1

Warramurrungunji's Children

يَأَيُّهَا النَّاسُ إِنَّا خَلَقْنَاكُمْ مِّنْ ذَكَرٍ وَأُنْثَى وَجَعَلْنَاكُمْ شُعُوباً وَقَبَائِلَ لِتَعَارَفُوا

Oh Mankind, we have created you male and female, and have made you into nations and tribes that ye may know one another.

(*Qur'an* 49: 13, Pickthal 1938 translation)

In the oral traditions of northwestern Arnhem Land, the first human to enter the Australian continent was the ancestress Warramurrungunji, who came out of the Arafura Sea on Croker Island near the Cobourg Peninsula, having travelled from Macassar in Indonesia. Her first job was to sort out the right rituals so that the many children she gave birth to along the way could survive. The hot mounds of sand, over which she and all women thereafter would have to purify themselves after childbirth, remain in the landscape as the giant sandhills along Croker Island's northern coasts. Then she headed inland, and as she went she put different children into particular areas, decreeing which languages should be spoken where. In the Iwaidja version of the story she said, *Ruka kundangani riki angbaldaharrama! Ruka nuyi nuwung inyman!* 'I am putting you here, this is the language you should talk! This is your language!', naming a different language for each group before moving on.

Language Diversity and Human Destiny

Ŋarri-watjpatjtjun djunbalal-ŋupan wultjimin
dhä-walwaldjun liliya waŋa
Dhuwanydja matha-gulkulkthunmaraŋala
dhuwalgindhu
buṯthurruna ŋari-watjpatjun waŋa ŋariŋarryun

Speech of different clans, mingling together...
Dhuwa moiety clans, with their special
distinct tongues.
People from Blue Mud Bay, clans of different
tongues talking together...
Words flying over the country, like the voices
of birds...
(Song 2, Rose River Cycle, Berndt 1976: 86 –87,
197–198,
retranscribed into modern Yolngu orthography)

Words of Wonder: Endangered Languages and What They Tell Us, Second Edition. Nicholas Evans.

The Judeo-Christian tradition sees the profusion of tongues after the Tower of Babel as a negative outcome punishing humans for their presumption, and standing in the way of cooperation and progress. But the Warramurrungunji myth reflects a point of view much more common in small speech communities: that having many languages is a good thing because it shows where each person belongs. Don Laycock quotes a man from the Sepik region of Papua New Guinea saying 'it wouldn't be any good if we all talked the same; we like to know where people come from.'[1] The Tzotzil oral traditions of the Mexican Chiapas give another twist to this view: 'while the sun was still walking on the earth, people finally learned to speak (Spanish), and all people everywhere understood each other. Later the nations and municipios were divided because they had begun to quarrel. Language was changed so that people would learn to live together peacefully in smaller groups.'[2]

I recently drove down the dusty road from Wilyi on the coast near Croker Island, to the inland town of Jabiru (Figure 1.1), while working with speakers of Iwaidja, the language in which Tim Mamitba had told me the Warramurrungunji story. The 200-km transect follows Warramurrungunji's path, traveling inland and southwards from the beach through eucalyptus savannah, stretches of tropical wetlands and lily ponds, and occasional sandstone outcrops whose caves hold vast galleries of rock paintings. It is a timeless landscape rich in wild food – magpie geese, fish, bush fruits, and yams. Its Aboriginal inhabitants can live easily throughout the year, finding all they need on their own clan countries. The few river crossings do not present geographical barriers. But Warramurrungunji's legacy of linguistic diversity is clearly here. In a few hours on the road, we passed through the territories of nine clans and seven languages from four language families, at least as different from each other as English, Portuguese, Russian, and Hindi (see Table 1.1).

To give a rough idea of how different the languages are at the two ends of this transect, consider the sentence 'you eat fish'. Taking Iwaidja from one end, and Kun-djeyhmi from the other, we compare *kunyarrun yab* and *yihngun djenj* – of which only the final *-n* in the two languages, which marks present tense in both, is historically relatable (or 'cognate'). Imagine I had driven from London to Moscow – 15 times as far. The Russian equivalent *ty esh rybku*, although incomprehensible to English ears, contains three cognate elements, at least if we cheat a bit by taking the earlier English version, *thou eatest fish*: *ty* (with English *thou*), *e* (with English *eat*) and *-sh* (with the older English suffix *-est* in *eatest*). And if I satisfy myself with a shorter trip to Berlin – still more than five times the Wilyi–Jabiru trip – we get the almost comprehensible *du ißt (isst) Fisch*, in which every element is cognate.

Some of these languages are now down to just a couple of speakers (Amurdak) or have recently ceased to be spoken (Manangkardi), but others are still being learned by children. Bininj Kunwok, the largest, now has about 2000 first language speakers as members of other groups shift to it. But the average population per language in this region is much smaller, probably less than 500 speakers. And many are even smaller: a study by Rebecca Green[4] on Gurrgoni, a few hundred kilometres to the east of the Warramurrungunji track, suggests it has been quite stable for as long as anyone remembers, never with more than around 70 speakers.

Each person from this region has one 'father language', which they have special rights in, by virtue of the clan membership they obtain from their father. This vests them with authority and spiritual security as they travel through their ancestral lands. In travelling to places that have not been visited for some time, clan members should call out to the spirits in the local language, to show they belong to the country. Doing this with visitors is the duty and right of a host, and this has been adapted in contemporary settings as 'welcomes to country'

Figure 1.1 Clans and languages in northwestern Arnhem Land.

