7 Grammar for Intersubjectivity

Let us take stock. Our investigation of the conditions and consequences of language began in chapter 1 with the concept of intersubjectivity. Then, in part I, we covered elements of the infrastructure for human interaction, without which language could not operate as we know it. These elements are *status* (chapter 2), *participation* (chapter 3), and *enchrony* (chapter 4). Then in the first two chapters of part II, we linked these elements of the infrastructure for language use to two core structural-functional features of language: *reference* (chapter 5) and *semantics* (chapter 6). This sets the stage for a discussion of *grammar*. This is the system perhaps most widely regarded as "core" to language, at least by twentieth-century linguists and cognitive scientists, yet perhaps least obviously linked to—as either consequence of, or condition for—intersubjectivity in human social affairs.

We outline the status of grammar in relation to our two core questions in this book. First, in what way does primary intersubjectivity make language possible? Second, in what way does language then enhance human intersubjectivity? We shall see that grammar presupposes intersubjectivity and in turn offers highly developed, fine-grained, and powerful means for enhancing human intersubjectivity and transforming our social lives.

7.1 Grammar Contains Diverse Subsystems with Diverse Natures and Functions

"Grammar" refers to a heterogenous set of rules and signs that people use productively in organizing words, phrases, utterances, and discourses.[1] It is more of an ecosystem—a set of interrelated systems—than a single, coherent system. The lexicon, which we discussed in the last chapter, consists of *open-class* stand-alone signs such as the nouns and verbs of English. They

are called open-class because they belong to classes that are large and open to receiving new additions, for example when new words are coined, or borrowed from other languages. In many languages, there are thousands of nouns, with new ones being adopted every week, while the set of pronouns numbers only a handful, and may see changes such as additions, mergers, or subtractions only once in a century or two.

Grammar can be said to contain everything *other than* the open-class vocabulary items of a language.[2] This covers at least two main types of functional structure. First are the generative patterns and constructions embodied in phrase structure rules, idiomatic templates, and other types of ordering principles that enable people to generate novel utterances using their inventory of words. These principles allow us, for example, to distinguish between *Kim saw you* and *You saw Kim*, or to describe an animal as not just a horse but more specifically as *that horse*, *my horse*, or *the black horse*. Second is the system of closed-class grammatical markers, for example, pronoun and demonstrative paradigms, case-marking affixes, agreement markers on verbs, and other kinds of bound morphological marking (e.g., evidentiality and modality), to be discussed below.

In the next three sections, we explicate three central functions of grammar: grammar links reference and predication (clause structure); grammar links narrated events to speech events (displacement); and grammar links language to language (reflexivity).

7.2 Grammar Links Reference and Predication: Clause Structure and Beyond

A simple but powerful function of grammar is to bring together the two major functional elements of language introduced in the last two chapters. Grammar supplies rules for taking words and combining them into propositions. These propositions are the basis of *the clause*, which has subject-predicate or theme-rheme structure.[3] A clause refers to something and then characterizes it in some way. Take the clause *Kim is asleep*. First, it points to *Kim*, setting this referent up as a theme or subject of joint attention (see chapter 5). Next, it introduces some new piece of information, something *about* Kim. This is the predicate. To understand this sentence, you focus your attention on Kim and then you learn that she is asleep.

Every language provides resources for organizing words into clauses in this way (though the specific rules for doing this differ across languages, sometimes radically; Sapir 1921; Bloomfield 1933; Greenberg [1966] 2005; Croft 2001; Evans and Levinson 2009; Dixon 2010; *inter alia*). In terms of our interests in this book, the simple function of linking the act of referring to the act of predicating provides an essential and powerful building block for the construction of intersubjectivity. Using the productive mechanisms of the clause, a person can not only draw someone's attention to something, they can convey new information about that thing. This increases shared information, which is a prerequisite for intersubjectivity.

But shared information itself is not sufficient. As we explained in chapter 1, for information to become intersubjective, the information must be shared *and* it must go on the shared record *that* the information has been shared (Lewis 1969; Clark et al. 1983; Clark 1996; Stalnaker 2002). We might both know that Kim is asleep, but it is only when the sharedness of the information is in the open—for example, when I say to you that she's asleep and you nod in understanding—that both of us become mutually accountable for our subsequent actions in relation to the proposition that Kim is asleep.[4] For example, when the fact is in the common ground, we don't have to explain why we are whispering or tiptoeing around the house. Conversely, we won't be surprised to be sanctioned for playing loud music.

So, grammar carries a huge functional load in building intersubjectivity by increasing common ground through productive generative mechanisms for linking subjects to predicates in novel propositions.

The power of syntactic structure in the clause is ramped up enormously when we add another level of capacity for specifying further details in our descriptions of events and situations. The subject of a clause isn't just a noun but a noun *phrase*. A phrase can be simple or complex. So instead of *Kim is asleep* we could say *My niece is asleep* or *My poor tired niece is asleep*. Because grammar is semantic in nature, the expressive alternatives provided by grammar have the same implications for intersubjectivity as the kinds of lexical contrasts we introduced in the last chapter. Grammar radically amplifies those possibilities for selecting among alternative framings of the situations we want to describe.

In the last chapter, we discussed the implications of word selection for accountability, for example of the participants in a situation being described.

