3 *Universal Properties of Interaction*

There was a time when it was taken for granted that the basic structure of all languages was underlyingly uniform, with distinct languages merely clothed, as it were, in different words and sounds.[1] Given the much greater richness of empirical data that we have today, this view is no longer tenable, but many scholars still take the view that the core of language involves a universal set of procedures. Along with this view has gone the idea that although the core of grammar is universal, the *usage* of language is culturally highly variable – there is a common tool as it were, put to different uses.[2] Here I will argue the contrarian position, that grammars are highly culturally varied (indeed are largely cultural constructs), while there are surprising and strong universals of language use. I've argued against the view that grammars are highly constrained by universal principles in Chapter 2. The reason for the contrasting strong universal base in language use, I will argue, is that many basic usage principles pre-date language as we know it, belong to our ethology, and share many commonalities with those of other primate species.

A brief word is in order about the notion 'universal'. For the human sciences, the key idea is of course that a universal rule or practice applies in all cultures and societies, and in all normal adults of the species. Linguistics has long had to weaken this concept, because nearly every rule or structure that has been proposed as universal has turned out to have some (and sometimes many) exceptions. Most linguists have therefore adopted a notion of 'statistical universal', a strong tendency for languages to be organized in a particular way. In addition, linguists have noted the extraordinary variety of languages, and have found it much more useful to divide them into (sometimes overlapping) types. So it is possible to say that if a language is of a particular type

[1] Chomsky 1986 on universal grammar; Hymes 1974 on a relativity of language use.
[2] Hymes 1974.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

(or has a particular important property), then most likely it will have another particular congruent property. This type of conditional statistical universal has turned out to be the most useful way to describe the striking linguistic diversity on the planet.

In contrast, in the domain of interaction it appears to be much easier to specify absolute or unqualified universals, of the kind that informal conversation in all languages will exhibit quite precise properties. If specified carefully, to allow for the superimposition of cultural patterning, these may be almost exceptionless behavioural norms. Given the huge variety of human ways of life, there may be no other domain where we are able to specify comparable near exceptionless norms of human behaviour, which surely speaks to the importance of this area in the human sciences.

3.1 The Cooperative Umbrella for Human Communication

One of the central puzzles in human evolution is the cooperative nature of human social life. The psychologist Michael Tomasello has argued that it is not sheer smarts that distinguishes us from apes, but the fact that our intelligence is geared to cooperation and joint activities, while apes are geared to competition.[3] This is not easy to explain in terms of standard evolutionary machinery, where natural selection favours individuals in competition with each other. How then do we explain human altruistic cooperative behaviour, both trivial acts of consideration (like giving a stranger route directions) or the ultimate sacrifice (like risking death to save a stranger from drowning)? In small-scale societies, where most of human prehistory was spent, nearly everyone is kin, so in helping one's neighbour one may be contributing to the success of one's own genes, and thereby passing on the cooperative urge. Such cooperation might be widened somewhat by hoping for reciprocity – if you share your hunt, and I share mine when successful, we may both be better off. This though requires trust, reinforced perhaps by punishment or purdah.[4] An alternative explanation is to presume that competition is at the level of groups rather than individuals,

[3] Tomasello 2014.

[4] Cooperation between kin was explored by Hamilton (1964) under the rubric of 'kin selection', while Trivers (1971) developed the theory of reciprocal altruism. These are generally accepted evolutionary mechanisms, in contrast to group selection, which may only be operative in culture-bearing species like humans.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

so by contributing to the success of the group one helps to guarantee the success of one's offspring. This 'group selection' remains controversial, but because human groups develop elaborate cultural innovations under cooperative conditions (as when advanced techniques of warfare allow the elimination of rival groups), this may indeed have contributed to the cooperative nature of human behaviour inside social groups.[5]

Regardless of its origins, human communication clearly takes place under the umbrella of cooperative assumptions. Indeed, this serves as a first argument to make the case that language usage abides by universal principles. Consider the fact that language is on close scrutiny pretty sketchy. I say 'The cat is on the mat'. You imagine a typical cat in a typical sitting posture in the middle of a mat on the floor. But that of course is not what I said – I could be talking about a dead cat lying on a mat, the cat might be stretching on a mat on a table, or it could even be a picture of a cat woven into the mat (as in *The maple leaf on the Canadian flag*). The word 'the' presupposes that you and I can identify a particular cat, either in the environment or from prior discourse – it requires contextual resolution. The word 'on' has a range of senses (as in *on the map*, *on Mars*, *on reflection*, and so forth), and the hearer must select one that makes sense. Thus, even a banal sentence requires imaginative fleshing out.[6] So what guarantees that your imagination matches the speaker's intention? The philosopher Grice suggested that we abide by a cooperative principle, which holds that other things being equal we should say what is relevant, timely, and true, while providing full information which is just sufficient for the purpose at hand.[7] Suppose instead I had said the more cumbersome 'The feline is positioned on top of the rug' – whatever picture that invokes it is not that of the familiar cat curled up in front of the door. How one says things matters in terms of what is invoked. Now there's no shortage of counterexamples to a principle of cooperation in talk, but nevertheless this principle does seem generally in operation by default. Suppose I ask my colleague in the office building 'Where's Dan?' and she answers 'He was in the street ten minutes ago', she implies by virtue of this principle that (a) she did see him, (b) she doesn't know exactly where he is now, and (c) he might not be in

[5] Boyd *et al.* 2005. See also Handley & Mathew 2020.

[6] See Searle 1978.