Table 1.1 Clans and languages along the 200-km track from Wilyi to Jabiru.[3]

Clan	*Language*	*Language family*
Murran	Iwaidja	Iwaidjan; Iwaidjic
Manangkali	*Amurdak*	Iwaidjan; Southern
Minaka	*Manangkardi*	Iwaidjan; Iwaidjic
Born/Kardbam (Alarrju)	*Bininj Kunwok (Kunwinjku dialect)*	Gunwinyguan (Central)
Mandjurlngun	*Bininj Kunwok (Kunwinjku dialect)*	
Bunidj	*Gaagudju*	Gaagudjuan (Isolate)
Mandjurlngun Mengerr	*Mengerrdji*	Giimbiyu
Manilakarr	*Urningangk*	
Bunidj Gun-djeihmi, Mirarr Gun-djeihmi	*Bininj Kunwok (Kun-djeyhmi dialect)*	Gunwinyguan (Central)

by the relevant custodians. It is said that many resources, such as springs, can only be accessed if you address them in the local idiom. For these reasons there are deep emotional and spiritual links between language and country. Travellers sing songs listing the names of sites as they move through the land, and switch languages as they cross creeks and other clan boundaries. In epics of ancestral travels, it is common to identify where the characters have got to simply by switching the language the story is told in – as if *The Odyssey* were told not just in Greek, but in the half-a-dozen ancient Mediterranean languages Ulysses would have encountered in his travels.

Throughout Aboriginal Australia, speaking the appropriate local language is a kind of passport, marking you – both to local people and to the spirits of the land – as someone known and familiar, with the right to be there. I once went out in a boat with Pat Gabori and some senior women to map a Kayardilt site a few kilometres off shore, in the company of a few children who did not know their ancestral language. Pat and the women called out in Kayardilt to the spirits and ancestors of the place, identifying themselves and introducing the silent children, and

Box 1.1 The many paths to multilingualism in up-country Cameroon

Jonas courting Gogo in Jilve village in the presence of other villagers. (Courtesy of Leslie Moore.)

Jonas, the boy in this photo, comes from the village of Jilve in the Mandara region of Cameroon, another region of daunting linguistic diversity, where people speak small 'montagnard' languages of the Chadic family, very distantly related to Hebrew, Arabic, and Egyptian. Here he is shown courting Gogo, the girl he wants to marry, at her mother's compound, with her girlfriends in attendance. They are speaking primarily in Mada,

explaining gently that the children's inability to speak Kayardilt did not make them strangers – they just hadn't learned the language yet.

A more extreme illustration of this principle comes from a story by Pluto Bentinck, another senior Kayardilt man, related during a Native Title claim. When asked if traditional law included sanctions to be taken against trespassers, he cited an incident during World War II, when a hapless white airman swam ashore on Bentinck Island after his plane had crashed into the sea. Pluto told me the man had said *danda ngijinda dulk, ngada warngiida kangka kamburij* 'this is my country, I just speak this one language', as he struggled ashore without his Kayardilt phrasebook. When I asked Pluto how he knew what the man had said, when he himself knew no English, he replied: *Marralwarri dangkaa, ngumbanji kangki kamburij!* 'He was an ear-less (crazy) man, he spoke your language!'. Speaking English on Bentinck Island, in Pluto's view, was tantamount to claiming it for English speakers. *Nyingka kabatha birdiya kangki! Ngada yulkaanda mirraya kangki kabath!* 'You found the wrong words! I've found the right words,

The Mandara mountains, Cameroon.

Gogo's paternal language. Mada is one of eight languages that Jonas speaks. Although he began learning Mada in order to court Gogo, the two of them already had two languages in common: Wandala (the local lingua franca) and Wuzlam, the first language of Jonas's father and of Gogo's mother. Prior to this visit, Jonas had prepared a list of topics of conversation and relevant Mada vocabulary, which he had noted on a piece of paper he brought with him but did not consult during the visit.

We know very little about how such impressive levels of multilingualism are acquired in small-scale preliterate societies, but Leslie Moore's pioneering ethnographic work (from which the earlier vignette is taken) has taught us something about how multilingualism works in Cameroon communities like Jilve. Besides the 'normal' acquisition of their mother tongue, children learn French and later English in school, the regional lingua franca Fulfulde from night-time story-telling by their elders, and the languages of neighbouring villages through self- instruction of the type we see here. From an early age parents ask their children to memorize messages, in languages they do not yet know, and to go to deliver them orally to people from neighbouring villages. Even young children develop a strong metalinguistic awareness, for example using knowledge of cognates in related languages to help them remember new vocabulary.

since forever!' he had replied to the man. Pluto continued: *Ngada bunjiya balath, karwanguni* 'and I clubbed him in the back of the neck'.[5]

Normal members of Arnhem Land society are highly multilingual, often speaking half-a-dozen languages by the time they are adults.[6] This is helped by the fact that you have to marry outside your clan, which probably means your wife or husband speaks a different language from you. It also means that your parents each speak a different language, and your grandparents three or four languages between them. The late Charlie Wardaga, my Ilgar teacher, was typical. Knowledge of Ilgar, Manangkardi, Marrku, Iwaidja, and Kunwinjku came to him from his grandparents and parents. Although he lived mostly on lands where Ilgar, Marrku, and Iwaidja were the locally appropriate languages, he married a Kunwinjku-speaking woman from a mainland clan and he would regularly speak Kunwinjku with her and her relatives, or when travelling to distant communities as a songman. In this system your clan language is your title deed, establishing your claims to your own country, your spiritual safety and success in the hunt there. Meanwhile the knowledge of other languages gives you the far-flung network of relatives, spouses actual and potential, ceremonial age-mates and allies, which makes you someone who counts in the greater world. This combination of highly developed multilingualism with strong attachments to small local languages is by no means an Arnhem Land oddity – around the world, it is common in zones of high linguistic diversity, like Nagaland in northeastern India, or the Mandara Mountains of Cameroon (see Box 1.1).

Most non-Aboriginal people are astonished when they learn how many demographically tiny languages etch their distinctive local domains across the Australian map. Modern citizens of industrialized countries like Britain or Japan take it for granted that they can use their languages with hundreds of millions of people and that a single language occupies the entire territory of their nation, bar dialect variation, immigrants, and one or two beleaguered minorities like Welsh or Ainu. For speakers of big languages, the question is: why are there so many languages in the world/in Papua New Guinea/in Australia/etc.? The naïve explanation sometimes offered, that they result from mutual isolation in distant valleys and gorges, just does not bear up. In Arnhem Land there are no significant geographical barriers at all, and the same is true in many other linguistically exuberant parts of the world, such as Southern New Guinea. And marriage patterns, in Arnhem Land, Southern New Guinea, or the Vaupés region of Amazonia, mean that several languages are spoken on a daily basis inside the one household (indeed, in the one bed!) – hardly a case of mutual isolation.