Recall the children disputing what happened just before a toy block structure tumbled down: A: *She poked it!* B: *I tapped it!* The generative nature of grammatical constructions means that a speaker's possibilities for alternative framings of a situation are not limited to the existing inventory of relevant words the language happens to furnish (e.g., *poke* versus *tap*), but are literally infinite.[5] This is important, in part, because accountability attaches not to brute facts but to formulations of those facts (consider, "You didn't talk to that man" versus "You didn't say hello to your friend").

Grammars also provide the means for extending basic clause structure with increasingly specific information in associated structures such as prepositional phrases or their equivalents. For example, we could specify the location of the event: *Kim is asleep in the next room*. The discipline of linguistics is heavily concerned with describing and understanding grammatical mechanisms for constructing clauses and their constituent phrases. Much of that research focuses on the organization of referential information in propositions that describe, or make claims about, events and situations beyond the specific context in which the proposition is being made.

We will not review the formal complexities of grammar's potential for infinite expression here (see, e.g., Van Valin and LaPolla 1997; Dixon 2010; *inter alia*). Our interest is to explicate the sense in which that potential plays an important role in enhancing human intersubjectivity.

7.3 Grammar Links En to Es: Displacement and Beyond

The capacity of grammar to link reference and predication is just the first of three important intersubjectivity-enhancing aspects of grammar. The second component is the role that grammar plays in linking the *narrated event* to the *speech event*.

Following Jakobson (1971), we adopt the following terminology. Given a linguistic proposition, the *narrated event* (En) is whatever event or situation the proposition refers to or is about. In this technical terminology, "event" covers any state of affairs, not just dynamic events. The *speech event* (Es) is the situation of speaking itself. If I say *Kim was asleep*, the past-tense form of the verb *was* is an explicit signal that the time of the narrated event (Kim being asleep) is *before* the time of the speech event (now). We use the term *origo* (Karl Bühler's term; Bühler [1934]1982) to refer to a point in

time-space-personnel (*here-now-me*), such that the Es origo (time of speaking, place of speaking, participants in speech act) may be different from the En origo.

The very possibility of communicating about events and situations beyond the here and now is one of the uniquely defining features of human language. Hockett referred to this as *displacement* in his 1960 article on the design features of language. Animal calls lack this capacity for displacement.[6] For example, a vervet monkey can make a call that means something like "Threat from the sky!" (the so-called eagle call; see Cheney and Seyfarth 1992). The call signals that there is a threat *here*, *for me*, and *now*: there is no distinction between the narrated event and the speech event. The vervet cannot communicate to its associates that, say, there was a threat from the sky earlier that day or that the threat is somewhere else than here (or, further, that there was *no* threat from the sky yesterday or that the caller thinks there is a threat but isn't sure). In systems of animal communication, the event of communicating (Es) has the same coordinates in terms of space-time-personnel as En, the event being communicated about. In other words, nondisplacing systems of animal communication do not have means for distinguishing Es from En.

By contrast, human language allows us to make propositions about events and situations that have different space-time-personnel coordinates from the speech event. This feature of language enables us to inform people about things that happened beyond their experience but which it might be useful to know. Suppose someone tells you *There is a wasp nest in that hedge*. Those words will prompt you to take care to walk around the hedge, not barge through it, and thereby avoid harm. Or suppose you collect some *Colocasia gigantea*, a wild forest tuber, not knowing exactly what it is, and you are told *My uncle nearly died from eating this*. Those words will relieve you of learning that lesson the hard way. This powerful capacity for informing others is often foregrounded when people emphasize the special utility of language. People can tell you things so you don't have to learn them from experience. This is central to the cumulative nature of human culture.

But saying things to people does more than simply update them on the informational content of propositions. It is part of the relationally strategic building of common ground, which is a form of investment in a social-affiliational economy. Gossip and related forms of talk are oriented to the

sharing of stances and the management of reputation. We not only report things that might have happened, we also use lexical and grammatical resources to frame and portray those events as good or bad, right or wrong. Telling people what others have said or done is an opportunity to share social values, to praise or to blame others, and to manipulate social affiliations by declaring and securing allegiances (Gluckman 1963; Haviland 1977; Goodwin 1991; Dunbar 1993; Enfield 2013, 2022). Similarly, the combination of infinite expressivity and displacement in language is the foundation of narrative more broadly (most gossip is already narrative, of course), which in turn is at the heart of social sense-making. Grammar supplies many of the tools we use for these functions.

Several of the subsystems of grammatical marking found in the world's languages are dedicated to marking relations between Es and En.[7] For example:

Tense: *Kim is asleep* versus *Kim was asleep*

Aspect:[8] *Kim fell asleep* versus *Kim has fallen asleep*

Modality:[9] *Kim is asleep* versus *Kim might be asleep* or *Kim must be asleep*

Evidentiality: *Kim is asleep* versus *Kim is apparently asleep*

Another such grammatical category is egophoricity (San Roque, Floyd, and Norcliffe 2018). This is a grammatical system that marks canonical versus noncanonical alignment between the person of a clausal subject (first versus second) and the speech act function of the clause in its context (assertion versus question). Consider table 7.1.