[7] Grice 1975.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

the building. This is because if she *did* know he was in his office, to be cooperative she should have said so: if you can be precise you should be. One application of these ideas that has been much studied is the use of words that form scales, like <one, two, three…> or <all, many, some>. If I you ask 'Are there any beers left?' and I say 'there's one', I conversationally imply not two. But what I say would be true even if there were two; it would however be misleading by the principle that I should cooperatively provide the full information. Similarly, 'Some of the passengers lost their lives' suggests that not all of them died, although if all of them died it would be true that some certainly did. This is how the sketchy message that language outlines is fleshed out into a complete and satisfying message.[8]

Grice suggested that under the cooperative principle there were four main maxims, which he called Quality, Quantity, Relevance, and Manner. Quality stipulates telling the truth, and by extension making genuine speech acts of all sorts (for example, requesting when one genuinely wants the requested item; promising when one actually intends to fulfil the promise, and so forth). Quantity requires that one produces adequate information, but not so much as to obscure the purpose. Relevance stipulates the timeliness and contingency of a response. Manner suggests making one's point as simply and clearly as possible. Grice went on to suggest that these maxims follow rationally from the cooperative principle, and so they may govern non-verbal interchanges too: if I'm helping you build a shed, when I gesture for a screw I don't want a nail (by Quality), nor do I want a great handful (Quantity), and I want it now not later (Relevance), and produced clearly, not wrapped up in a brown paper bag (Manner). There are various more recent schemes reducing these maxims, but they aim to achieve the same effect.[9]

Given this rational derivation, and assuming that the use of language is mostly for cooperative purposes, there is reason to suppose the cooperative principle is universal across all languages – providing one takes into account the fact that in certain circumstances it is culturally appropriate to depart from these rules of thumb. For example, I say 'Hi, how are you?' and you say 'fine' – you may in fact not be fine at all, but in ways that are not appropriate to mention, as we

[8] Levinson 2024 sketches the full gamut of ways for meaning more than you say.

[9] Horn 1984, Sperber & Wilson 1986, Levinson 2000.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

mutually know. Similarly, there are ethnographic reports of cultures where a joking mode negates the veracity of all that is said, or where a traditional greeting is ‘Where are you going?’ and it is equally traditional not to give the true answer.[10] In such circumstances, participants know to relax their expectations; elsewhere they know to activate them. If we are sharing out the chocolates and you say ‘there are three chocolates left’ when in fact you know there are four, we will consider you are a cheat. Although there have been many studies of how these principles work in particular languages and a presumption that they do, there is in fact no careful cross-linguistic survey that empirically establishes their universality. But from first principles one can suppose they must be operative in all cultures. For a start, it would not be possible for the next generation to learn a language in the absence of this cooperative stance: if I taught my child Mary that ‘rabbit’ means ‘rhino’ on one day, ‘lion’ the next, and ‘tree’ on the third occasion, she would never grasp its meaning – veracity and constancy of reference is essential for grasping a new word. Similarly, relevance is crucial for learning the use of language – if my poor child hears ‘Hello’ at random, she won’t know how to use it. If when I say ‘two’ I sometimes mean ‘five’, sometimes ‘one’, she will likewise be flummoxed. On these bases we are on fairly safe ground to suppose these principles are generally applicable in all cultures. This system is crucial, along with gesture, for giving relative precision to the interpretation of utterances, which themselves tend to be brief, vague, and unresolved.

3.2 Timing and Turn-Taking

In Chapter 2 we noted that the rapid alternation of speakers in casual conversation – the core niche for language use – is one of the central design features of human communicative interaction. Human communication in its most central form therefore consists of short bursts of speech alternating across participants. We noted that this design feature has the virtue of allowing us to correct an interpretation revealed by a response in the following turn, in this way permitting the ‘sketchy’ character of language to be effective despite its actual ‘lossy’ character, like a highly pixellated image. And it provides the basis for

[10] See, e.g. Keenan 1976, Senft 2018.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

the contingency that is another crucial design feature, so that a question can be followed rapidly by an answer. Although most of what we say is delivered strictly in our own turns, not all human vocalizations obey turn-taking constraints – these include emotional cries, laughter, sobs, in-breaths, sighs, and the like, and these may belong to an earlier evolutionary stratum, similar to the involuntary cries of apes, which are also often delivered in overlap with a conspecific's vocalizations. They might be considered the fossils in our communication system.

Turn-taking may seem at first sight rather trivial, but the system turns out to be fascinating and complex. Turns are of no fixed length, but are on average around 2 seconds long. How do we know when the prior speaker has finished speaking, given that there is rarely more than the very smallest gap between speakers? The system in casual conversation appears to work with rules like this: a turn at talk properly consists of a fairly minimal unit, typically a sentence or fragment of one; anyone can speak next, unless a particular next speaker has been called upon; on hearing the completion of a unit, the first participant who jumps in gets the right to a turn, and others should desist. In addition, overlaps (two speaking at the same time) are generally minimized, and can carry social opprobrium.[11]

It might seem that turn-taking with rapid alternation and minimal overlap is rationally motivated. After all, wouldn't my speaking over your words mask those words for both me and bystanders? But experiments show that speech is actually a rather poor mask for other speech (otherwise, after all, cocktail parties would be impossible).[12] So, the 'one at a time' rule that can easily be observed probably has deeper origins that we will return to.

The central puzzle is that the alternation between speakers is very rapid, sometimes instantaneous, rarely taking longer than a second, and on average occurring in about 200 ms, that is, a fifth of a second. This is the length of a single syllable, or the duration of a blink. This is about the fastest human reaction time to a single-choice response (a two-choice response is nearer to 350 ms). If what was exchanged was always as simple as 'Hi' – 'Hi!', that speed would still be surprising. But the majority of utterances are sentences or parts of them and such units have a complex structure and meaning which has to be

[11] Sacks, Schegloff, & Jefferson 1974. [12] Miller 1947.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

composed by the speaker and interpreted by the hearer. On the face of it, both that composition and comprehension has to be done within the 200 ms gap, if I am to understand and respond to you on time.

