loss of linguistic diversity is sometimes compared to the worldwide loss of biodiversity: Jonathan Loh and David Harmon argue in a recent paper, for instance, that 'the world's languages, as a group, are more severely threatened than three vertebrate taxa: mammals, birds or reptiles. Languages, globally, are at least as endangered as the most highly threatened vertebrate taxon, the amphibians.' Pessimists predict the demise of about 90% of currently spoken languages by 2100; optimists say it will only be about 50%. Here are a few figures from the Catalogue of Endangered Languages (EL-Cat): there are currently 3,220 endangered languages (but as I will discuss later, this figure is elastic), 635 extinct or dormant languages, and 100 extinct language families – families of related languages that have all vanished. And a language dies – becomes dormant – every three months or so.

In the following chapters I address the obvious questions: What is language endangerment? Why does it happen? How does it happen? Who cares? So what? And what, if anything, can be done about it?

A complete map of endangered languages would have markers on every continent (except Antarctica, which has no settled population), as well as on islands scattered throughout the Pacific and other oceans. In fact, UNESCO's *Atlas of the world's languages in danger* fits this description, featuring an interactive map that can be searched by country and by language name. Examination of a global view of that map shows dense clusters of endangered languages in most areas; the only exceptions are places in which (as far as we know) there never were many languages, such as Greenland and parts of northern Asia. As we will see, the causes of language endangerment are diverse. They range from the death of all the speakers (by massacre, starvation, or devastating introduced diseases) to social, political, and economic pressures that lead people to give up their minority language in favor of a majority language. These causes are viewed from a macro-historical perspective; discussion of detailed case studies that sort out culture-specific causes of LANGUAGE SHIFT and endangerment will be limited in this book.

Before turning to more specific questions about language endangerment, I should make two general points – or, perhaps a better term, warnings – about the nature of the book. First, I have made no attempt to conceal my personal view that the loss of any language is a disaster. It is certainly a

disaster to science. Language (it is often said) is a window into the human mind, and it's clear that we'll get more benefit from this language window if we come to understand seven thousand languages fully than if we have only seven hundred to three thousand five hundred languages to comprehend. Every language in the world differs significantly from every other language, so no language is redundant in scientists' efforts to understand human cognition through understanding the workings of human language. Admittedly, the loss of a community's language is not necessarily a disaster for that community: not all communities mourn the disappearance of their heritage languages (though their grandchildren often do). But when the heritage language is an indispensable part of the community's identity, of community members' sense of self, and an indispensable repository of the community's culture, losing the language is a disaster indeed. In any case, the large number of language revitalization programs that have sprung up all over the world – programs driven by communities' eagerness to save their languages – attests to the fact that a great many endangered-language communities do view the loss of their language as a very bad thing.

My second point (or warning) is about my fairly frequent use in this book of terms like “language death”, “extinct language”, and “dying language” for the worst fate of an endangered language. I also use “salvage linguistics” to refer to linguistic fieldwork undertaken after the last fluent speakers of the target language have died. The use of such terms is controversial. Many endangered-language activists, both community members and linguists, prefer more benign terms like “sleeping language” or “dormant language” to the grim (apparent) finality of “dead language”. As Nancy Dorian observed recently, as early as the 1970s ‘many language communities found it offensive to have the label “dying language” attached to their ancestral speech form by outsiders who had no direct connection to the ethnic community and often no personal experience of the language at all’. Dorian cites Bernard Perley (a member of the Maliseet First Nation in New Brunswick, Canada) for the argument that metaphors like “dying language” are ‘detrimental to small indigenous communities’ sense of vitality and restorative possibility’. Perhaps my habit of using terms like “dead language” stems from my professional persona as a historical linguist, a student of language change: historical linguists are used to thinking and talking about instances of (mostly permanent) language death through the ages, though in general we are referring to languages that were last spoken several millennia ago rather than just a few decades ago, or yesterday. In any case, both the grim terms and the benign terms are used in this book: I have tried to reserve the grim terms for general discussions and to use the euphemistic terms especially for languages that no longer have any speakers but that are, or are likely to be, targets for revival efforts. To me it makes good sense to consider a language that is being revived to have been dormant rather than absolutely dead. But sadly, some – probably most – languages that lose all their speakers are dead for all time. I hope that readers will not be offended by either set of terms.