In English, the person of the subject and the speech act function of the utterance are independently expressed, but they are not independent in practice. Two of the four possibilities in the table are pragmatically marked, meaning that they are far less frequent in usage, and when used they can

Table 7.1
Logical combinations of subject (first versus second person) and speech act functions (assertion versus question) in a sample of clauses with the predicate "hungry" in English (San Roque et al. 2018, 66)

	First-person subject	Second-person subject
Assertion	I am hungry.	You are hungry.
Question	Am I hungry?	Are you hungry?

sound unusual and thus invite special inferences. These are the shaded cells, in the bottom left and top right of table 7.1. Because a person has privileged access to knowledge and experience about themselves, this leads to two natural correlations: (1) Propositions with first-person subjects will usually occur in the speech act function of assertions, and (2) propositions with second-person subjects will usually occur in the speech act function of questions. In most languages, evidence for these correlations is found solely in frequencies: instances of the shaded cells will be far less frequent than the unshaded cells. See table 7.2.

As table 7.2 shows, of a sample of "hungry" clauses in English with first-person subject, 100 percent of these are assertions and none are questions. And of those with second-person subject, questions (i.e., "Are you hungry?") account for 95 percent of cases. Put another way (and recall we are focusing only on clauses with first- or second-person subjects), if an assertion is made that someone is hungry, only seven percent of these will have a second-person subject, while 93 percent will be first person, and if a question is asked as to whether someone is hungry, then 100 percent of these will have a second-person subject.

This pattern is surely not surprising, but we note that in English it is evidenced only in these frequency measures and is not marked in the grammar. But in some languages, there is overt grammatical marking of this asymmetry (San Roque et al. 2018). With a system of grammatical marking of *egophoricity*, speakers will use one type of marking to signal that the correlation of person and speech act function is the canonical one (so, marker 1 is used with both first-person-subject statements and second-person-subject questions) while the other type of marking is used in the noncanonical, unexpected situation (i.e., marker 2 is used with first-person-subject *questions* and second-person-subject *assertions*).

Table 7.2
Distribution of subjects (first versus second person) and speech act functions (assertion versus question) in a sample of clauses with the predicate "hungry" in English fiction texts (San Roque et al. 2018, 66).

	First-person subject	Second-person subject
Assertion	171	12
Question	0	249

Systems such as these add information to the basic subject-predicate content of clauses and sentences, giving further weight to the intersubjectivity-enhancing effects of language. By adding more specific information, the accountability implied by the speech act is fleshed out. In languages with systems of evidential marking, speakers are accountable for marking the evidentiary basis for what they're saying: if they haven't seen the event for themselves, they should make this explicit using a "hearsay" or "inferential" marker or something similar. In languages with multiple distinctions in tense (e.g., recent past versus past), speakers are accountable for marking this distinction accurately. At the same time, such marking will also allow people to highlight or downplay accountability for narrated-event participants.

To summarize, grammar gives us the capacity to productively build novel phrases, clauses, and sentences, and to do so in ways that (1) orient to the distinction between Es and En and (2) finesse that distinction in subtle ways. These functions of language massively amplify the language's capacity to enhance human intersubjectivity. Grammatical systems such as tense-marking, evidentiality-marking, egophoricity-marking, and modality-marking are dedicated to specifying certain kinds of links between a speech event and a narrated event. These semantic specifications are conveyed by grammatical elements in closed-class systems. They convey information that is not limited to En, the narrated event.[10] They provide information about Es, the speech event, and therefore they orient to language itself. This takes us to a third relation between grammar and intersubjectivity, the language-unique function of *reflexivity* (Hockett 1960).

7.4 Grammar Links Language to Language: Reflexivity and Beyond

So far, we have talked about structure at the level of clauses and sentences. But a focus on this level alone—endemic in formalist linguistics—would be insufficient. This is because people don't talk in isolated sentences (Foley and Van Valin 1984). Instead, in the enchronic context of language usage, language comes in the form of grammar-hewn chunks in sequences of social interaction. If you utter a clause or a sentence, it will never simply point to and characterize some state of affairs. The utterance will always be occasioned by something in a social interaction, and it will always occasion something next. The speech event is not an isolated point. It is a link in a temporal chain. Every contribution is at least two things beyond its informational

content: (1) Looking back, it is a response to, and interpretation of, what just came before, and (2), looking forward, it sets the conditions of relevance for, and interpretation of, what comes next. This is most obviously true when each move is produced by two different people, as in a back-and-forth conversational structure like a question-answer sequence. But it is also true when the series of moves is produced by a single person, as when a speaker uses language to do something that takes more than one turn, such as giving a set of instructions or narrating a series of events.

So, our first point is that in discourse, people don't just understand each contribution but understand that it orients to its "cotext" (i.e., linguistic environment, as opposed to "context" more generally). Language is always linked to language.[11]

The idea that language can be about language is made most explicit in the phenomenon of quoted speech. Suppose it turns out that Kim wasn't asleep after all and when I find this out, I sanction you for misleading me. I say, *You said that Kim was asleep*. My utterance refers to your previous utterance. This kind of function is not available to even the most sophisticated animal communication systems. No vervet monkey can hold another vervet monkey to account for, say, giving a wrong signal or not giving it quickly enough. No bee can communicate to another that their waggle dance was unclear or false. But with language, humans do this sort of thing all the time.

To talk about the things that others say, we need a dedicated set of syntactic strategies for *embedding* others' utterances into our own utterances. In the example just given, the verb *say* takes the complementizer *that*, which in turn takes as its complement a finite clause whose propositional content is that of your original utterance. A good deal of the machinery of complementation strategies in the grammars of the world's languages is dedicated to supplying ways of saying things about things people say (Vološinov [1929] 1973; Goffman 1974; Lucy 1993; Dixon and Aikhenvald 2006; Rumsey 2021; Zuckerman 2021).