But maybe we are approaching the problem from the wrong end. It makes more sense to turn the question round and ask, not why Melanesia, the Amazon, Arnhem Land, Cameroon, or the Caucasus have so many languages, but rather why Europe or parts of Asia have so few?

Indeed, there are good reasons to believe that our little transect through Arnhem Land is a good representation of how we humans have been for most of our past – not just for the 99% or so of our history up to 10 000 years ago when we were all hunter-gatherers, but also for much of the time that followed. This is because the dawn of agriculture, although it led to an explosion in human populations, did not automatically lead to the development of much larger societies. Speech communities got a bit bigger, but it was rare for them to exceed the few score thousand that could be held together as a homogeneous unit without the panoply of state control that only began with our incorporation into large centralized political entities like the Roman Empire or a modern nation state. Mapping the million or more years of human history onto a 24-hour clock, incorporation into large centralized states did not start for any human society before 10 minutes to midnight (five millennia ago in the Fertile Crescent). For many groups it has only begun to happen in the last few seconds.

The island of New Guinea and its Melanesian surrounds, a few hundred kilometres to the north of Warramurrungunji's territory, is a good illustration of a region almost completely made up of Neolithic agricultural societies, with no centralized states until recent colonization by Europeans and Indonesians. Its population of around 10 million people speaks some 1150 languages – under 10 000 people per language. In the Central Highlands, where the population density is highest thanks to intensive agriculture and pig breeding, elaborate networks of production and ceremonial exchange have gradually bound people together into larger speech communities. The more intensified the system, the more speakers per language. But, even in the most elaborate and intensified highland Papuan communities, the average number of speakers per language rarely exceeds 40 000. And in many other parts of Melanesia languages of that size seem unimaginably large: the nation of Vanuatu (total population 308 000), most of whose population are village agriculturalists, counts 135 languages[7] – an average of just over 2000 speakers per language for the whole country. Apart from the recently developed national lingua franca, Bislama, its biggest language (Lenakel) has just 12 000 speakers and only 13 languages have 5000 speakers or more.[8]

We see small languages wherever in the world societies have lain beyond the homogenizing reach of great empires. But the situation is most extreme where groups can maintain themselves self-sufficiently without needing to call on the hospitality of others. The village cluster of Archib (population 1237) in the Caucasus is the only place in the world where the Archi language is spoken – a language whose morphology is so complex that it has been calculated that a single verb possesses more than 1.5 million inflected forms. Most of its inhabitants will be born, married, and laid to rest in this one village cluster, basing their economy on specially adapted mountain sheep, which they tuck up on freezing winter nights into special underground sheep folds called *małłi* (see Figure 1.2). Or, in northwestern California, the entire territory of the Chimariko people and their language consisted of a 20-mile stretch along a narrow canyon of the Trinity River.[9] Until the Gold Rush, their economic self-sufficiency on this small patch was assured by the rich salmon stocks in the river.

Figure 1.2 Archi men herding sheep into a *małłi* or underground sheep fold. (Courtesy of Marina Chumakina.).

There is evidence from many parts of the world that small groups in favoured areas did not simply rely on the drift of time to carry their languages apart from those of their neighbours. In northern Australia the reigning ideology is that each clan should have its own distinct language variety. This then sanctions the investiture by tribal elders of variant forms as proper to their local languages, driving along a relentless diversification.

Peter Sutton, working on clan identities in the Cape York Peninsula of Australia, reports cases where the fission of clans is rapidly followed by the emergence of new language varieties. In settings where fewer than 100 people may speak a 'clan-lect', one or two powerful individuals can readily impose what may have started out as individual idiosyncrasies, and seed the emergence of a new system. In Iwaidja, many forms of nouns and verbs mutate their initial consonants – 'his or her arm' is *bawurr*, from the root *mawurr*, for example, whereas the corresponding words in related languages like Charlie Wardaga's language Ilgar keep the original *m*. The mutated forms are based on an obscure 'miscellaneous' gender so rare in all the languages of the family that it would never have won out as the standard form by processes of normal change. More likely, at some point in the past, Iwaidja speakers deliberately extended the use of the miscellaneous gender to set their language apart from their neighbours, on the 'you say tomahto, I say tomato' principle.

In New Guinea language differentiation is sometimes fostered even more deliberately. When we compare the Uisai dialect of Buin (1500 speakers), on Bougainville Island, with the other dialects of Buin (about 17 000 speakers in all), we see that it has completely flipped over all of its gender agreements:[10] all the masculines have become feminine, and all the feminines have become masculine. Because no known mechanism of normal linguistic change could produce this effect, Don Laycock has suggested that 'an influential Uisai speaker innovated a linguistic change to differentiate his community from the rest of the Buins.' Again we see how much influence a single individual can have in a small speech community. For another Papuan language, Selepet, we actually have a reported instance, witnessed by linguist Ken McElhanon, where one community decided at a meeting to replace the standard Selepet word *bia*, for 'no', with the word *bunge*, to differentiate themselves from other Selepet villages.[11]

Although for illustrative purposes we have concentrated on decisions affecting a single word or grammatical feature, this is just the thin end of the wedge. William Thurston studied 'esoterogeny' – the engendering of difference and linguistic obscurity – with Anem speakers on the island of New Britain, off the New Guinea mainland. He found that 'esoterogenic' languages tend to streamline pronunciation in ways that make the overall structure harder to see, comparable to saying *dja* for *didja* from *did you* in English. They replace clear regular relationships with 'suppletive' (totally irregular) ones, revelling in alternations like *good:better* at the expense of the more transparent *big:bigger* style. They have huge numbers of opaque idioms, of the *kick the bucket* type, and entrench prescriptive traditions that limit flexibility of language: 'you must speak this way to be a member of our community!' They also elaborate terminology to make subtle distinctions, and speakers take more pride in the greater richness of their language than the neighbouring language of Lusi in this regard.