## 1.1 How can we tell when a language is endangered?

A language is clearly endangered when it is at risk of vanishing within a generation or two – that is, when its last fluent speakers are elderly, when few or no children are learning it as a first language, and when no one is learning it as a second language. Some experts call a language **MORIBUND** when it is no longer being learned as a first language: a language that is not being transmitted to younger generations cannot outlive the last generation of native speakers. This has been the fate of many Native languages of North America, for instance, and the same fate threatens most of the remaining indigenous languages of the United States and Canada. The same is true of most of the indigenous languages of Australia and of thousands of other minority languages in most countries all over the world.

Some sources give a more elaborately fine-grained classification of language endangerment. Perhaps most prominently, UNESCO's *Atlas of the world's languages in danger* lists six categories of languages according to their status. Here is UNESCO's Language Vitality and Endangerment Framework:

- **safe:** 'language is spoken by all generations; intergenerational transmission is uninterrupted'
- **vulnerable:** 'most children speak the language, but it may be restricted to certain domains (e.g., home)'
- **definitely endangered:** 'children no longer learn the language as mother tongue in the home'
- **severely endangered:** 'language is spoken by grandparents and older generation; while the parent generation may understand it, they do not speak it to children or among themselves'
- **critically endangered:** 'the youngest speakers are grandparents and older, and they speak the language partially and infrequently'
- **extinct:** 'there are no speakers left'

Other scales of language endangerment can be found in the literature. Lenore Grenoble and Lindsay Whaley, for instance, propose a six-way scale that is similar (though not identical) to the UNESCO framework: Safe; At Risk; Disappearing; Moribund; Nearly Extinct; and Extinct.

A list of this type, though clearly very useful, is necessarily too condensed to cover the territory thoroughly, so let's look more closely at ways of gauging a language's status. The most obvious criterion for determining whether or not a language is endangered – the overall number of speakers – turns out to be useless as a predictor unless there is evidence of changes in speaker numbers over time. Most readers of this book will be most familiar with one or more major languages that have millions of native speakers and, in a few cases, many more millions of nonnative speakers. English, for instance. But languages of small communities are not necessarily endangered; some language communities with

just a few hundred speakers have held steady at that size for centuries or even millennia. The Pirahã language of Amazonia in Brazil is an example: in 1977, the community had, and as far as experts can tell had always had, about 200 members, all of whom spoke Pirahã and only Pirahã; since then the community has grown, so that there are now about 750 members, but most of them are still (or were until very recently) essentially monolingual. Until the last decade or so, some community members knew a few words and phrases of Portuguese, but none had any proficiency in Portuguese or any other language besides Pirahã. Pirahã is therefore in no danger of being lost in this community, although the community itself, under pressure from mainstream Brazilian society, is in a precarious situation. It is probably safe to say that the language will disappear only when and if the community disappears. Nor does a large speaker population necessarily guarantee safety from endangerment: see the discussion of Navajo immediately following.

Speaker age is a more reliable indicator of endangerment: the more young speakers there are – the more children there are who are learning the language as a first language – the less likely the language is to disappear any time soon. Unfortunately, for many of the world's languages we have no evidence of speakers' ages, so it is often difficult or even impossible to apply this criterion.

There are, as the various scales of endangerment indicate, degrees of endangerment. Even languages that have many thousands of speakers and are still acquired as first languages may reasonably be considered endangered if dwindling numbers of children learn them in each generation. The Navajo language of Arizona, New Mexico, and Utah is a case in point. In 1990, the United States Census Bureau reported 148,530 Navajo speakers (according to the *Ethnologue*), out of a tribal population of 219,198; 7,616 of these Navajo speakers were reported as monolingual in Navajo. This is an impressively large number of speakers for any Native language of North America, though the lack of information about level of fluency makes it difficult to interpret the significance of the figure. Still, especially compared with the many Native North American languages whose speaker numbers are in the double or even single digits, the large number of Navajo speakers makes the language look robust. Unfortunately, the evidence of child language acquisition tells a different story. A 1969 study found that 95% of Navajo six-year-olds spoke Navajo at that time; but in a 1992 study of 682 Navajo children in Head Start preschool programs, more than half of the children (54%) were reported to be monolingual in English, and as of 1998 only 30% of six-year-old children spoke Navajo as their first language. This sharp decline in first-language acquisition of Navajo is reflected in other studies too. The rapidly decreasing number of first-language learners of Navajo makes it clear that the language is in danger of disappearing within a few generations. And if even Navajo – which in 1990 had more speakers than any other indigenous North American language – isn't safe, every other indigenous language of North America is also at risk.