Beyond the grammatical tools for quoting others' speech, there is a much more expansive and open-ended set of mechanisms for linking language to language. These are the mechanisms for building narrative.

Narrative is key to human collective sense-making and it would be impossible without the grammatical mechanisms that languages provide for achieving cohesion over larger texts (Bruner 1991; Ochs and Capps 2009; Dancygier 2012). As we have noted, people typically produce series of utterances rather

than isolated signals. We are about to see some of the ways in which the resources of grammar provide explicit ways to connect moves in linguistic sequences.

First, we note that the primary, highest-level kind of glue that joins these utterances together is not marked in grammar but is imposed by the human capacity for assuming associations among adjacent signs, and for using general processes of inference to arrive at the most likely interpretations. Suppose we simply juxtapose two propositions (uttered at midday):

Kim is asleep. She worked the night shift last night.

Human inferential processes are such that we will not take these propositions to be unconnected. We will infer, by abduction, what the relevance of the second part is to the first part, and thus what the speaker must have wanted to say by adding it. Most likely, we would take the statement that Kim worked the night shift last night to be an explanation of, or justification for, *why* she is sleeping at midday. Note that this kind of inference draws heavily on broader, intersubjectively-shared common ground (e.g., knowledge about shift work, about daily sleep needs and normative sleep patterns, etc.) to figure out what kinds of connections between propositions are most likely to be intended and why.[12]

Our always-on abductive inference engines—using a sort of association heuristic—will go a long way to providing a general degree of coherence among the many propositions that together, and in a given order, make up a discourse. But the resources of grammar can add a great deal of further detail and refinement, giving speakers the means to constrain hearers' interpretations. In the example we just gave, we included one such grammatical resource that lends structural cohesion between the two clauses. That is the resource of pronouns. Look again at the example:

Kim is asleep. She worked the night shift last night.

In the second clause, the word *she* is understood to refer to Kim, though in itself it only specifies a third-person feminine singular. It picks up its reference precisely by its placement in a discourse sequence (Halliday and Hasan 1976). This word is one of a closed class of grammatical words in the set of personal pronouns in English.

While this use of pronouns is a universal strategy for textual cohesion, the semantic distinctions made in pronoun systems can vary widely, depending on the (often historical) cultural concerns of the communities in

which the language has evolved. The English pronoun system is relatively simple among languages in the number and type of semantic distinctions made (singular/plural; first, second, or third person; and masculine/feminine in third person singular). Different languages will add different additional information.

Some languages add distinctions relating to kinship. For example, in some Australian Aboriginal languages there are alternative dual pronouns, one pronoun marking a pair of people who are in a generationally harmonic relationship (e.g., in the same generation such as two brothers, or two generations apart such as a grandmother and granddaughter), the other marking that they are in a disharmonic relationship (e.g., adjacent generation such as a father and son) (see, inter alia, Sapir 1913; Hale 1966; Merlan and Heath 1982; McGregor 1996; Evans 2003a, 2003b). This is information about participants in En, the narrated event, but the pronouns also function to link utterances that stand in a certain relationship to each other within the enchronic course of Es (Silverstein 1987; see also Nichols 1984).

Another kind of meaning distinction that language might mark in pronouns relates to politeness. In many European languages, there is a choice between distant/formal and familiar/informal pronouns in second person reference (e.g., *tu/vous* in French, *du/Sie* in German; see Brown and Gilman 1960). This kind of distinction is oriented more to Es, the speech event, signaling something about the relationship between the speaker and addressee, independent of what is being said about the narrated event.

There are many other kinds of grammatical device for explicitly marking cohesion across clauses. One way is to explicitly mark the relationship of conjunction between clauses or sentences. In the following examples from English, the underlined words mark distinct meaning relations among propositions: *She got on her horse and rode home; He was so upset that he had to be alone; His crop failed because the soil was poor; She forgot the words so she hummed the tune.* In another kind of case, a relationship between clauses is marked by a combination of a word and an affix: *He was eating lunch when they arrived.* All languages have these kinds of interclausal linkers. Some have patterns of marking that operate at greater distance across many propositions in a narrative sequence. Many languages of Papua New Guinea feature a grammatical pattern known as clause chaining.

In a typical Papuan clause-chaining construction, a series of events is predicated in a number of verbs used in sequence. All of the verbs are

marked with a grammatical suffix, which marks that the verb is *medial* in the chain (i.e., non-final). Medial verb marking can specify agreement-related values of person and number of the narrated-event participants but does not specify tense, aspect, or mood. For their tense/aspect/mood values, these medial verbs depend on the inflectional marking of the final verb in the chain. Here is an example from Ku Waru:[13]

(1) (Ku Waru) (Rumsey, Reed, and Merlan 2020, 4)

Olyo med maket-ma-nga pu-p

we down.there market-plural-genitive go-medial

Kalyip baim te-p

peanut buy do-medial

No-b pilawa lyi-p no-

eat-medial flour.balls get-medial eat-medial

Pu-mulayl

go-future/first-person-plural/definite

"We'll go down to the markets and buy peanuts and eat them, and get some flour balls and eat them and then we'll go."

All languages have ways to indicate such interclausal relations, lending structure to narratives and increasing informational specificity. This increases common ground among interactants and also adds specific content to the accountability being intersubjectively established, both among speech event participants and among narrated-event participants.

