1 *The Role of Social Interaction in the Human 'Carrying Capacity' for Language and Culture*

1.1 An Epiphany

Some years ago, I was working as a linguistic anthropologist on a remote Pacific island, home to a unique language and culture unrelated to anywhere else. As I was coming from bathing in the river, a young man came up and, smiling, shook my hand. He gestured in a complex way while whistling through his teeth, and I quite quickly realized he was completely deaf, and gesture was his sole means of communication. He seemed to be saying by gesture that he lived over the hill in a village on the far side, and he knew all about me, and how I had come by boat from far away (he pointed across the ocean) with lots of luggage (he mimed the unloading and carrying of bags). I found that by pointing and gesturing I too could ask questions. I caught my breath. Wait a moment, I thought, we don't share a language, and we don't even share a culture, we just share a fleeting moment in time and space. But we seem to understand one another! This should not be possible, if all the things that linguists and philosophers have said about language and communication are true: communication should only be enabled by a shared conventional code with the kind of properties that make a language – arbitrary symbols, grammar for combining them, and ways of talking about other times and places using these devices. But there we were, two humans from different ends of the planet, able without a shared language or background to connect.

The young man and the island he comes from will play a role in the chapters that follow. But this observation – the possibility of communication without a shared code or set of conventions – opens up a wealth of potential insights into the shared humanity that underlies the rich variation in language and culture, and indeed the wellsprings of all those characteristics that distinguish us from our primate cousins. This book follows this scent where it leads.

1.2 Unity in Diversity: A Central Puzzle in the Human Sciences

Humans are by any measure an unusual species. We are wildly successful (humans outweigh the biomass of all other non-domesticated mammals together by nine times) and have colonized every continent and almost every niche on the planet. All eight billion of us are biologically closely related, more so indeed than neighbouring groups of chimpanzees are to one another. But, partly by adaptation to those varied niches, we have differentiated in superficial characteristics, in body shape, skin, and hair, but above all in cultural form – clothing, customs, languages, technology, subsistence modes, and beliefs. While variation might be expected, there are no other animals that just by virtue of belonging to different social groups have totally distinct diets (some humans live purely on milk, meat, and blood, others solely on plants), distinct mating and kinship systems, and incommensurable communication systems. Cultural specialization has overridden the underlying biological unity with a kaleidoscope of variation. The degree of variation is surprising, even to seasoned observers: things we in the West take for granted, for example the nuclear family and the recognition of fatherhood, representational art, private property, money and markets, or even languages with nouns and verbs, are absent or moot in some other societies.

Human languages and cultures are so diverse that they resist the extraction of simple universal architectures. Although the celebrated linguist Noam Chomsky has influentially argued that all languages have an underlying commonality of structure, that may amount by his own admission to little more than the ability to put two bits together to form a larger structural and meaningful unit. Nearly all the substantive shared structural properties of language that have been proposed as universals turn out to be weak tendencies with many exceptions, and the observable tendencies may often have more to do with common historical origin than with structural or cognitive necessity. There remain, however, discernible tendencies for languages to prefer some patterns over others, for example a consistency in word order, but these have the character of biases rather than hard constraints.[1] The

[1] It has recently become possible due to better databases to assess the strengths of these biases; see e.g. Verkerk *et al.* (submitted). When geographical proximity

same sorts of problems are faced in the domain of kinship, where exceptions to the role of the nuclear family, recognized paternity, marriage, and the like can be found. As we get better and better databases for languages and cultures, the absence of simple generalizations becomes reinforced, and the strength of the signal of relatedness due to particular historical descent and borrowing becomes stronger.[2] Insofar as anthropology or other human sciences have ambitions to be a general science of humanity, it will be necessary to look deeper, not at the structural forms exhibited in language and culture, but at the processes that produce them.

In this book I propose that the foundations of language, and thus the many aspects of culture that depend on it, lie in a parallel system, namely the social interaction which provides the cradle in which languages are learnt, forms the major constraints that languages must meet, and plays a crucial part in shaping the nature of human communication systems.