During Thurston's research on Anem he found that 'some of the boys had devised a competitive word game aimed at exposing one another's ignorance of the name for an obscure vine or bush; in order to keep ahead, boys were asking older people, secretively, for words they could use to try tricking other boys.' All these forces conspire to maximize difference between one language and its neighbours – although I should stress that, up till now, we have no more than the sorts of anecdotal evidence reported here, and systematic studies of the causes and processes of change in small languages are badly needed.

Small-scale societies in such parts of the world are economically self-sufficient, and proudly form the centre of their own social universe without needing to defer unduly to more powerful outside groups. Their constructive fostering of variegation – which holds social groupings to a small and manageable size, and keeps outsiders at a suitable distance – is not offset by the need to align their language with large numbers of other people in the world. The great Swiss linguist Ferdinand de Saussure saw language as being pulled in opposite directions by the 'spirit of the steeple' – the parochialism of showing which little community you belong to – and the 'spirit of wider communication'. But, for those small-scale societies able to subsist mostly on their own resources, the force of the steeple is dominant.

Language Diversity through Time and Space

The classic estimate of the world's population on the eve of the Neolithic, 10 millennia ago, is 10 million.[12] Combining these figures with a very generous 2000-speaker maximum for hunter-gatherer languages suggests that, on the eve of agriculture and fixed settlements, there were already 3000 to 5000 languages in the world – roughly the same number as now, even though the population was less than 0.5% of its current level. If we assume 1000 speakers per language, a more realistic figure in my view, the number doubles to between 6000 and 10 000.[13] Levels of language diversity whose full magnitude we can barely grasp have been with us for a very long time.

Going much further back in time, to the population bottleneck about 150 000 years ago that preceded the long trek out of Africa, Rob Foley estimates from mitochondrial DNA that there were probably between 10 000 and 20 000 women of reproductive age – say 50 000 humans all up, who on the language-size estimates given here would already have been speaking 10–20 distinct languages (and possibly 100 if we go with the plausible language population of 500). Already then, at a point when humans went through a population bottleneck probably caused by environmental crises, just before their fortunes turned and they ventured forth across the planet, there were scores of languages. Since our forebears probably began to speak and develop languages long before that,[14] the figure 150 000 years ago is likely to have included tongues that were already quite different from one another.

Let us come back to the emergence of agriculture at the beginning of the Neolithic. As agriculture spread around much of the world from then on, it is likely that the increases in language populations of cultivating groups – perhaps to New Guinea-like levels of around 10 000 – would have been more than offset by the explosion in overall world population, so that the number of languages in the world may have risen to 10 000 or 15 000. However, scholars like Colin Renfrew and Peter Bellwood[15] have argued that the bearers of agricultural expansion would have been just a small number of groups who had made a radical cultural transition. From the hunter-gatherer perspective, agricultural life looks pretty unappealing. In return for the security of regular crop supplies you have to put up with a poorer and less-varied diet, monotonous year-round residence in a circumscribed area, and the diseases that come from having so many people living in relatively close quarters, on top of each other's excrement. Most hunter-gatherers would have walked away from the deal. Inexorably, though, the 'demic expansion' that agriculturalists could feed from the increased food yields they drew from the land would gradually have squeezed out or assimilated the original populations of hunter-gatherers.

The expansion of these few agriculture-based lineages would have produced the 'Renfrew–Bellwood' effect: a decrease in deep-level diversity, i.e., in the number of unrelated stocks or deep lineages, as clusters of closely related languages spread outward from the dozen or so foci where agricultural complexes were developed. This would have obliterated the deep-time variability that was there beforehand. If agricultural expansion is old enough, however, there is time for significant new multiplicity to develop, as has happened with the Afroasiatic languages stretching from Hausa in Nigeria to Hebrew and Arabic in the Middle East, or the Austroasiatic family in Southeast Asia. Each of these families probably goes back to the dawning of agriculture in their respective regions, and contains languages so different that it has taken many years of work to demonstrate their relatedness – a topic we return to in Chapters 5 and 6.

We can see the Renfrew–Bellwood effect clearly in Indonesia where, over most of the archipelago, no linguistic trace remains of the hunter-gatherers who must have occupied its fertile lands until the coming of the Austronesian agriculturalists a few thousand years ago. It is also clear in New Guinea, where the whole fertile highland cordillera along which root-and-pig farmers expanded over the past few millennia is occupied by a single 'Trans New Guinea' family, albeit one with around 400 member languages. New Guinea's most mind-boggling lineage diversity is found in lowland areas like the Sepik and the Trans-Fly region, north and south of the cordillera, where people either practice a much less intensive form of agriculture or mix agriculture, fishing, hunting, and gathering.

We have seen already, though, that even in the densely populated New Guinea Highlands, and even after 6000–7000 years of intense agriculture, there were no really large languages. This is because it was only with the advent of centralized and then industrialized state societies that a few languages began to spread to the point where they counted hundreds of thousands and then millions of speakers. Unfortunately, most of these expansionist new societies had no interest in recording anything about the languages of the peoples they subjugated, as we will see in Chapter 2. But we can get some idea of what the world was like as the first great empires emerged by looking at the Italian peninsula in the fifth and sixth centuries BCE. There, under Greek influence, a number of different civilizations developed their own writing systems in time to leave some record of their languages before they were all sucked into the Latin-speaking vortex of the Romans.