So far, we have concentrated on grammatical means for achieving cohesion in series of propositions in narratives, where each move represents one of a number of happenings in the narrated event. This is textual cohesion. It is needed if we are to make sense of monologues. But in the enchronic context that dominates our use of language, cohesion is needed not just between propositions in a narrative, but between propositions in back-and-forth sequences of interaction, that is, in the speech event involving two or more participants.

The most fundamental expression of cohesion between propositions produced by different participants in a speech event comes under the rubric of *sequence organization* (Schegloff 1968, 2007a; see chapter 4, above). The assumption of relevance allows us to infer meaningful relationships between propositions in a sequence produced by a single person. Similarly, in dialogue,

the assumption of relevance allows us to infer the probable meaning relationship between adjacent propositions. Consider the following:

1. Person A: *Is it cold there?*
2. Person B: *It's freezing.*

Again, these aren't disconnected propositions. We understand the second utterance, from Person B, to be not just an assertion about some narrated event but an *answer* to the question posed by Person A. Now notice that in the enchronic context of dialogue, the connection of relevance between the two utterances isn't only a cue for our interpretation of what the second proposition means in relation to the first. It also generates real-time rights and duties for the speech event participants. Once a person addresses a question to somebody, as in the example just given, then the addressee of that question is subject to an interpersonal imperative that Schegloff (1968) termed *conditional relevance* (see chapter 4). Because of this imperative, Person B in our example is momentarily obligated to address the question. They should answer it if they can. If they can't answer it then they should at least provide a response (even if that is "I don't know"). If they do not—for example by simply not responding, or by saying something entirely unrelated to the question—then the answer that line 1 had projected will be "officially absent," which in turn will mean that Person B may be held accountable for not having fulfilled the duties implied by their status as "addressee of a question" (see chapters 2 and 3). The complexities of conditional relevance go beyond our current scope. We raise the issue here to point out that conditional relevance is often brought about, regulated, or refined by the grammatical resources of the language being used.

7.5 Grammar Matters for the Human Cognitive Capacity for Intersubjectivity

Research on language and social intelligence over the last few decades has discovered multiple kinds of relationship between language and the psychological basis for intersubjectivity, often referred to as theory of mind. In their 2014 review of this field of research, de Villiers and de Villiers (2014, 313) define theory of mind as the "ability to understand that other people have minds, and those minds contain beliefs, knowledge, desires, and

emotions that may be different from [one's own]." This is obviously an important part of the human infrastructure for intersubjectivity. Research on the link between language and theory of mind has found evidence for three main types of connection, focusing on the development of theory of mind in children (much of which takes place in the first four years of life).

A first link between language and theory of mind concerns the semantic content of frequently used words and expressions about other people's inner states or first-person experiences. We use words with children early on to discuss what they want and think, how they feel, and so on. Such talk draws our attention to others' inner states and also encourages us to build hypotheses—largely based on our own experience—as to what those states are and what they are like, and ultimately to build folk-psychological models of the inner life of others (see Nelson 2005; Hutto 2008).

A second link is in the information about other minds that typical conversational language use provides, assuming the kinds of relevance-based, associational inferences described above. De Villiers and de Villiers (2014, 314) cite the following exchange:

1. Person A: *I am going to have chocolate spread on my toast.*
2. Person B: *That's Marmite!*

Marmite is a savory food spread made from yeast extract. It is a salty, yeasty concoction that might, to the untrained eye, look like chocolate spread. The exchange provides evidence that Person A's utterance in line 1 is not merely a proposition but is treated as evidence of the person's belief, namely, that the spread is made of chocolate. Here, Person B discerns that it is a false belief, and they act accordingly, informing Person A of the truth. (This action itself reveals an attention and orientation to others' inner desires; namely, the assumption that people would want their false beliefs to be corrected.) This is grounded in the fundamental idea that speech acts have conditions of appropriateness, many of which have to do with the speaker's state of mind (Austin 1962; Searle 1969). Thus, when a person makes a statement, asserting a proposition, this conveys—by presupposition—that they believe the statement to be true. (Of course, they may be lying, but our vulnerability to being fooled by lies is grounded precisely in our usual assumption that people's statements are grounded in their sincere beliefs.) Listeners learn to adopt the working presuppositions as a matter of principle in the cooperative context of conversation (Grice 1975; see also Harris 2005).

A third link between language and theory of mind brings us back to this chapter's focus on grammar. This link is made by the grammatical resources that a language provides for expressions that explicitly represent our own and others' states of mind. These features of grammar are akin to, and often identical to, the structures used for quoting the speech of others, mentioned above. An example is *I forgot that Kim was asleep*, where the proposition that Kim was asleep is embedded in the clause as a complement of the verb meaning "to forget."

Languages provide means for *embedding* one clause in another.[14] This function is especially important for intersubjectivity because it is our tool for thematizing and characterizing people's beliefs and desires, and thereby for publicly coordinating around the things that people think, want, and feel. These are the grammatical resources for *complementation* (Dixon 2010, vol. 2, 370ff).

Verbs of cognition such as *want* and *know* may occur in simple clauses, each taking a noun or noun phrase as their complement:

I want water.

You know the answer.

But they also often take a complement that is a clause in itself:

I want to leave now.

You know that Kim is asleep.

And such embeddings can be further embedded:

You know that I want to leave now.