Similarly, it is impossible to compile a definitive list of endangered languages for the world or any sizable part of it. Speaker numbers do not provide conclusive evidence, both because they tend to come from census reports (which have self-report data and which rarely include questions about degree of fluency) and because they provide little or no information about the age of the speakers.

Moreover, vigorous revitalization efforts have taken root in many communities, and some of them have achieved at least a partial reversal of language decline. The most spectacular success story is also the oldest one: Hebrew, dead for almost two thousand years as a language of everyday communication, was revived during the twentieth century to become the dominant language of modern Israel. It has now been learned by several generations of Israeli children as a first language, and although it is hardly the same as the Biblical Hebrew that formed its main foundation, it is nevertheless a recognizable offshoot of the ancient language. Among more limited recent success stories is that of Maori, a Polynesian language that is now one of the official languages of New Zealand. The Maori people have invested heavily in educational and other organizations designed to support their language, including preschool LANGUAGE NESTS and Maori-language immersion elementary schools; in recent decades, largely or entirely as a result of their activism, the government of New Zealand has also invested heavily in programs designed to reverse the sharp decline of Maori. The results are mixed. The perceived status of Maori is still relatively low, in part because of the urbanization of ethnic Maoris and especially because of the overwhelming dominance of English in New Zealand's political, economic, and social life. The number of Maori speakers has increased since the 1960s and, crucially, that number now includes younger people. Bernard Spolsky, writing in 2009, said that 'Māori looks like one of the few threatened languages that can be confident about its next hundred years.' This optimistic view suggests that the increasing prominence of preservation and revitalization programs around the world will (with considerable luck) point to improvement in the condition of some languages that are now endangered, ideally to the point of long-term stability.

Another complication in any effort to define 'endangered language' concerns the status of endangered dialects. The difficulty is most evident in the case of dialects that are spoken outside a language community's traditional homeland. Hungarian, for instance, is obviously not an endangered language: it has millions of speakers in Hungary. But Hungarian immigrants in the United States are shifting to English, so that all varieties of American Hungarian are moribund. Even within a language's home territory, dialects – especially nonstandard dialects – often disappear; in many countries with mass education in a dominant STANDARD DIALECT (officially or unofficially acknowledged as the "correct" form of the language), nonstandard dialects have vanished or are vanishing rapidly. This is one of the primary effects of standardization: all dialects other than the newly standard one are likely to suffer. To most people, linguists as well as laymen, the loss of a dialect seems less important than the loss of an entire language.



But dialects, like whole languages, are carriers of culture. A community that loses its unique dialect inevitably loses a part of its culture and its local identity within the larger community; and loss of a dialect, like the loss of a language, robs science of potential insights into the workings of the human mind. (This isn't the place to enter into a detailed discussion of the distinction between two dialects of the same language and two separate but closely related languages. Briefly, the only viable linguistic criterion for making the distinction has to do with mutual intelligibility: if there is some, they're dialects; if there is none, they're separate languages. But often this common-sense criterion doesn't give a straightforward answer because of various sociopolitical considerations that affect perceived intelligibility; and the saying 'A language is a dialect with an army and a navy', though definitely not universally applicable, is not wholly invalid.)

Finally, the criterion that is most useful in deciding whether a language is endangered or not – the ages of the speakers – does not apply to all languages. PIDGIN languages, which are by definition learned and used solely as second languages and for limited purposes (such as trade with outsiders), have no native speakers: no children learn them as first languages. It's true that most known pidgins are now endangered, if they have survived at all. But the reason is unconnected with the age of learning; it has to do instead with changing worldwide economic and social conditions. The Age of Exploration, during which pidgins sprang up on African, Asian, and American coasts when Europeans began visiting them, ended several centuries ago, as did the earlier period of Arab expansion, which also gave rise to a number of pidgins. Pidgin languages have been reported from other contexts as well, such as the pidgin that was once used by (Native American) slaves of the Nez Perce tribe of the Columbia River plateau in northwestern North America. Populations are more settled nowadays, and most pidgins are being, or have been, replaced by multilingualism and/or shift to the languages of former colonial powers. The other fate of a pidgin language is development into a CREOLE LANGUAGE – that is, the acquisition in a settled multiethnic community of a native-speaker population, until the language has so many native speakers that it can no longer be considered a pidgin. This is what happened to Tok Pisin, now one of the official languages of Papua New Guinea: it started life in the nineteenth century as a pidgin (the name literally translates as 'talk pidgin'), with a vocabulary drawn primarily from English and a compromise grammar based mainly on indigenous AUSTRONESIAN languages. It became so embedded in the life of the country, including increasing use as a home language when speakers of different indigenous languages married, that it is now a creole. By contrast, the seventeenth-century pidgin that was used between Delaware (Lenape) Indians and European missionaries and settlers from several different countries is long dead, and Lenape itself is gravely endangered.