But that is impossible – the speech production system isn't nearly that fast. Under experimental conditions it takes at least 600 ms to produce a simple noun phrase, and over a full second to prepare a sentence. Much work has been done on what happens in the mind during the process of formulating and outputting speech, and there is a detailed analysis of the chronometry by psycholinguists like Willem Levelt. If a participant is shown a picture and asked to name it as fast as possible, it takes on average about 200 ms to fixate the relevant concept, 75 ms to find the word, and 325 ms to mentally encode its form – all before anything comes out of the mouth after a minimum of 600 ms (see Figure 3.1). If the picture is unfamiliar the whole process will take nearer to 1,000 ms, and a short sentence about 1,500 ms.[13]

These processes can't be drastically truncated: the mental lexicon (the dictionary in our heads) is likely to have over 30,000 words to select from,[14] and 100 or so muscles have to be coordinated for the actual output. What is remarkable is that it is as fast as it is. In fact, the speed of speaking breaks Hick's Law, which holds that response latency increases in a non-linear way with the number of alternatives to choose among (and just think of the vast number of words we control). In contrast to speech production, comprehension is much faster – people can understand speech sped up by a factor of three or more. Understanding an utterance allows multiple processes to work in parallel (word recognition, parsing, disambiguation, contextual inference). By contrast, key elements of production have to be serial – you first have to think what to say, then find the words, and then code them into articulatory movements and so on. These processes can operate incrementally, that is, by starting one process and rapidly moving on to the next while the first is still chugging along in the background. Nevertheless, the production of speech remains brutally slow compared to comprehension.

[13] Levelt 1989, Griffin & Bock 2000, Bates *et al.* 2003, Indefrey 2011.

[14] Brysbaert *et al.* 2016 show the average American knows over 40,000 English words. Unwritten languages may offer fewer words, but the largest dictionary compiled of an unwritten language still has over 30,000 words.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

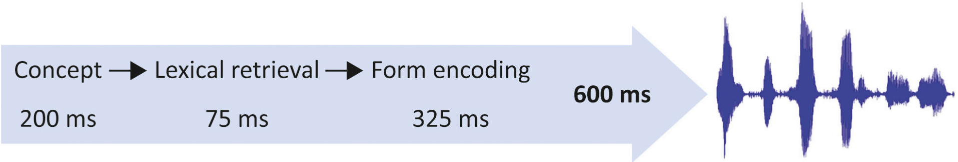

Figure 3.1 Mental chronometry in the production of a single word (Levinson 2016 after Levelt 1989, Indefrey 2011). It takes 600 ms or more before articulation begins, starting with retrieval of the concept, then the word form, and then the preparation for articulation.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

This is a paradox: people are responding at a speed three times faster than the thought processes involved (within 200 ms instead of the 600 ms minimal word production time)! Although people sometimes buffer the response with *uhms*, *wells*, and the like, the only general escape from the paradox is to assume that people start planning their response as early as they can, well before the end of the incoming speech. It is perhaps an uncomfortable thought that in conversation our interlocutors are hatching their responses before we are half way through what we are saying! But in a series of experiments with brain imaging we have shown that this is indeed what people do – at some point, often half way through the incoming turn from another speaker, recipients are busy planning their response. This of course involves predicting how the half-finished incoming turn will unfold. If the planned response is ready to go before the other has stopped speaking then, to avoid overlap, the response must be held until there are key signals of the end unfolding.[15]

There are deep ramifications of having a system like this. It implies that in conversation we routinely double task – while listening we are also involved in speech planning. The diagram below (Figure 3.2) sketches this period of overlapping comprehension and production in the listener, the next speaker.

Humans are notoriously bad at double tasking, which is why using a mobile phone while driving a car is dangerous. But in conversation the double tasking is more extreme, because both comprehension and production use much of the same linguistic machinery, so it is more like juggling balls while dribbling a football! How we achieve this, apparently without extraordinary effort, is currently a mystery.

It is worth stressing that this remarkable speed of response can only work if the addressee is working hard to predict how the incoming turn is going to unfold. Languages have remarkable statistical properties that help this along: the probabilities of a particular completion narrow as the utterance unfolds (you can check the relative probabilities of continuations by typing, for instance, ‘fly a *’ into Google N-Gram Viewer – top responses will be ‘a plane’, ‘a kite’, and so on).[16] Conversation analysts have long recognized that this ‘projection’ of what the other is going to say plays a crucial role in turn-taking, and

[15] Levinson & Torreira 2015, Levinson 2016, Meyer 2023.
[16] Kuperberg & Jaeger 2016.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

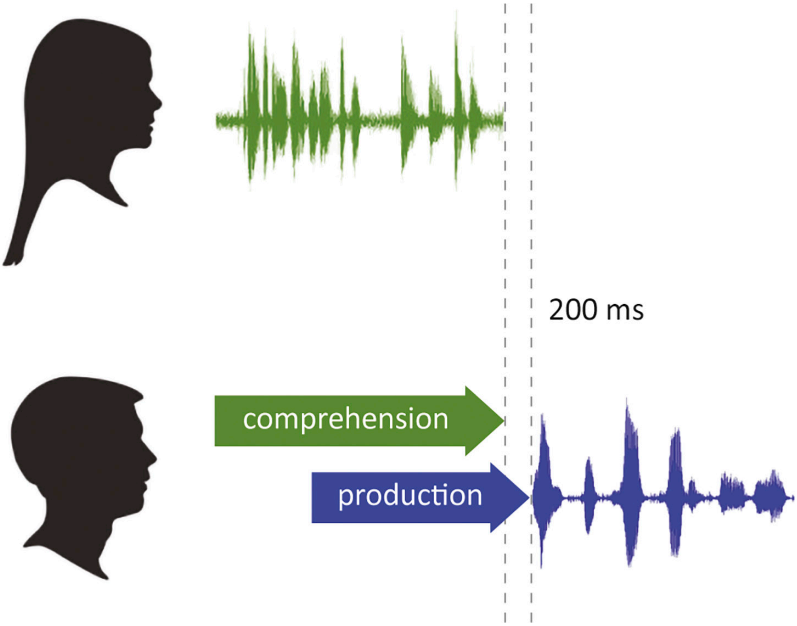

Figure 3.2 The listening interlocutor in conversation has to plan a response early in order to respond in as little as 200 ms, and so must plan in partial overlap with comprehension of the incoming turn (after Levinson 2016).