The ability to control this kind of complex embedding of others' inner states and perspectives presupposes the ability to conceptually isolate and manipulate others' inner states and perspectives. In the example just given, the speaker linguistically represents not only what the main subject (you) knows (a first-order representation of another person's inner state) but also the subject's representation of another person's inner state (a second-order representation: you know what I want). But crucially for the argument we are making in this book, not only do these kinds of linguistic structures *presuppose* certain capacities for intersubjectivity, they also help to *enable* the underlying human capacities for intersubjectivity. For young children who are developing those capacities, "mastery of the appropriate linguistic structures gives the child a new ability to reason about the contents of

others' minds" (de Villiers and de Villiers 2014, 314). But taking this out of the domain of individual psychology and into the public realm of enchronic interaction, we can say that such structures massively amplify the capacity of people to build intersubjectively shared knowledge and to hold one another accountable. This aligns with our argument that, firstly, primary intersubjectivity makes language about other minds possible, and secondly, in turn, that as language about other minds develops, this further enhances human intersubjectivity, taking it to a new level.

While the complex technical details of syntactic embedding are beyond our scope here (see Foley and Van Valin 1984; Givón 1984, 1991; Van Valin and La Polla 1997; Dixon 2010) it is nevertheless worth pausing briefly to consider the ways in which the subtleties of meanings in different mental predicates and speech act verbs are associated with differences in grammatical structure and behavior. Consider the English verbs *want* and *wish*. Their meanings are similar enough that they are often interchangeable without significant consequence:

John wants to excel.

John wishes to excel.

Other contexts, however, reveal an important difference:

I wish that I had met my grandmother.

**I want that I had met my grandmother.*

As Dixon (2014, 32) writes, "The verb *want* is only used of something which is achievable, and it is restricted to a complement clause introduced by *to*, as in *I want to meet Barack Obama*. There is a more wistful sense associated with *wish*, relating to something that is not possible, and it will then take a complement clause introduced by *that*" (see also Dixon 2005). In addition, while *wish* subsumes "want," it also contains additional semantic components relating to *knowing* (Wierzbicka 1988, 132ff), and so it is compatible with English *that*-complements in a way that *want* isn't.

We can gain further insight into this point by looking at inner-state verbs that can occur with both types of complement structure. Consider the verb *remember*:

I remembered to lock the door.

I remembered that I had locked the door.

(Cf. *I wanted to lock the door* vs. **I wanted that I had locked the door*.)

The event of remembering is relevant both to achievable subsequent action (and hence compatible with a *to*-marked complement) and to the situation of a proposition being the object of my thought or knowledge (marked with *that*).[15] Other words, like *hope*, introduce further nuance: although one can say *John hopes to excel*, one cannot say **I hope that I could have met my grandmother*. This is because *hope* cannot take complements that express counterfactual propositions.

The point is this. Meaning differences between the inner-state predicates that create, maintain, and regulate our shared conception of mental life are not solely in the semantics of the relevant words but also in the grammatical frames within which the words may occur. Such inner-state predicates no more point to preexisting mental states than speech act verbs point to preexisting types of act (see Enfield and Sidnell 2017a and 2017b). Rather, such words and grammatical constructions analyze, construe, and construct mental states and, in turn, structure the ways in which we talk and think about how the mind works.[16]

7.6 Intersubjectivity Is Not for Minds but for Agents

The significant scholarly literature about language and theory of mind has focused on questions of individual psychology, albeit psychology in the social domain. Here is de Villiers and de Villiers's (2014, 325) summary conclusion of their review of the role of language in theory of mind development:

> A breakthrough arises with the child's ability to represent, via complex grammatical language, the contrast between the content of his own beliefs and knowledge and those of others and to use this contrast to reason about others' behavior. Being able to talk about minds leads to a richer theory, one that continues to help make sense of social situations.

This summary is in line with our enhanced intersubjectivity argument, but it puts more emphasis on the development of individual mental capacities. It is about the individual's ability to "represent," "reason about," "talk about," and "make sense of" others' beliefs and desires. But there is something important beyond these individual capacities and we want to draw attention to it. That thing is action. Having linguistic means for representing the thoughts, desires, and words of others is not only a way to understand other people, it is a way to make claims about their social accountability, and in turn to position them, and oneself, in a set of fundamentally social relations. And then one acts accordingly.

So, if someone says *You know that Kim is asleep*, they are not just stating what you know, they are drawing your attention to something that is relevant for evaluating your behavior. For example, if you are not taking care to be quiet when moving about the house, the statement *You know that Kim is asleep* might be an admonition. This might contribute to cementing the speaker's publicly registered stance that you are an inconsiderate person, which in turn may affect your relationship with that person and, more broadly, your reputation.

And if someone says *She thought it was chocolate spread*, they are not just stating what she thought. They might be explaining or justifying her behavior to somebody else. They might be using this as evidence for her outsider status (if, say, she is an American lodging with a family in England). They might be using this to make a case that she is careless or dim-witted. These aren't just matters of our individual capacity to mentalize about others' minds. Social intelligence is more than this. It includes the motivations and capacities for creating and manipulating statuses and relationships, and for enabling social action.

Through processes of the kinds we have reviewed, a child who is acquiring language and other aspects of social competence will draw on language, and to a large degree on grammar, to achieve the enhanced intersubjectivity that characterizes human social life. In so doing, as Katherine Nelson (2005, 28) puts it, the child is thus able to "enter the community of minds." We agree, but we would add that it is not enough to be a mind among minds. The child who has developed linguistically enhanced intersubjectivity is enmeshed in frameworks not only for mutual knowledge but for mutual values, ethics, and actions. The community they enter is a community of accountable agents.