Other languages that fall outside the scope of endangered-language designation are ancient languages that survive only in limited domains of usage. Most of these are religious languages, dead for centuries or even millennia

as languages of everyday communication but preserved in religious contexts. Hebrew was one of these, but no longer. Latin, Classical Arabic, and Coptic – the last descendant of Ancient Egyptian, whose latest known use as an ordinary spoken language was sometime between the fourteenth and seventeenth centuries (see Chapter 3 for further discussion) – are examples. Sanskrit is a borderline case: it is the ancient sacred language of the Hindu religion, and as such it is used in religious contexts and codified in a form that was essentially fixed in the fourth century BCE. But the 1991 Census of India reported 49,736 fluent speakers of Sanskrit (in self-report data), and efforts are being made to revive it as a spoken language. There are several “Sanskrit villages” in India where Sanskrit is the everyday language of communication; for instance, in the state of Karnataka, in the village Mattur (also spelled Mathur or Mathoor), the entire population is said to be fluent in Sanskrit. Given the fact that several hundred other spoken languages compete with Sanskrit in India – almost thirty of them with more than a million speakers each – it is hardly likely that Sanskrit will ever become the country’s major spoken language; the conditions that favored the remarkable rise of Modern Hebrew (a new nation established by a beleaguered people, a linguistically mixed population uprooted from their homes and with no common language) do not exist in India.

Still, if we exclude pidgins and ancient languages preserved in religious contexts, the age of the speakers is generally a good indicator of a language’s prospects for long-term viability. As a typical example, consider Veps, a FINNIC language spoken in northwestern Russia, not far from Lake Ladoga. In 1926, 94.7% of ethnic Veps people spoke Veps as their first language; by the 1960s the percentage was down to about 60%, and it has continued to decline ever since. Nowadays no children are learning Veps as a first language, and there are few or no young people in the villages where Veps is still spoken. The language has no official status. Although legally it can be used in schools, it is mainly ‘taught as a hobby school subject or as a second foreign language twice a week’. Unless current revitalization efforts succeed in reversing its course, the language will soon disappear.

## 1.2 Where are all the endangered languages?

The short answer to this question is, everywhere. Endangered languages are found on every continent (except Antarctica). Counting endangered languages is as difficult as defining the category, mainly because far too little relevant information is available for many parts of the world. At best we have speaker figures from national census reports; but even these, as noted earlier in this chapter, are unreliable as indicators of a language’s status as endangered or not. The most comprehensive source that gives speaker numbers for all known languages is the *Ethnologue* ([www.ethnologue.com](http://www.ethnologue.com)), a product of the missionary



organization SIL International. The speaker numbers in the *Ethnologue* are often outdated, and any figures are bound to be approximate in any case; still, even with its inevitable limitations, this is an invaluable resource for anyone who wants to find out about the location(s) and general demographics of any language.

One searchable category on the *Ethnologue* website is “nearly extinct languages”, which includes all languages that are spoken by just a few elderly people. The page lists about 516 of these in 58 different countries (the language count is approximate because the entries are organized by country and some of the languages are spoken in more than one country). In other words, according to the *Ethnologue* figures, more than 500 of the world’s approximately 7,000 languages are about to disappear; in fact, many have surely already disappeared, because some of the *Ethnologue*’s sources are twenty or thirty years old. Of the 516 nearly extinct languages, 46 are in Africa, 170 in the Americas, 78 in Asia, 12 in Europe, and 210 in the Pacific. The latter region includes both Australia (with 168 languages – but this is an exaggerated figure, as we will see) and Oceania (42 languages). These counts do not include nearly extinct dialects of languages in which one or more dialects are not nearly extinct. Australia is the country with by far the largest number of languages in the *Ethnologue*’s “nearly extinct” category, followed by the United States (68), Indonesia (31), and Brazil (30). The two countries in Africa with the largest numbers of nearly extinct languages are Nigeria (15) and Cameroon (13); in Europe only one country, Russia, has more than a single nearly extinct language, and Russia has only two. These figures of course do not include languages that were extinct before the *Ethnologue* was compiled, and we know that there were many of these in North America, Australia, and probably also Russia, as well as in most or all other parts of the world.