have noted compelling cases where one speaker finishes the turn of another, as in:[17]

<5> (from Lerner 1996:241, = indicates continuation with no discernible gap)
A: 'When the group reconvenes in two weeks ='
B: '= they're gunna issue strait jackets'

What then has to happen inside our heads is something like this: as the other is talking, we try to detect the point or action being made. As soon as we have detected that (is it a question, a criticism, a joke?), we can go to work on a response, finding the words, the syntactic frame, and finally the phonological code – using brain imaging, we can see it goes as far as this long before we open our mouths.[18]

[17] On our surprising ability to correctly complete others' sentences see Lerner 1991, 1996, 2002.
[18] Bögels, Magyari, & Levinson 2015.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

Speaking is powered by air whistled through the speech organs – so unless we have enough residual air, we'll also need to take a breath. Here we use a special interrupt of the autonomic system which drives our breathing by measuring gas exchange, a voluntary interrupt that uses a nervous pathway unique to humans.[19] Now we are ready to speak, but our interlocutor may not be finished yet – to detect this we must continue to listen, parse the incoming stream, and check for signals of closure. Such signals may consist of a lengthened syllable, a falling intonation, or a multimodal signal like the hands returning from a gesture to resting location. As soon as the signal is detected, we can shoot – activating the vocal organs (we can see this happening using ultrasound).[20] Since any human minimal response time is around 200 ms, that accounts for the normal tiny gap between utterances.[21] The histogram in Figure 3.3 shows a typical European pattern of the response speed in conversation – what it shows is that there is a minority of overlaps (less than 5 per cent of the speech stream) mostly very short (modal length 100 ms), and the modal response (the most frequent type) is with a 200 ms gap. Given that turns are of no fixed length and typically novel in content, this is amazing precision, and a remarkable system is organizing all this complexity of cognition and physiology.

An interesting question is what's the hurry? What impels us to respond so quickly that we have to go through the effort of this double tasking, already getting ready to speak before the other has finished? We noted at the outset that the turn-taking system seems to be rule governed – each speaker gets a turn, usually a minimal unit like a clause, and the first person other than the current speaker to speak next, gets the next turn. But if no one else jumps in, the first speaker can continue – on average the first speaker waits about an additional 150 ms to give others a chance.[22] So if you want to speak, you have to get on with it! An additional pressure is that, given the normal pace of turn-taking, longer gaps come to have semiotic significance. We all

[19] McKay *et al.* 2003.

[20] Bögels & Levinson 2023.

[21] Levinson & Torreira 2015.

[22] Ten Bosch, Oostdijk, & de Ruiter 2004, Ten Bosch, Oostdijk, & Boves 2005. All of these numbers can clearly vary a bit across interlocutors and circumstances, but same-speaker continuations occur on average with a gap at least a third longer than gaps between different speakers.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

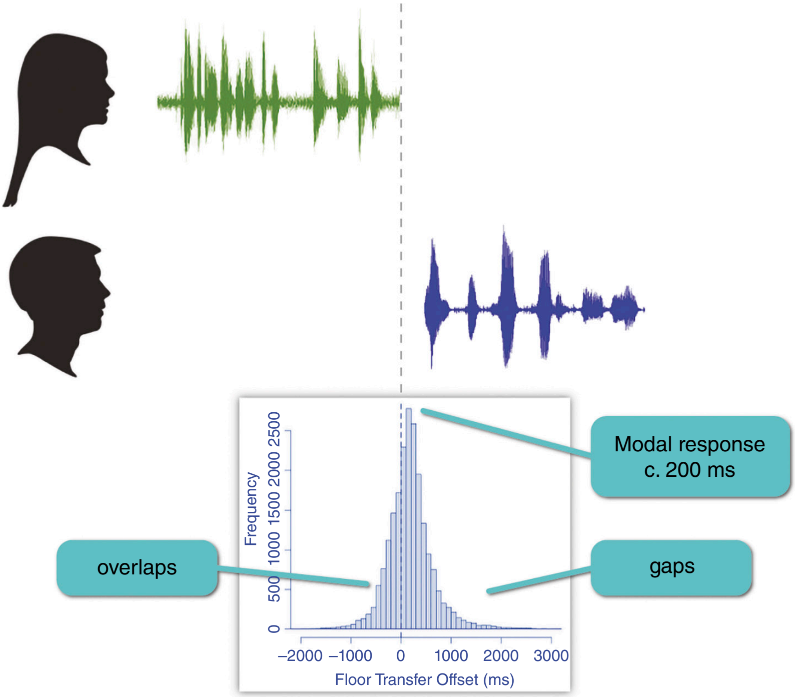

Figure 3.3 Typical response timings for an English conversation. The histogram shows overlaps to the left of the dotted line (that is, overlaps with the end of prior speaker's turn), and gaps to the right, in increments of 100 ms (after Levinson 2016). The modal or most common response is close to 200 ms after the end of the prior turn (0 on the x-axis).

know that silence after an accusation can betoken guilt.[23] But even slight delays after an invitation or a request will start to give rise to negative inferences, such as a reluctance to accept or carry out the request. To avoid such inferences, responses need to be produced on the dot.[24]

We can now appreciate that turn-taking is an elaborate system, involving the orchestration of multiple physiological and cognitive resources. But is this, like languages and other aspects of culture,

[23] UK law now qualifies the right to silence along these lines, allowing adverse inferences from silence (Criminal Justice and Public Order Act 1994).

[24] Using brain imaging, it's possible to show that as the length of the silence increases the strength of the inferences increase too (Bögels, Kendrick, & Levinson 2015).