7.7 Where Grammar Comes From

In this chapter we have been concerned with the role that grammar plays in enhancing human intersubjectivity. By focusing on features of grammar such as pronominal systems, markers of interclausal relations, speech act markings, and complementation strategies, among much else, we see how grammar achieves the following things, above and beyond what words alone can achieve:

- linking words to other elements in a clause
- linking speech events to narrated events
- linking clauses to clauses in larger sequences; forming texts such as narratives
- linking utterances to utterances (e.g., in reported speech)
- representing and expressing people's inner states, such as beliefs and desires
- regulating statuses of speech event participants

Our observations about the intersubjectivity-enhancing properties of grammar in this chapter have been limited to structures and processes in two distinct temporal-causal frames. One of these is the synchronic frame, to use de Saussure's term. Synchronic descriptions of the elements, structures, and rules of grammar as a system that hangs together are entirely abstract compared to the temporal dimensions of language use. We have described grammar in synchronic terms, but at the same time, for the purposes of understanding grammar's role in intersubjectivity, and in particular its relation to matters of the speech event, we have also found it necessary to invoke the enchronic frame, the frame that foregrounds the role of language in constructing moves in sequences of social interaction (see chapter 5). The enchronic frame is, we argue, at the confluence of linguistic and social processes occurring in all other causal-temporal frames. The enchronic frame should be privileged in any analysis of language and intersubjectivity because it constitutes the bottleneck through which all language (and all signs) must pass.

Let us put this into the broader context of research that seeks to explain why grammatical systems are the way they are. Attempts to explain observed language structures will begin with synchronic description, as we have just defined it. While we include the synchronic frame in a set of diverse causal-temporal frames for language (Enfield 2014, in press), it stands apart from other frames because it is the only frame that is in fact neither causal nor temporal. It is purely relational.

Once you know the set of relational patterns and rules that hold among linguistic signs in a language system, the question then arises: How to explain them? Different kinds of answer will appeal to different causal-temporal frames. One kind of answer appeals to the diachronic frame. This

refers to the population-level historical processes by which grammars evolve, over decades and centuries. As an example, consider the robust finding from descriptive work on the world's languages that there is a strong correlation between (1) the relative order of verb and object in clauses and (2) the relative order of adposition and noun in adnominal phrases. English conforms to this correlation, with verb-object order in clauses ($Kim_{subject}$ [saw_{verb} you_{object}]) and preposition-noun order in adnominal phrases ($at_{preposition}$ $home_{noun}$). Japanese shows the opposite correlation, with object-verb and noun-postposition order. Most languages (approximately 94 percent) follow one or the other of these patterns.[17] There are different explanations for why this is.

A diachronic explanation might begin by observing that in many languages adpositions historically evolve from verbs. What used to be a verb-object structure in a clause retains its original ordering but changes its structural/functional role to become an adposition-noun structure. What used to be a verb is now a preposition. In this account, the correlated structures are not independent. One has been derived from the other, over a timescale longer than an individual human lifetime.

Another kind of explanation appeals to processes or patterns in the *microgenetic* frame. This frame is the causal-temporal setting of cognitive processes: the processes of perception, attention, memory, motor control, and reasoning that individuals draw on in the real-time production and comprehension of language, or indeed of any waking moment of action or interpretation in a sociocultural setting. When Greenberg ([1966] 2005) first noted systematic correlations of word order patterns in languages of the world, he referred to the dominant correlated typological patterns (noted above) as *harmonic*. This idea was grounded in principles of individual cognitive processing of language. A harmonic relation in grammar is "connected with the psychological concept of generalization" (97).[18]

This idea is rooted in the general tendency for any entropy-resisting system such as a human body and brain to minimize energy cost for maximal benefit. In the first half of the twentieth century, the linguist George Zipf formalized and tested a "principle of least effort" in individual language processing, linking it to emergent properties of linguistic systems such as the robust tendency for more frequently used words to be shorter (see Zipf 1935, 1949).

Data such as the relative frequency of use of grammatical options in a language are seldom included in synchronic descriptions. For example, a synchronic grammar would state that in Sanskrit, nouns may be inflected for three numbers: singular, dual, and plural. It is implied that the three options are structurally equivalent. But the three options are used with radically different frequencies, introducing strong asymmetries in their values in the grammatical system. Greenberg ([1966] 2005) reported that 70 percent of Sanskrit noun tokens in texts are in the singular, 25 percent are marked for plural numbers, and only 5 percent are marked for dual numbers. This conforms with a cross-linguistic universal tendency for the singular to be the unmarked member of the set of markers of grammatical number. This often means that the singular is not only the most typical or most often encountered number, it is also often literally "unmarked" in that no formal marker is present (e.g., *dog* versus *dog-s* in English, where singular takes "zero" marking). This insight was made famous by Zipf more than seventy years ago, when he noticed that more frequently used words require less effort to process (Zipf 1949; see also Bybee 2010).

Hawkins (2004) includes [singular → plural → dual] among a set of *performance frequency rankings* that are accounted for by pressures on the microgenetic processing of language. Hawkins argues that patterns of grammar correlate with patterns of cognitive and motor performance. He proposes three ways in which the organization of grammars tends to maximize frequency: (1) minimize the domains to be marked (e.g., if there are three numbers, only mark two of them, and ensure that the most frequent one does not get any marking), (2) minimize the linguistic forms to be processed (i.e., make the markers short), and (3) "select and arrange the linguistic forms so as to provide the earliest possible access to as much of the ultimate syntactic and semantic representation as possible" (9).