Inscriptions in pre-Roman Italy attest between 12 and 15 distinct languages, quite different from one another, and belonging to four branches of at least two distinct families – three branches of Indo-European (Celtic, Italic, and Greek) plus Etruscan, which was non-European. The Romans did not actively try to stamp out other languages – indeed, the retention of other languages by non-Romans favoured the policy of *diuide et impera* 'divide and rule'. Umbrians, for example, continued to make inscriptions in their language for centuries after Roman annexation. But eventually the power and status of Latin prevailed, particularly after all residents of Italy became Roman citizens in the middle of the final century BCE. At first, other groups would just have used Latin for 'outside' purposes, but gradually the centralizing power of Rome 'relegated the local speech, just as it did political initiative and concerns, to a secondary, subordinate, and ever retreating position'.[16] We do not know exactly when the last speakers of Oscan, Umbrian, Etruscan, and other languages of the peninsula finally passed away, but the elimination of all non-Latin languages from the Italian peninsula is likely to have been almost complete by the time Pontius Pilate washed his hands of another death in another part of the Roman Empire.[17]

Emerging kingdoms in many parts of the world – in Egypt, Arabia, Persia, Mali, China, Korea, India, Mexico, and the Andes – had similar impacts on the smaller peoples in their domains. It

was probably in this period – beginning around 2000 BCE – that the first languages with more than a million speakers emerged. Expansionist agricultural–military complexes like the Bantus in the southern half of Africa obliterated vast mosaics of diversity. And then, from 1492, European colonial expansion began to take its toll. Little more than half a century after the Spaniards reached Cuba and Puerto Rico, the Arawakan language Taino would cease to be spoken, although some of its words have survived as loanwords into Spanish (*cacique* 'chief') and others passed further into English (*barbacoa* 'barbecue'; *canoa* 'canoe'; *tabaco* 'tobacco'). Thousands of other languages around the world would suffer a similar fate, leading to the accelerating loss of linguistic diversity we see today, and the concomitant dominance of the dozen or so languages with more than 100 million speakers.

Where the Hotbeds Are

The upshot of what we have been discussing is that language diversity is now distributed very unevenly around the world. On one estimate,[18] 17 countries hold 60% of all languages, although these countries make up only 27% of the world's population and 9% of its land area.[19] Table 1.2 shows two slightly different rankings of the top 25 language-diverse countries – a calculation of the sheer number of endemic languages,[20] and an alternative measure showing the number of linguistic lineages, which is a better measure of deep-level language diversity. It also shows the top 25 countries for biological diversity, for reasons we will return to later.

As the preceding discussion should have made clear, the current distribution of languages reflects many influences. The effects of each region's history have been superimposed on original patterns that are likely to have shown even closer parallels between linguistic and biological diversity:

(1) An original stratum of deep-time language diversity goes back to when all humans were hunter-gatherers. This is visible in regions where people have remained hunter-gatherers until recently. Here there is high language diversity on both measures (i.e., total number of languages and number of independent lineages) except in spread zones such as deserts and other less favoured regions that show the effects of repeated recolonization and cultural pressures to extend intercommunicating networks.

(2) A second stratum results from small-scale agricultural expansion since the end of the Neolithic in some regions, although it was much more recent in some areas. This expansion wiped out hunter-gatherer languages but re-grew a more recent pattern of diversification – with more or less lineage density depending on the time-depth of agriculture – leaving a pattern of large numbers of languages groupable into deep-level families like Indo-European or Austronesian.

(3) The effects of state formation, between 3000 BCE and 1000 CE depending on the area, produced a steady fall in linguistic diversity. In most countries of East Asia (Korea, Japan), Europe, North Africa, and the Middle East the low scores are due to comparable effects over the past two or three millennia; the high score still found in China reflects minority populations at the edges of Han control, especially in Yunnan province where the number of reported languages has grown substantially over the past two decades.[21]

Table 1.2 The top 25 megadiverse countries, for two measures of language diversity and for vertebrate species.

Rank	*Endemic languages*		*Endemic linguistic lineages*		*Endemic higher vertebrate species*	
1	Papua New Guinea (PNG)	856	PNG	74	Australia	1,346
2	Indonesia	688	Indonesia	57	Mexico	761
3	Nigeria	480	USA	52	Brazil	725
4	India	390	Brazil	35	Indonesia	673
5	Australia	353	Australia	30	Madagascar	537
6	China	304	Peru	26	Philippines	437
7	Mexico	293	Sudan	24	India	373
8	Cameroon	241	Colombia	21	Peru	332
9	USA	227	Bolivia	19	Colombia	330
10	Brazil	218	Mexico	16	Ecuador	294
11	Dem. Rep. Congo	186	Venezuela	14	United States	284
12	Philippines	168	Russian Fed.	12	China	256
13	Vanuatu	135	Ethiopia	12	PNG	203
14	Russian Federation	117	Canada	12	Venezuela	186
15	Chad	112	Chad	11	Argentina	168
16	Tanzania	107	India	10	Cuba	152
17	Peru	103	China	9	South Africa	146
18	Malaysia	98	Argentina	9	Zaire	134
19	Nepal	94	South Sudan	8	Sri Lanka	126
20	Myanmar	82	Ecuador	8	New Zealand	120
21	Ethiopia	82	Nigeria	7	Tanzania	113
22	Colombia	82	Chile	7	Japan	112
23	Canada	73	Turkey	6	Cameroon	105
24	Sudan	72	Solomon Is.	6	Solomon Islands	101
25	Solomon Islands	71	Paraguay	6	Ethiopia =	88
					Somalia =	88

(4) Most recently the effects of expansionist colonization by Europeans and the elites of the nation states they created have been to wipe out indigenous linguistic diversity in many of their colonies. Uruguay, Cuba, Haiti, and all islands of the Caribbean – ironically, the only sea to be named after an indigenous language (Carib) – have the dubious distinction of having completely silenced their indigenous languages.

If we look at countries like Australia, the United States or South Africa, colonized by Europeans in the past few centuries, we get very rapid rates of language loss under the impact of English.

Comparably rapid rates of language extinction are occurring in much of Brazil under the impact of Portuguese, in Siberia under Russian, in the Sudan under Arabic, throughout Indonesia under Indonesian, and even in some quite remote parts of Papua New Guinea under Tok Pisin, the newly developed national lingua franca.

Looking back beyond the recent flattening of multiformity by colonial languages – and the scores in Table 1.2 largely bracket this off by giving known numbers of languages at the moment of colonial contact – we can see strong correlations between linguistic and biological diversity. Arizona linguistic anthropologist David Harmon first looked at this correlation in an important 1996 study, and since then his findings have been replicated world-wide on a country by country basis,[22] confirmed for Africa at a coarse resolution,[23] and at quite fine resolutions for the Americas[24] using passerine birds as the index of biodiversity and sampling geographical cuts down to squares one degree wide. More recent approaches to this question have used broad ecological areas instead of the rather accidental boundaries given by countries, and again found strong correlations.[25] Ten of the world's top dozen 'megadiverse' countries on biological measures also make it into the A-league of the world's top 25 countries for endemic linguistic diversity.[26] Harmon's work also makes it clear where both types of diversity are concentrated: Central and South America, tropical Africa, South and Southeast Asia on through Indonesia, Melanesia, and Australia to the Western Pacific.