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

highly variable across cultures? The answer is it varies a little in timing, but is remarkably stable across populations and cultures. We have examined the patterns in many languages,[25] and one way to see the similarity across cultures is to compare the graphs of response times (like the English histogram above in Figure 3.3), with overlaps on the left, gaps on the right, and the highest point the modal response time (see Figure 3.4). Using a controlled context, namely responses to yes/no questions in natural conversations, the findings show that the modal (most common) response was between 0 and 200 ms for all ten languages. The means show greater variation (indicating more spread, as can be seen visually in Figure 3.4) but they cluster around 200 ms. There have been many ethnographic reports of cultures where conversation sometimes appears very slow, amongst which are Nordic cultures. Travel writers warn Americans that their rapid-fire turn-taking will be perceived as arrogant: ‘while the Finnish are notorious for the slow pacing of their conversations and their extreme comfort with what would otherwise be considered painfully uncomfortable periods of silence, it is a trend present to a lesser extent across all of the Nordic countries.’[26] A Danish sample is shown in Figure 3.4, and responses here are indeed slower, averaging 470 ms, but this is only a quarter of a second slower than the cross-cultural average. It seems that we are deeply tuned to the ‘metabolism’ of our own culture’s conversational practices, so that we find small deviations very remarkable.

Other features of this study are interesting – across all ten languages, responses to questions take at least twice as long if they avoid answering or provide an answer that is not in the expected direction. This cross-cultural tendency to delay the unwelcome response means that delays will likely trigger negative inferences, which again will force the pace of conversation if those are unintended. In the study, nonverbal responses were counted in the timings. As a result, Japanese responses emerged as very fast, given the Japanese cultural favouring of early nodding and other rapid feedback (so-called *aizuchi*), reminding us that verbal interchange is naturally a multimodal practice. Indeed, the presence of visual signals by the questioner sped up answers in all languages.

[25] Stivers *et al*. 2009.

[26] Travel blog from https://virtualwayfarer.com/nordic-conversations-are-different/

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

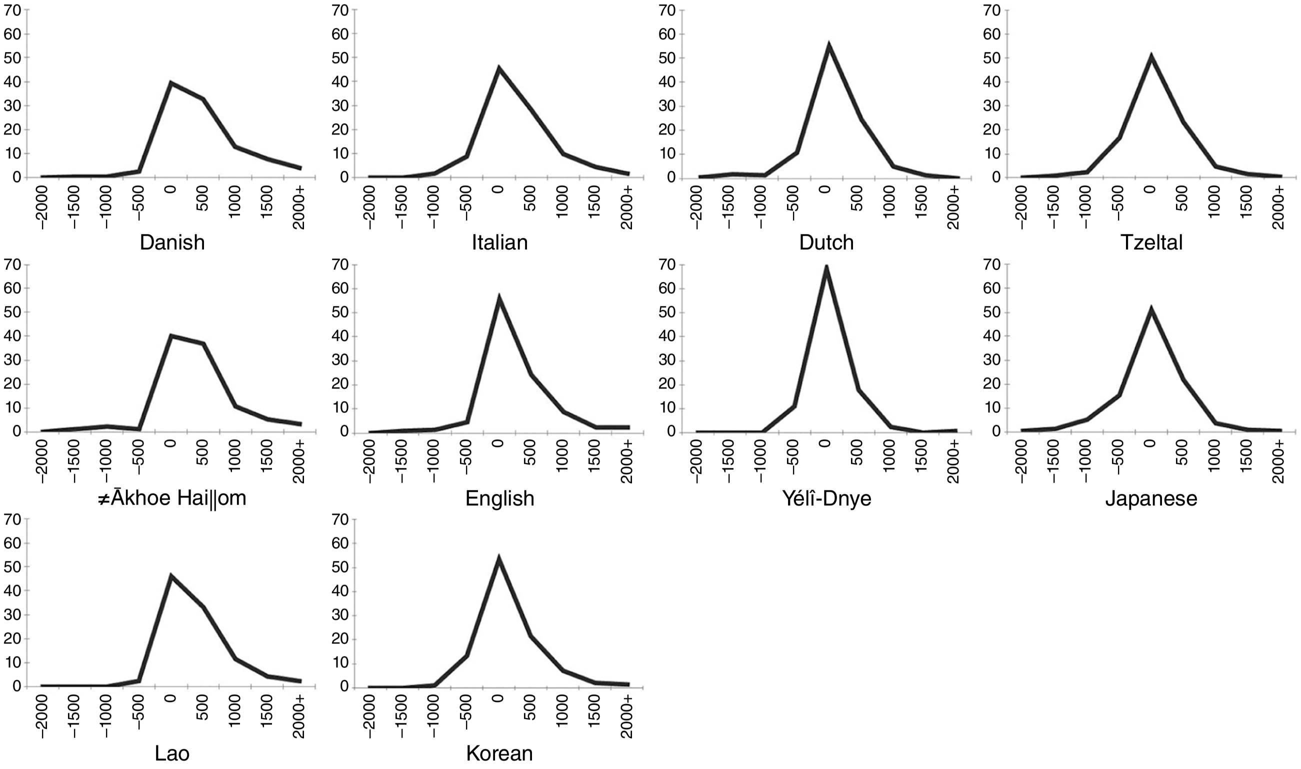

Figure 3.4 Conversational response times in ten languages. The languages are Lao, Korean, Danish, Italian, Dutch, Tzeltal, Akhoe Hai//om, English, Yélî Dnye, and Japanese; the timings are of responses to yes/no questions (from Stivers *et al*. 2009). The peak is the modal (or most common) response just after the end of the prior turn.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