Let us consider an example of how this kind of efficiency-based microgenetic argument has been applied in explaining grammatical structure. Since Greenberg's pioneering work comparing grammatical structure across the world's languages, it has been known that languages have a preference for prefixes over suffixes. Fifty-five percent of languages prefer suffixes, while only sixteen percent prefer prefixes (Martin and Culbertson 2020, citing Dryer 2013). It has been argued that this asymmetry follows from biases in human perception: "the beginnings of words are most salient to the human

speech-perception system, and this privileged position is reserved for the most important content: the stem" (Martin and Culbertson 2020, 1107). This argument appeals to the microgenetic frame of language processing, in which words are processed in milliseconds, as they are incrementally perceived. When you are listening to a person speaking, as information is coming in, millisecond by millisecond, you need to identify each next word in a sequence as soon as you can if you are to understand what's being said. In this context, if the stem comes first in an inflected word (as in *walk-ing*), the uniquely identifying part of the word comes first. You hear *walk* before you hear *-ing*. This favors faster word recognition and processing, and it decreases costs for language users, likely resulting in fewer incorrect understandings.[19]

We want to emphasize that this account is not purely about individually-situated cognitive processing. It is also supported by arguments from the enchronic frame (in ways that are hardly recognized in the psycholinguistics literature). If a certain organization of grammar tends to make it more difficult for hearers to recognize words in the flow of time, this would not only require more effort for listeners to process speech, it would also make such processing more prone to error. Considered purely from a microgenetic point of view, this creates an efficiency problem. But considered in the enchronic context of the speech event, with its socially interacting participants, the problem is not only about efficiency. If it leads to more disfluency, this would be expected to impact the always-relevant social-relational aspects of the speech event as well.

Himmelmann (2014) invokes the enchronic frame in his argument that suffixation is preferred from the point of view of the listener whose task includes, among other things, projecting the course of a current speaker's turn for the purpose of participating in the turn-taking of free-flowing conversation (see chapter 4). He presents an argument for the historical development of suffixes that is grounded in the two-party interplay of turn-taking and its implications for the prosodic and grammatical dependency of markers and the elements they mark.[20]

In a similar vein, Roberts and Levinson (2017, 402) develop an account of the historical evolution of constituent order across languages in the context of processing pressures that are "imposed from turn taking in conversation." An agent-based modeling experiment provides support for their contention that "interactive turn-taking in conversation must impose constraints on cognition" and "these may have implications for the way in which

languages change over time" (424). This view is in line with the argument long promoted by Schegloff (1989, 2006, inter alia), that the structures of grammar are shaped in the interactive context of conversation. Thus, Schegloff says (1989, 144), we must understand those structures to be "adaptations to the turn-at-talk in a conversational turn-taking system with its interactional contingencies."

We now close this brief review of some interacting biases that operate upon processes of language usage and transmission to give rise to grammatical structure as we know it. We conclude that ultimately, grammar is a cultural-evolutionary outcome of social interaction under intersubjectivity. The biases that have determined its shape are those set by the enchronic context in which intersubjectivity operates. Grammar is the population-level output of biased evolutionary processes that must pass through the enchronic bottleneck (chapter 4), and in turn grammar constitutes the shape of that bottleneck for subsequent iterations of the process.

7.8 Conclusion

We have argued that the foundations of grammar in human language are intertwined with human intersubjectivity, in the two senses that we have pursued through this book: (1) Grammatical systems would be impossible without primary intersubjectivity first being established in the social contexts in which language is used and evolves, and (2) once grammatical systems are established, they amplify, enrich, and enhance human intersubjectivity in turn. As de Villiers (2007, 1858) writes: "the interface between language and Theory of Mind is bidirectional." Because grammar is the outcome of linguistic usage in social interaction, any existing grammatical system has necessarily been created by interactions among people in a world of intersubjectivity. In turn, a grammar provides coordinates for, and constraints on, the localized nature of intersubjectivity and the horizons of interactional possibilities for social associates who inherit that system.

Consequences of Language

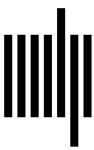
From Primary to Enhanced Intersubjectivity

By: N. J. Enfield, Jack Sidnell

Citation:

Consequences of Language: From Primary to Enhanced Intersubjectivity
By: N. J. Enfield, Jack Sidnell
DOI: 10.7551/mitpress/14795.001.0001
ISBN (electronic): 9780262372749
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Enfield, N. J., 1966– author. | Sidnell, Jack, author.
Title: Consequences of language : from primary to enhanced intersubjectivity / N. J. Enfield and Jack Sidnell.
Description: Cambridge, Massachusetts : The MIT Press, [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2022006408 (print) | LCCN 2022006409 (ebook) | ISBN 9780262544863 (paperback) | ISBN 9780262372732 (epub) | ISBN 9780262372749 (pdf)
Subjects: LCSH: Social interaction. | Intersubjectivity. | Anthropological linguistics. | Semantics.
Classification: LCC HM1111 .E536 2022 (print) | LCC HM1111 (ebook) | DDC 302—dc23/eng/20220428
LC record available at https://lccn.loc.gov/2022006408
LC ebook record available at https://lccn.loc.gov/2022006409