The Wellsprings of Diversity in Language, Culture, and Biology

> [Continued loss of biocultural diversity will] staunch the historical flow of being itself, the evolutionary processes through which the vitality of all life has come down to us through the ages.
>
> (Harmon 2002: xiii)

The arguments for conserving diversity are similar whether we consider the loss of a rare bird or tree species, a body of cultural knowledge that will soon be forgotten, or an endangered language. Since Darwin, we have begun to articulate, at the scientific level, what most cultures have had enshrined in their aesthetics and cosmologies for a long time: that variety is the reservoir of adaptability.

Having a genetically homogenous population of Cavendish bananas is great for maximizing yield and efficiency, but it just takes one new strain of fungus to wipe out the world-wide population. The traditional (agri)cultural practices that were displaced by the technology-driven Green Revolution, with its fertilizers, standardized seeds, and productive new breeds, are increasingly being seen as having strengths that were often overlooked 40 years ago: drought resilience, disease resistance, lower demands on local water tables – and, ironically, higher yields if we factor in water consumption rather than just tons of yield per area. The many 'landraces' of traditional agriculture around the world are now being genetically archived,[27] so that at least it is possible to draw on their genetic diversity. But we also need all the cultural knowledge that grew up around them. Which variety should be planted in which conditions? What grows best where? Which is resistant to which crop disease? Which are good companion plants? Agriculture, like all technology, only works when artifacts marry know-how, and if we only store seeds without the accompanying knowledge we still have an impoverished picture.

Within the western scholarly and scientific traditions, we can identify two types of attitudes to global knowledge. One is universalizing, and sees it as possible to incorporate all knowledge into

the world language of the era – Latin, Arabic, French, and English have all had their turn – spoken by a 'unified mankind within a single unified realm, subscribing to a universal value system'.[28] The other recognizes the strength and richness that comes from distinct traditions that can never be straightforwardly mapped onto a single value system speaking a single world language: 'any reduction of language diversity diminishes the adaptational strength of our species because it lowers the pool of knowledge from which we can draw.'[29]

Joshua Fishman outlines this alternative view, showing how it has been developed by a succession of thinkers from Vico and Herder to Boas, Sapir, and Whorf:

> the entire world needs a diversity of ethnolinguistic entities for its own salvation, for its greater creativity, for the more certain solution of human problems, for the constant rehumanization of humanity in the face of materialism, for fostering greater esthetic, intellectual, and emotional capacities for humanity as a whole, indeed, for arriving at a higher stage of human functioning ... the great creative forces that inspire all humanity do not emerge out of universal civilization but out of the individuality of separate ethnic collectivities – most particularly, out of their very own authentic languages.
>
> (Fishman 1982: 7)

Fishman's wording makes it clear that this is neither a plea for the thousands of particularistic small societies to become mutually isolated museum pieces, nor for a few cute local words to be lifted into a world language like English to form a sort of linguistic theme park. Rather, it recognizes the deep creative interactions and synthetic insights that are revealed when we look at one language or culture through the prism of another.

I will assemble the evidence for the humanistic and scientific value of Warramurrungunji's bequest as this book unfolds. But to close this chapter, as we have been talking about the relation between linguistic diversity, species diversity, and ecology, let us look at some of the ways in which small languages hold detailed biological and ecological knowledge, which generations of speakers have gradually discovered and recorded in their languages. These examples will also show how the loss of our linguistic heritage is intimately tied up with the loss of cultures and habitats. The sorts of knowledge and vocabulary I am about to turn to are typically the first things to be lost when speakers of a language are shifted from their traditional lands to a reserve in Oklahoma, a rubber plantation in Malaysia, a ghetto in Ibadan, or pushed into a sedentary lifestyle where they no longer practice their traditional ecological knowledge.

Words on the Land

> [I]ndividuals draw on cultural resources to structure and accomplish problems with which they engage in everyday socially organized activities.
>
> (Saxe and Esmonde 2005: 173)

Small languages and societies have kept their place in the world by being finely tuned to their local ecologies and amassing a rich fund of knowledge about them. Much of this has been carried forward only in their languages. Many aspects of their traditional knowledge are still unknown to western science, and in fact languages are arguably the most important and distinctive of the 'cultural resources' that Saxe and Esmonde are referring to in this section's opening quote.

Consider Seri, spoken by around 500 hunter/gatherer/fisherpeople in Baja California in Mexico (see Figure 1.3). This is most probably an isolate language without known relatives,

Figure 1.3 A group of Seri people in what outsiders call the 'desert', but which they call *heheán* 'place of the plants'. (Arizona State Museum, University of Arizona, James Manson (photographer), JWM ASM-25114, reproduced with permission of the Arizona State Museum.).

although some linguists argue it is a southern outpost of the Hokan languages of California. In the course of documenting the Seri lexicon, linguists Edward and Mary Moser were told by Seri speakers about their use of eelgrass (*Zostera marina L.*) as a source of grain, leading to the involvement of ethnobotanist Richard Felger. The resultant research was published in *Science* with a title appropriately including 'discovery of its nutritional value by the Seri Indians'.[30] The authors concluded that this is the only known case of a grain from the sea being harvested as a human food source, and emphasized its considerable potential as a general food resource for humankind, which can be cultivated without fresh water, pesticides, or artificial fertilizer. Despite its potential importance in a world likely to need new crops, this crucial knowledge had been known only to members of the tiny Seri world, held inside their language. Many other words in Seri contain information about the treatment, products, and harvesting of eelgrass. For example, the month of April is called *xnois iháat iizax* 'moon of the eelgrass harvest', and the onset of harvest time is signalled when the black brant bird known as *xnois cacáaso* 'the foreteller of eelgrass seed' dives into the sea to feed on the plant.