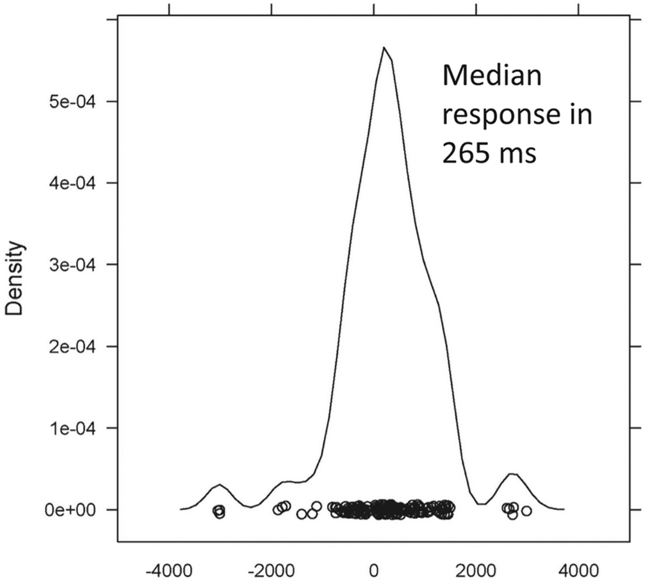

Figure 3.5 Response timings in the Dutch sign language NGT (after De Vos, Torreira, & Levinson 2015). The distribution is represented in a density plot (like a smoothed histogram) with timings very similar to spoken language.

Sign language is particularly interesting in this context. Sign languages, like spoken languages, vary considerably in lexicon and structure and they do so across spoken language boundaries (British Sign and American Sign, for example, have different cultural origins). Sign languages have the distinctive property that there's no shared channel in production and reception: making signs involves motoric action that is decoded visually. With spoken language, if two people speak at the same time, the two auditory signals may interfere, whereas in signed language in principle simultaneous broadcast and reception might be possible. That might predict that turn-taking in Sign might be very different. However, that is not the case. Figure 3.5 shows a Dutch Sign (NGT) sample of response times to questions, and the timings fall in the middle of the range from our spoken samples in Figure 3.4. Earlier it had been reported that signers do not adhere closely to the 'one at a time' rule of spoken conversation,[27] but this

[27] Coates & Sutton-Spence 2001.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

impression arises because the hands take time to get into the signing position and time to return to rest. Once those anticipatory movements (akin to taking a breath in speech) and retraction are taken into account, sign languages seem to obey exactly the same kind of timing as spoken languages.

3.3 Universals of Conversational Repair

One of the miracles of conversation is how well we seem to understand one another. Researchers in linguistics and artificial intelligence point out all the myriad ways in which utterances are ambiguous, vague, and their speech act or point unclear. Once our understandings diverge, they are likely to run far apart – you perhaps thought I was talking about London Ontario and I was talking about London UK. It is thus essential to clear up uncertainties as soon as they arise. Conversation analysts noted early on a systematic set of procedures that achieve this. First, there is a trigger – for example, a word or phrase that is not heard or understood. Immediately after the speaker of the troublesome words finishes his or her turn, the recipient can then issue a *repair initiator*. If possible this precisely locates the problem, for example, by giving its context, as in *John forgot to do what?*, or if the whole of the prior utterance was not heard or understood, then a general or open repair initiator like *huh?* or *what?* is used. In a study of twelve languages from all quarters of the world, it was found that repair of these kinds occurs regularly about every 80 seconds in conversation, which shows just how crucial a mechanism it is.[28]

Just as with turn-taking, it is the details that show that this is an interesting universal mechanism. As noted, there are two kinds of repair initiators – specific and general. The specific type is itself divided into a specific question (like *John forgot what?*) or a guessed solution, an offer as it were (like *John forgot the key?*). Note that the offer type merely requires assent, while the question type puts the original speaker of the troublesome item to more effort. Outside trouble-prone contexts (such as a noisy environment), specific repair initiators are considerably more frequent than open ones like *huh?*, which only occur about a third of the time. In general, it seems that, at least in the

[28] Dingemanse *et al*. 2015.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

dozen languages studied in the aforementioned study, participants in conversation use the most specific repair initiator they can, so requiring as little effort as possible from the original speaker of the troublesome phrase: guesses or offers only require assent, specific questions like *Who did?* only require a name or a phrase, while the general or open type will require a complete repeat or rephrase. What is interesting is that across all the languages repair requesters are altruistic, doing as much as they can to make the repair quick and easy for the original speaker of the trouble source. This is a sign of the collaborative coordination that is the hallmark of human communicative interaction.[29]

Another finding is that the default open repair initiator, the *huh?* word, itself has remarkable similarity in form across languages: it is always a monosyllable with an unrounded low front central vowel, delivered with questioning intonation, and if it has a consonant at all it is a glottal onset */h/* (as in 'hot') *or /ʔ /* (a glottal stop as in the middle of disapproving 'uh-uh!'). Although it thus has a claim to be a universal word, the details of the vowel quality, the intonation and the consonant (if any) are fitted to the phonology of the language.[30] This universality may either be due to natural features that lend themselves to the job (for example, a location of the tongue close to resting position, allowing a quick intervention) or just possibly because it belongs to a small group of surviving elements of some protolanguage. It is interesting that there are also close cross-cultural similarities of 'filler' words like *um*, *hm*, *err*, and the like.

3.4 Universals of Action Sequences: The Source of Recursion

We earlier introduced the notion that utterances perform actions, such as questioning, requesting or complaining, and noted the complexity of the inference from the words to the action. Nevertheless, human interaction is structured by sets of contingent actions – including question-answer, request-compliance, or greeting-greeting – technically called adjacency pairs. This is, if you like, the 'syntax' of

[29] In effect, they blame themselves for the hearing or understanding failure, in this way acting according to the principle of the 'virtual offence' discussed in Chapter 5.