The tension between human diversity and the underlying unity of the species forms a central dilemma of the social sciences. Let me illustrate with an anecdote. In June 1848, HMS *Rattlesnake* sailed from Cape York in Queensland to the Louisiades, an archipelago lying to the east of New Guinea. On board was a young medic, Thomas Henry Huxley, lovesick from separation from his spouse-to-be in Sydney. The ship's first landfall was Rossel Island, but due to the encircling barrier reefs (and an apprehensive and fatally sick captain) they could not land – and in fact no Western vessel had ever penetrated the reef. But what the young Huxley noted was that both the houses on Rossel and the canoes were of different construction than those of the next island – 'they can hardly have two fashions of canoes in islands twenty miles apart?' he remarked in his diary. Days later, inside the reef about a hundred miles on, near Nimowa Island, he noted a canoe from Rossel Island, and he found he could communicate with the locals easily enough to barter fresh vegetables, water, ethnographic specimens, and the like. Communication, up to a point, was possible without a shared language or culture. Although Huxley didn't then know that,

and shared ancestry are controlled for, the patterns tend to halve in strength, but are still often discernible.

[2] On languages, see Evans & Levinson 2009, Levinson & Gray 2012, Skirgard *et al*. 2023.

he and the islanders last shared an ancestor perhaps 40,000 years ago, yet despite this huge separation in time and space, some mutual understandings and simple negotiations could be achieved.

There is our dilemma – cultural diversity and human commonality – a presumption of similarity that makes communication possible without a shared language. Had he been able to land on Rossel he would have found there a language entirely unrelated to the one spoken at Nimowa where he had anchored (indeed unrelated to any existing language, including sounds not found in any other language), and a cultural system remarkably more complex than the Nimowa one, with a bilineal kinship system, a full Olympiad of gods celebrated in verse who inhabit protected reserves, and a baroque shell 'money' system that was to exercise more than a few economic anthropologists in decades to come.[3] The language of Rossel Island, called Yélî Dnye, has the largest phoneme inventory – ninety contrastive sounds – of any language in the Pacific.[4] Forty-seven of these sounds involve nasality, with air sent through the nasal chamber (like English *m* or *n*). Just 550 km across the ocean there is another island, Bougainville, with thirty-odd languages; one of these is called Rotokas, which has, on one count, just eleven contrastive sounds in the entire language and no nasal consonants at all, one of the smallest and most unusual inventories in the world. Overall, the cultural diversity of New Guinea is extraordinary: it has some 1,300 languages, or a fifth of the world's total, of at least forty different, unrelated families.[5] A crucial part of the peculiarity of our species is that we are the only known animal with a communication system that differs across social groups so fundamentally at every level, from sound to meaning to syntax. That pliability may be part of our evolutionary success, reflecting adaptations to local ecology and social systems and to cultural technology. Huxley went on to become 'Darwin's bulldog', the man who first noted that birds were dinosaurs, and the pre-eminent academic scientist of Victorian Britain, insisting against the tenor of the times that racial differences were entirely superficial, but with little to say about the cultural and linguistic diversity that his ethnologist colleagues were busy cataloguing at the time.

The superstructure of language seems on all the evidence to be largely socially constructed in sound, meaning, and grammar. Languages

[3] Armstrong 1928, Liep 2009, Levinson 2022a. [4] Levinson 2022a.
[5] Palmer 2018.

differ of course, just as customs do, but unrelated ones do so in such profound ways that it is hard to find things that they all have in common (see Chapter 2). But there are also anatomical and neurocognitive adaptations for language – most obviously in the human vocal apparatus itself, with the resonant vocal chamber and the agile tongue, highly controllable glottis and epiglottis, giving us an infinite phonetic space within which languages construct a sound system. Even our breathing control system has a special voluntary divert to power the speech apparatus. And then there are brain adaptations of which the most anatomically prominent is the extension of the arcuate fasciculus, the white fibre bundle connecting (roughly) Broca's area (behind the temple) to Wernicke's area (above the ear). Further, there are genes that from various language disabilities we know contribute crucially to our language capacity. What is striking though is the gap between all these physiological underpinnings and the diversity of languages – as if evolution went on holiday before finishing the job. And yet, obviously, there's a puzzle here: how can these physiological adaptations have arisen without closing the arc from brain to behaviour, that is, without a functioning communication system? How can we have evolved such developed infrastructure to adapt to a moving target, the kaleidoscopic variety of languages?

I will argue that the answer is that there was an antecedent communicational infrastructure which allowed the processes of evolution to slowly build out structures at both ends, the cultural and the biological, a system which still to this day completes the arc, and which remains relatively unappreciated and understudied. There are alternative explanations; for example, perhaps the species once had a single communication system like all other mammalian species, and then cultural evolution invaded and diversified it (the Babel story). Or perhaps the physiological adaptations were only pre-adaptations for language, for example being originally evolved for singing for mates. But I will leave those alternatives for others to explore.