All around the world indigenous people transmit, through the words and expressions of their languages, the fruits of millennia of close observation of nature and experimentation with its products. In Arnhem Land there have been a number of cases where the impulse for western natural scientists to recognize new species has come from indigenous traditions of taxonomic naming. The large and striking Oenpelli python, long known to Kunwinjku speakers as *nawaran*, is one such species: it was first incorporated into western scientific taxonomy in 1977 as *Morelia oenpellensis*.[31]

Meanwhile a host of native bees of great economic importance to local Aboriginal people, each with their own honey and wax types, has long remained unidentified by entomologists. Until work published by Aung Si and Margaret Carew in 2018, their English identification in bilingual dictionaries had to be limited to the inadequate designation *Trigona species*.

Traditional cultures also contain detailed knowledge of the healing properties of plants, transmitted in local languages. The recent discovery of a drug, prostarin, effective against HIV-type 1 goes back to a conversation between Samoan tribal healer Epenesa Mauigoa and ethnobotanist Paul Allen Cox about traditional medicinal uses of the stem of a particular tree, *Homalanthus nutans*. The fact that Cox had learned Samoan as a missionary's son was a key fact in enabling this conversation. Comparable curative potentials abound across the immensely variegated world of traditional ethnobotany, and their full investigation requires the collaboration of traditional healers, ethnobotanists, and linguists.

A recent study by Cámara-Leret and Bascompte (2021) tallied up the total known 'medicinal plant services' – pairings of a medicinal plant species and a medicinal subcategory (e.g., Ficus insipida + digestive system) across three regions with high biocultural diversity – North America, New Guinea, and Northwest Amazonia. Of the total of 12 495 such 'medicinal plant services' they found, across a sample of 236 indigenous languages, 75% were linguistically unique – in other words, just one language in their sample records this connection. They conclude: 'Our finding of high uniqueness in indigenous knowledge and strong coupling with threatened languages suggests that language loss will be even more critical to the extinction of medicinal knowledge than biodiversity loss.'

Vocabularies of indigenous languages are not just guides to plant use, but often also show the ecological links between plant and animal species. Throughout Arnhem Land the spangled grunter fish bears the same name as the native white apple tree, *Syzygium eucalyptoides*, because this fish eats the fruits that fall from this tree into creeks and billabongs:[32] in Kunwinjku both are called *bokorn*. Knowledge of this link is of obvious value to anyone who happens to be out fishing for spangled grunter: look for the tree, and in the water below you are likely to find its fish 'mate'. The languages of Central Arnhem Land abound in such pairings, making them a veritable fisherman's guide to the area.

As another example of traditional ecological knowledge, consider the Arrernte language of the Alice Springs area, where various types of grub are an important source of food. Arrernte[33] has a special method of naming grubs after the bushes where you can find them: *atnyeme* 'witchetty bush' yields the *atnyematye* 'witchetty grub', *utnerrenge* 'emu bush' yields the grub known as *utnerrengatye*, and you can work out for yourself the name of the grub found in *thenge*, the ironwood tree.

A host of examples like these lead out into the vast ethnobiological wings of our Library of Babel. But all such knowledge is at great risk, as long as it is only available in little known languages spoken by just a few hundred people, since a shift to another language can cut off its transmission. Once we go over to calling the *bokorn* fish a 'spangled grunter', and the *bokorn* tree a 'white apple', our words no longer deliver the ecological link between them.

Further Reading

Trigger (1987) and Garde (2008) contain detailed discussions of locally appropriate language uses in Aboriginal speech communities. Feil (1987) gives demographic figures for the New Guinea Highlands showing the correlation between intensification of agriculture and language

size. The fascinating multilingual situation in the Amazonian Vaupés is discussed in Jackson (1983) and Aikhenvald (2002). Moore (2004) examines how children become multilingual in up-country Cameroon, and Thurston (1987, 1992) gives a rare and entertaining watch-it-happen study of the processes and results of 'esoterogeny' in New Britain. Sutton (1978) discusses Cape York clan-lects, and the origins of Iwaidja mutation are in Evans (1998); Evans (2003a, 2018a; 2019) set out a model for how ideologies of difference can drive the emergence of clan and other variation. For the situation in pre-Roman Italy see Pulgram (1958) and Robb (1993) on ancient Europe more generally. Ostler (2005) gives a wide-ranging discussion of the fates of languages and past empires in the historical period. On the Taino language of Cuba see Álvarez Nazario (1996).

Nettle (1999) is a pioneering essay on the causes of language diversity, while Collard and Foley (2002) compare the distribution of biological and linguo-cultural diversity, pointing out the need to take historical forces into account on top of the basic ecological determinants. Maffi (2001, 2005) treats the links between language and biodiversity at a global level, Turvey and Pettorelli (2014) do this in great detail for New Guinea, and Hua et al. (2019) provides a recent study eliminating the methodological difficulty of spatial autocorrelation that has been a problem with earlier research on this topic.

The classical world-wide listing of the world's languages is Ethnologue 2005 (www.ethnologue.com; Gordon 2005), maintained by the missionary organization Summer Institute of Linguistics, though its speaker-population estimates are sometimes significantly flawed;[34] an alternative source, arguably now superior to Ethnologue, is Glottolog, which also contains the best current world-wide classification (Hammarström et al. 2020). *The World Atlas of Language Structures* (Dryer and Haspelmath 2013) is a mine of information on the geographical distribution of 144 linguistic features, and allows you to compile your own maps and tables from a vast database.

On Seri ethnobotany see Felger and Moser (1973, 1985); the latter includes descriptions of how eelgrass seeds are harvested. Arvigo and Balick (1993), Balick and Cox (1996), Cox and Balick (1994), and Sofowora (1982) discuss ethnobotany and traditional herbalism. The Arrernte examples are from Wilkins (1993), and the book it is in (Williams and Baines 1993) contains an interesting collection of papers on the now-burgeoning topic of traditional ecological knowledge; for an exemplary discussion of the interweaving of ethnobiological and linguistic knowledge in the Solega people of Southern India see Si (2016). Good discussions of other highly specific indigenous biological knowledge, and arguments linking the loss of linguistic and biological diversity, are in Loh and Harmon (2014) and Nettle and Romaine (2000).