Let’s look more closely at the status of Australian languages, using estimates for speaker numbers that are more recent than the *Ethnologue* figures, which date from 1970 to 1983. The starting point would be the number of Indigenous languages (including both Aboriginal and Torres Strait Islander languages) that are estimated to have been spoken before Whites arrived in Australia: most estimates give a figure between 200 and 300. By 1990, 64% of these languages were either extinct or reduced to a handful of elderly speakers; 28% were “severely threatened”, and a mere 8% were “relatively healthy”. As of 1990, 20 Indigenous Australian languages were considered “strong” (according to the criterion “spoken by all age groups regularly”); by 2001, three of those twenty were endangered. At the other end of the scale, 35 Indigenous languages were still spoken in 1990 by just a few elderly people each, and by the turn of the millennium all those languages had probably become extinct. The number of surviving Indigenous languages in 2000 would then be about 90. According to 1990 figures, 8 Indigenous languages still had 1,000 or more speakers each; the largest speaker counts were for Central Torres Strait (3,000–4,000 speakers), Arrernte (3,000+), Pitjantjatjara (3,000+), and Warlpiri (3,000+). By now the speaker numbers for most or all of these eight languages are likely to be lower,

perhaps much lower. In their 2001 survey of the state of Indigenous Australian languages, Patrick McConvell and Nicholas Thieberger emphasize the dramatic ‘decrease in the percentage of Indigenous people speaking Indigenous languages from 100% in 1800 to 13% in 1996’, a trend that was especially rapid in the decade from 1986 to 1996. They comment, ‘If these trends continue unchecked, by 2050 there will no longer be any Indigenous languages spoken in Australia.’ Even if the rate at which speakers and languages are lost diminishes (as McConvell and Thieberger believe it will), the prospects for the long-term survival of Australian Indigenous languages – which are not only the carriers of Indigenous cultures but also display some linguistic features that are totally unknown elsewhere in the world – are bleak.

Another sobering aspect of the *Ethnologue*’s list of nearly extinct languages is that it includes nine of the forty languages classified by the *Ethnologue* as ISOLATES – that is, languages that are not known to be related to any other language: Itonama and Leco (Bolivia), Muniche and Taushiro (Peru), Puelche (Argentina), Yámana (Chile), Kutenai (U.S. and Canada), Yuchi (U.S.), and Ainu (with fifteen speakers in Japan as of 1996; the last speaker of the Sakhalin dialect of Ainu died in 1994). To this list one might add the Lenca language of El Salvador, which is listed in the *Ethnologue* as ‘unclassified’, and some other unclassified languages are also likely to be isolates. If an isolate dies, its linguistic and cultural riches are lost for all time: it leaves no relatives that might continue some of its rarest and most interesting features.

The Australian case is perhaps the most striking example of sweeping loss, both in terms of the overall number of extinct and nearly extinct languages and in terms of the potential loss to science and the broader study of human culture and society – of information vital to the construction of viable theories of universal aspects of language structure and of information about songs, narratives, kinship systems, and other products of individual and social creativity. Australia is hardly the only example, however; other groups of languages are in a comparably dismal situation. To give just one example, the Salishan LANGUAGE FAMILY of the U.S. and Canada has, or had, twenty-three known languages. Of these, eight are listed by the *Ethnologue* as nearly extinct (Bella Coola, Clallam, Coeur d’Alene, Lushootseed, Sechelt, Snohomish, Squamish, and Straits Salish), and seven may already be extinct (Lower Chehalis, Upper Chehalis, Cowlitz, Nooksack, Quinault, Tillamook, and Twana). That leaves just eight languages in the family that are in neither of these end-stage categories (Comox-Sliammon, Halkomelem, Shuswap, Lillooet, Thompson, Colville-Okanagan, Columbian, and Montana Salish-Kalispel-Spokane), all of which are “merely” gravely endangered. Three of the family’s five branches are in great danger of dying out entirely (the Bella Coola and Tillamook branches, each comprising just one language, and the four-language Tsamosan branch).

The numbers and distributions of nearly extinct languages are of course just the tip of the iceberg: overall estimates of endangered languages are much higher. The only possible conclusion is that language endangerment is a pervasive