The communicational infrastructure that seems to provide the crucial bridge is what I have called 'the interaction engine', a term of art intended to indicate that this is not a 'module' or single adaptation but a loose assemblage of capacities and proclivities, the sort of thing an alien ethologist would note of humans: they huddle together, look each other in the eye, gesticulate, make fleeting facial expressions, blink systematically, and most prominently take turns issuing bursts

of vocalization coupled with facial and gestural indications, doing this thousands of times a day in closely timed alternation. This is the matrix in which the great bulk of language is used, and crucially in which it is learned. From an energetics perspective, it is a huge diversion of energy away from the biological essentials. I will show later that this interaction is much more systematically organized than it seems at first sight, and that this system is – unlike the very languages it affords – remarkably uniform across all human social groups. But the present point is that it provides the keystone to the arc, the connection between the biology on the one hand, and the culturally evolved communication systems we call languages, on the other.

1.3 The Interaction Engine: A Base Providing the 'Carrying Capacity' for Language and Culture

I will argue, in a nutshell, that the keystone in the arch between the cultural elaboration of languages on the one hand and the biology supporting communication on the other is an interactional base which essentially provides the 'carrying capacity' for language and culture. This base, the interaction engine, consists of a package of ethological properties, a collection of pan-human predispositions. Although many authors have pointed to the special role that individual parts of this system contribute to language (for example, awareness of other minds, or the importance of inference, gaze, or gesture), few have thought about it as an elaborate package of many rather different capacities.[6] We cannot treat all these elements here, but we can pick out four ingredients that will play a special role in what follows:

1. Multimodality
2. Timing
3. Contingency
4. Intention recognition or 'mind reading'.

A few remarks characterize the key role of each of these factors here, but these topics are central to the book and will recur again and again.

By multimodality is meant the use of multiple simultaneous channels (auditory, visual, tactile) during communication, for example the

[6] A partial exception is Clark 1996, a highly readable introduction to language use.

gestures, smiles, nods, and winks that accompany speech. This is not restricted to humans of course – the proud robin puffs its chest and sings with its beak pointing to the heavens, or an aggressive goose lowers its head and honks. The multimodal nature of human communication may be self-evident, but it is downplayed by the emphasis on the written word – in face-to-face conversation, not only are the full resources of vocal expression (the voice quality, the intonation) employed beyond the mere sequence of phonemes, but the hands, the face, indeed the full ventral surface of our bodies exposed by our bipedalism are involved in communication. Notice that most of these expressive possibilities are in abeyance when listening – indeed gesture with the hands is a signal of who is speaking. The existence of manual sign languages shows that our extraordinary communication prowess does not rely fundamentally on the vocal stream. There are biological adaptations to this multimodality, including the white sclera of the human eye which makes it clear when we have the gaze of our interlocutors.

The second crucial characteristic is timing. Timing plays a central role in human communication – spoken language is only possible because of the extraordinary orchestration of over a hundred muscles, each playing a particular role on a millisecond-by-millisecond scale. But critical in interaction is response timing. In informal conversation, responses take place on average within about 200 ms (milliseconds), that is quite literally the blink of an eye. The implications of this will be explored further, but what will be shown is that this speed of response can only be achieved by predicting the end of the other's turn at talk. We have a delicate sensitivity to timing in interaction – a hesitation of a few hundred milliseconds after a request is likely to be taken as a 'no'. The in-built transmission delays in video communication upset our native rhythm and help to make this a tiring form of interchange.[7]

Responses of course have to be appropriate to the prior turn – a greeting requires a counter-greeting, a question an answer. This third property, contingency, is crucial, and it is not trivial because, firstly, language is infinitely generative, so utterances are often novel, making the selection of an appropriate response anything but automatic. Secondly, the contingency is built not on surface form but on intent, or

[7] Boland *et al.* 2022.

perceived action. For example, an exchange might go 'Where did you get that amazing dress?' 'Oh, thank you, you are the second person to admire it', where the first utterance is understood as a compliment rather than a question. The contingency of such actions has, it turns out, its own syntax, as we will see in the chapters that follow.