[30] Dingemanse, Torreira, & Enfield 2013.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

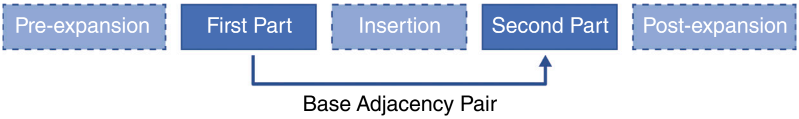

Figure 3.6 The possible expansions of the base adjacency pair (after Kendrick *et al.* 2020). The three slots for expanding a sequence of paired utterances like question-answer or request-compliance (described by Schegloff 2007).

human communication, the principles that make it a coherent activity. Curiously, given its importance, it is an understudied area, and once again we are indebted to the conversation analysts, and in particular Emanuel Schegloff, for most of what we know.[31] And as with turn-taking and repair, it is the intricacies of the details that make the phenomena distinctive of human communication across cultures.

The pair of actions, the initiating action and its contingent second, form the basis of at least one core part of this system. We can take the question-answer pair as canonical. What a question does is set up an expectation for an answer, the absence of which can be complained about and pursued; and where the addressee cannot provide an answer, an explanation is due. These pairs are thus normative rather than statistical in character.

It has been noted for English conversation that around this core or 'base' pair (such as question-answer) elaborate structure can be built, by expanding the sequence before, in the middle, or after the pair, as in Figure 3.6.

Pre-expansions are themselves typically formed of adjacency pairs, as in:

<6> (adapted from Clift 2016:77)
Tom: 'Emily!' <-Pre-First
Emily: 'What?' <- Pre-Second ('Go ahead')
Tom: 'Why have you got the playstation in your room?' <-Base-First
Emily: 'Because I am using it' <-Base-Second

Here 'Emily!' is a summons and 'What?' is both an answer to the summons, and recognizes that the summons is a preamble to a second action. Summonses can stand before any type of action, but

[31] The essential reference is Schegloff 2007. See also Levinson 2013b for review.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

other pre-expansions are more specific to what follows, as in the pre-expansion below (repeated from Example 2.3):

<7> (adapted from Schegloff 2007:30)

N: 'Whatcha doin''	<-Pre-invitation
C: 'Not much'	<-Go-ahead
N: 'Y'wanna drink?'	<-Invitation
C: 'Yeah'	<--Acceptance

Here the pre-action *Whatcha doin'* has some generality over proposals, invitations, requests, or the like. Their function is often to adumbrate what's coming up in a way that allows that to be truncated if it turns out to be inappropriate. They can be fully specific as in *Can I ask you something?*, which interestingly is not usually followed directly by a question but by another pre-action. In the example below there are thus two pre-sequences:

<8> (simplified from Schegloff 2007:47)

B: 'But- (1.0) wouldju do me a favour? Heheh'	<--Pre-Pre-First
J: 'e(hh) depends on the favor::, go ahead,'	<--Pre-Pre-Second (Go Ahead)
B: 'Didjer mom tell you I called the other day?'	<--Pre-request (Check on Preliminary)
J: 'No she didn't'	<--Pre-second
B: 'Well I called.'	
J: 'Uhuh'	
B: '.hhh 'n I was wondering if you'd let me borrow your gun'	<-- Request

This introduces us to the fact that the structure in Figure 3.6 has a potentially recursive quality, which becomes particularly evident in the insert sequences. A simple example, which we met earlier, in Chapter 2, of an inserted adjacency pair within an adjacency pair is as follows:

<9> (from Merritt 1976:333)

A: 'May I have a bottle of Mich?'	<-First-Part-Base
B: 'Are you twenty-one?'	<-First-Part-Insert
A: 'No'	<-Second-Part Insert
B: 'No'	<-Second-Part-Base

This nested or centre-embedded structure, we pointed out earlier, has been much lauded as a critical property of the human mind bestowed

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

on us by linguistic structure, as in *The boy John saw ran away* where one sentence (*John saw the boy*) is embedded in another (*The boy ran away*). In fact, this recursive structure is typically much more elaborate in conversational sequence structure than in language syntax. The following example shows three degrees of embedding, one pair inside another pair three times. This example already exceeds the depth of all syntactic embeddings in linguistic structure (at least as attested in spoken language),[32] and conversational examples can be found up to six centre-embeddings deep.[33] Notice that in this example repair is involved – the exchange takes place in a noisy environment, a sandwich bar.

<10> From Merritt 1976: 79
S: 'Next' ← Request to order
0 C: 'Roast beef on rye' ← Order
1 S: 'Mustard or mayonnaise?' ← Q$_1$
2 C: 'Excuse me?' ← Repair Initiator (RI$_1$)
3 S: 'What?' ← Repair on RI
3 C: 'Excuse me?'
2 'I didn't hear what you said' ← RI$_2$
1 S: 'Do you want mustard or mayonnaise?' ← Q$_1$= Repair
C: 'Mustard please.' ← A$_1$
0 S: ((provides)) ← Compliance with order

Finally, there are possible expansions after the base adjacency pair. These can be of a minimal kind, marking information receipt, as in:

<11> (after Schegloff 2007:121)
A: 'You wan' me bring you anything?' <-First-Part-Base
(0.4)
B: 'No: no: nothing.' <-Second-Part Base
A: 'Okay' <Minimal post-expansion

Or post-expansions can be another adjacency pair (again with all the possibility of insert expansions), as in:

[32] Karlsson 2007, using online data bases, found that three degrees of embedding were found in only thirteen sentences in the whole of Western literature, and two degrees of embedding in only three English spoken recordings to that date.

[33] Levinson 2013a.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

<12>
A: ‘Is Al here today?’ <--First-Part Base
B: ‘Yeah’ <-Second-Part Base
(2.0)
A: ‘He is?’ <-First-Part Post-expansion
B: ‘Well he was’ <-Second-Part Post-expansion

To summarize, around the simple structure of an adjacency pair, an elaborate sequence structure of actions can be built by repetitive or recursive application of the possible insertions.