Notes

1 Laycock (1982: 33).
2 Gossen (1984: 46–47).
3 Based primarily on the author's own field notes, with additional input from Murray Garde (pers. comm.), Mark Harvey (pers. comm.), and Birch (2006).
4 Green (2004).
5 When he told me this story in the late 1990s, Pluto was a gentle old man, around 80 years old. Violent as it is, his story promised to provide crucial evidence in an upcoming Native Title claim. Against similar claims elsewhere in Australia government lawyers had advanced the argument that, if traditional laws were only described as ideals without case studies of sanctions taken against violators (including white people), their evidentiary status was weakened. Although his evidence was potentially important, it could also have led to charges being mounted against him

for a murder that until now had not appeared on the radar of the Australian justice system. Asked whether he was willing to submit his statement and take the risk of perhaps being sent to prison, Pluto looked around the room of the old people's home and laughed – 'ah, so they might send me to prison, eh?' Undeterred, he signed off on his affidavit with a thicket of spidery marks, grasping the pen like a hand axe. In the end he passed away just days before the official hearing commenced, and he was buried in his own country on Bentinck Island. Parts of his affidavit, including the statement recounted here, were read out at his funeral.

6 See Singer and Harris (2016) and references therein.

7 François et al. (2015).

8 Lynch and Crowley (2001: 6).

9 Campbell (1997: 122).

10 Laycock (1982).

11 Nettle and Romaine (2000: 88).

12 Lee and DeVore (1968); a similar figure is given in Hassan (1981), although some scholars, such as biological anthropologist Rob Foley (pers. comm.), think a figure of 6–7 million is more likely.

13 Daniel Nettle (1999: 102), who goes through more careful estimating procedures than I have given here, nonetheless comes to a very similar estimate, giving 'the late Palaeolithic language diversity as between 1,667 and 9,000 languages'.

14 I base this date on the assumption that the ability to use language only evolved once and must thus not postdate the archaeologically attested dispersal of modern humans, coupled with extrapolations from the degree of typological divergence of modern languages paired with known dates like the settlement of Australia some 50 000 years ago. (Though see Evans (2018b) on the idea of 'polysemigenesis' in which different intermarrying groups each evolve some design features of language that are then integrated through multilingual intermarriage.) Though the topic remains controversial, there is growing evidence that other hominins also had language, in particular Neanderthals (Dediu and Levinson 2018); if this is true then the date for the emergence of language may be substantially older, since humans and Neanderthals diverged some 650 000 years ago and on this argument language in some form would have already have been present.

15 See e.g., Bellwood and Renfrew (2002).

16 Pulgram (1958: 268).

17 The Etruscan language of the Romans' original mentors is unusual in that we know exactly when it was last spoken. In 408 CE Rome was threatened with destruction by Alaric, king of the Goths. '[S]ome Etruscan priests went to the emperor, offering to perform certain magic rites and recite Etruscan prayers and incantations to ward off the enemy. But they were unsuccessful, for Rome was sacked, and it was the last time the Etruscan language was spoken' (Bonfante 1990: 328–329).

18 Nettle and Romaine (2000).

19 This estimate is from Nettle and Romaine (2000). Their ranking of language-diverse countries differs somewhat from that given in the first column of Table 1.2; they include Ghana (72), Benin (51), Vietnam (86), and Laos (92), and omit a number of others.

20 My source for data in the third column is Harmon (1996), whose own sources for species are derived from WCMC (1992: 139–141). Figures for the first two columns are from Glottolog; I thank Harald Hammarström for crunching the data to get these rankings. 'Lineages' here represent groupings at the level of Germanic or Romance within Indo-European, or Semitic within Afroasiatic – I do not count the higher-level groupings because they take longer to establish and we are further from having an equivalent understanding of how these work for all the world's languages 'Higher vertebrates' is just one possible measure of biodiversity; these numbers include mammals and bird, reptile, and amphibian species. Species figures for the following countries do not include reptiles, because the number of endemic species is not reported in the source table: United States, China, Papua New Guinea. Column 1 and column 3 figures for Ethiopia include Eritrea.

21 The rankings and numbers in this edition have shifted somewhat since the first edition of this book, reflecting shifts in our knowledge of the world's languages. The idea for the first and last

original table came from Harmon (1996), whose own sources for endemic language numbers are derived from Grimes (1992), and for species from WCMC (World Conservation Monitoring Centre) (1992: 139–141); in the first edition I calculated the second column figures from the appendix to the first edition of World Atlas of Language Structures (Dryer 2005), descending to the second-level clades, e.g., 7.6 (Semitic) or 9.10 (Slavic), and counting how many such families were represented in each country on their (admittedly incomplete) list. Since then, more up to date language figures have been compiled via the online resource Glottolog, which is what I have used here. The most noteworthy shifts are the large growth in the number of languages in China, reflecting many recently discovered in remote villages in Yunnan province in particular (Erard 2009), and the growth in the number of lineages in several countries – especially Papua New Guinea, Sudan, South Sudan, and Chad – as new isolates are discovered and as more sceptical appraisals of old 'lumping' classifications discard older unfounded groupings.

22 Harmon and Maffi (2002).
23 Moore et al. (2002).
24 Manne (2003).
25 Stepp et al. (2004).
26 Note: 'endemic diversity' counts only languages whose original historical association is with where they are spoken: Hokkien speakers in Malaysia, Hindi speakers in Fiji, and English, Spanish, Portuguese and French speakers anywhere in the New World will be discounted. Clearly if we count immigrant languages in cosmopolitan first-world cities the picture would be changed.
27 And, to be fair to the scientists of the Green Revolution, many were involved in establishing this process of archiving land-races.
28 Fishman (1982: 6).
29 Bernard (1992: 82).
30 Felger and Moser (1973).
31 Torr (2000: 4).
32 Evans (1997).
33 You can hear what these words sound like on the online Arrernte dictionary at https://arrernte-angkentye.online.
34 See Gerrand (2007) for some examples of problems with Ethnologue's speaker-population estimates.