The ascription of an action to an utterance is, outside simple routines like greetings, a complex inference of likely intent. The form of the utterance constrains the inference, as crucially does its context, both local and global, textual and extra-linguistic.[8] In constrained contexts, likely moves and previously attested responses can guide inference, but outside these contexts this 'mind-reading' ability has resisted scientific understanding. The same or a similar kind of inference is involved in the action domain, where you hold out your glass and I refill it with wine. Without this, the rapid sustained cooperation essential for most human activities would not be possible, whether hunting together, building something together, or lifting something together.

These ingredients – multimodality, timing, contingency, and intent recognition – help to build a system of rapid, relevant responses with attribution of purpose or intent. It is within this system that infants learn their first language, and learn the rules for appropriate conduct. Through it they also inculcate the values of their society. In short, it is this system that carries language and culture across generations, thus providing our unique carrying capacity for learned communication systems and learned behaviours. The chapters that follow amplify this brief characterization of the interaction engine and trace its impact on the origin of language and human social life.

[8] There is a large literature here, see Levinson 2013b and 2024 for references.

2 *Human Communication and the Interaction Engine*

2.1 The Puzzle of Language Diversity

There are over 7,000 languages in the world today. Just as species can be grouped into taxa with their known relatives, languages can be grouped into language families, which are the largest groups where their member languages can be shown to be related. The 7,000-odd languages belong to over 400 such families – the relation between these families goes so far back in time that we cannot easily recover it.[1] In general we can track relatedness of languages through their vocabulary or structure to about 10,000 years back, although occasionally we may be able to go a little farther. Over the last quarter of a million years, an estimate for the time depth of anatomically modern humans, there have been perhaps half a million languages. Languages do not just differ superficially, as if they were the same basic structure in a different set of sound clothing as it were. Instead, they differ in every possible way: some have as few as a dozen distinctive sounds, others twelve times as many (depending a bit on how you count), some have such complex morphology (ways of building words from words) that a whole English sentence can be expressed in a single word, while others have no morphology. The literature is full of claims that all languages exhibit some structure, but these claims are based on inadequate samples. We now know for example that basic sentence structure is highly variable, and that all possible phrase orders or even no set phrase order can be found in different languages. If structural diversity is like a wild garden, so is meaning: languages differ wildly in the concepts they choose to lexify (encode as words) or grammaticalize. Even the human body, one of the few universal objects, is segmented quite differently

[1] Data from Glottolog (Hammarström *et al.* 2021), which lists 7,606 languages in 425 families or isolates (languages without any relatives, of which there are 181). Some authors think they can discern many connections between these families, but the scientific basis for these larger groupings is weak.

into named parts in different languages. Imagine a language with no words for numbers greater than three, or no words for left and right, or no words for relatives beyond parents and siblings – these are all attested. Instead of strong universals, we find enormous diversity, and instead of a limited set of alternatives, relatively unconstrained variation, albeit with tendencies for structural coherence.[2]

Now contrast all known animal communication systems. In most cases these have a finite set of signals with an instinctive basis, triggered by recurrent events in the environment (for example, threats) or biological needs (such as advertising for a mate or defending a territory). In some species, including songbirds, the shape of the signals may be partially learned, but the 'meaning', the function or triggering events, are fixed. In short, they lack the structural complexity (like the use of meaningless elements to construct meaningful ones, and complex hierarchical patterning), the indefinitely extended meanings, and critically, the deep variation across groups that human languages exhibit.

Examining the structural diversity of human languages shows a number of things. First, despite their cultural nature, languages seem to evolve or change remarkably like biological evolution, inheriting traits faithfully across generations, inventing new structures, losing old ones, although they hybridize more like plants than animals. Secondly, if we imagine a 'design space' constructed from all the known parameters of linguistic variation, languages can be shown to have spread out to explore many of the far corners of this space; only related languages are likely to cluster closely together.[3] What this suggests is that there are relatively few constraints on the directions that languages can evolve in, providing they retain learnability for the next generation. Most extraordinary of all, human languages can flip from the oral mode to the gestural as in the sign languages of the deaf, without losing expressive finesse.

The unparalleled variability of human communication systems does not argue against a biological basis for language: it is patently clear

[2] See Evans & Levinson 2009, Hammarström 2010, Levinson 2003b. On some recurrent patterns and biases, though, see Verkerk *et al.* in press.

[3] Harald Hammarström and I performed this experiment (in an unpublished work) on a sample of languages from the Nijmegen Typological Survey, but this can now be replicated on much larger typological samples, as in Skirgård *et al.* 2023.