The question now arises whether this is a culturally specific set of conversational routines, or something much more fundamental and cross-culturally applicable. This question was tested across samples of conversation from a dozen languages of eleven different language families from all corners of the globe.[34] It was important to see whether each of the six types of expansion sequence distinguished in the study were exhibited in every language based on a sample of conversation from each. The finding was that all the types were found in nearly all of the languages, and where there were missing cases they were likely due to the relatively small size of the corpora used.

Here then is an apparently universal syntax with full recursive power based not on the surface form of language but on the actions that the utterances perform. As mentioned, the depth of recursive centre-embeddings found in conversational sequences far exceeds anything found in the syntax of any language, the absence of which has previously been put down to memory limitations. That explanation now looks unlikely, but it does seem that the cooperative nature of conversational interaction makes it easier to keep track of where one is in the embedding structure. This suggests indeed that the ultimate source of recursive structures in our minds actually lies outside them, in the interactions we conduct routinely. Finally, it is worth pointing out that there are vanishingly few generalizations about human conduct that have the precision that we find in conversational structure. Generalizations from economics, social organization, demography, or criminology will be phrased in terms of probabilities, and we saw that linguistic generalizations are also normally of the kind ‘If a language has property X, then probably it has property Y’. Sequence structure

[34] Kendrick *et al*. 2020.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

does seem to be a remarkable organization that structures our communication regardless of language or culture, and through that our social lives.

3.5 Possible Universals of Action Types (Speech Acts)

A question that has fascinated both linguists and philosophers is this: What are all the kinds of things we can do with language? Are there intrinsic limits, and if so what are they? Is there a core set of universal functions? We have already mentioned (in Section 2.3) that there are some recurrent patterns across languages, but also clear cultural specializations. Here we delve a little further into these questions.

Starting from first principles, consider the philosophical notion of a proposition, a description of a (possible) state of affairs. Then we can imagine a range of attitudes one can have to a proposition – believing it is true (as in an assertion), checking whether it is true (as in a question), relying on someone to make it true (as in an order), urging people to make it true (as in an exhortation), wishing it were true (as in a prayer), advising against making it true (as in cautioning), and so forth. Now, the grammar books say that assertions, questions, and orders are universally realized in universal sentence types, namely declaratives, interrogatives and imperatives.[35] But this is not strictly true: firstly, there is no fixed relation between the form and the action – a question can request ('Can you pass the water?'), declare ignorance ('How would I know?'), express wonder ('How did they do it?'), and so forth. Secondly, about a third of languages (and the majority of sign languages) do not mark polar or yes-no questions in any systematic way except perhaps intonationally.[36] A few, such as Hopi perhaps, don't even have clear Wh-words, instead using indefinite statement forms, like saying in effect 'Someone came?' for 'Who came?'. Similarly, by no means do all languages have an imperative – Nunggubuyu (an Australian aboriginal language) for example, just uses the future tense for that function ('You will go' can be understood as 'Go!'; Modern Hebrew uses the same strategy, while Rapanui (spoken on Easter Island) uses the present tense).[37] Many others do not have a form restricted to the second-person addressee – they have a whole paradigm including first- and third-person imperatives,

[35] Aikhenvald 2016. [36] Dryer 2013.
[37] See Sadock & Zwicky 1985, König and Siemund 2010.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

the set not being really distinct from statements of moral obligation. Thirdly, even when there is a dedicated form, perhaps for questioning, it is very often not used – for example, in samples of languages that have a dedicated yes-no question format, 50 per cent of the time it is not used in yes-no questions, while the majority of Wh-questions do not do questioning at all![38]

Despite these caveats, there is nevertheless a tendency for languages to have at least the three distinct formats of sentence – declarative, imperative, and interrogative (the latter usually of two unrelated types, yes-no and Wh-format) – recognizably associated with asserting, ordering, and questioning. That the distinctions are statistical tendencies is, as we noted earlier, typical of what are called language universals. The reasons for the tendencies are, presumably, that there is a universal grammar of human motives – these actions are recurring uses to which language is put: we want to tell people about things, get them to do things, and find out what they know about things. The psychologist Michael Tomasello has noted that, whereas apes might well be said to gesture imperatively, so pointing to things in their presence, they lack the complex syntax that would articulate statements or questions, and indeed the motivations.[39] It is after all the coding of propositions that endows language with its informational efficacy. One possible solution to how humans evolved this propositional core is addressed in Chapter 4.

Earlier we noted that we can entertain propositions for different purposes, and beyond the golden three principal sentence formats, languages differ widely in the specialist formats they provide for action coding. Some languages have special forms for exclamations, curses, benedictions, wishes, or warnings. Cultural evolution devises special institutions geared to the local social systems, and these provide specialized uses of language, for instance, for legal hearings and judgments, educational purposes, and religious and magical rituals. It is these institutions that make it possible by using a special form of words to constitute a marriage, cause a divorce, make war, or condemn someone to death (see Chapter 5).

But what gets canonized in a special grammatical form is always a small subset of the functions that language performs. That is because

[38] Stivers, Enfield, & Levinson 2010. On the general form-function mapping problem see Levinson 2013b.

[39] Tomasello 2008.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004

beneath the machinery of grammar the interaction engine is always operative, searching for the point of utterances, in order to be able to respond to the underlying action. Although social motives may in principle be limitless in kind, they do of course have a probability distribution that will play a role in attributing actions to utterances in a local setting.

In this chapter we have reviewed some of the many ways in which human communicational interaction shows remarkable constancy across languages and cultures. The details, whether the precision of turn-taking, the repair system, or the rules of sequence structure, reveal an intricate system. It is this systematic base which makes it possible for an infant to work its way into its natal culture, for adults to maintain a meeting of minds during conversation, or indeed for a traveller to successfully mime what they want to eat in a foreign restaurant. The question that we now turn to is where this system comes from and how it arose in human evolution.

Downloaded from https://www.cambridge.org/core. IP address: 216.73.216.63, on 06 Aug 2025 at 13:09:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.004