

8 Conditions and Consequences of Language

We have argued for two interconnected claims: first, a primary form of human intersubjectivity has made language possible in our species, and second, the emergence of language has, in turn, been transformative, resulting in the enhanced intersubjectivity that characterizes human societies and cultures today. Despite the impressive social capacities of other species, such as the other great apes, their interactions are not undergirded by the same elaborate interactional architecture of accountable intersubjectivity seen among humans.

The difference is a consequence of language. Our species' dynamic, language-enhanced form of intersubjectivity is foundational to all forms of historically constituted and transmitted knowledge. All macro-level institutions and networks, we argue, are ultimately the product of context-situated enchronic encounters among invested human agents.

There would be no enhanced intersubjectivity without accountability and there would be no accountability without language. Accountability and language are thoroughly intertwined, in at least two ways. One is that language use and linguistic meaning are themselves normatively regulated and the distinctive, indeed unique, properties of language make it subject to a starker and more fine-grained form of accountability than any other form of human conduct. This is partly because of the discrete, granular properties of utterances that allow for precise segmentation, isolation, and identification of what has been said, even to the scale of individual grammatical markers.¹ The other is that language can figure as an instrument, rather than object, of accountability, as we explored in part II.

Let us finish not with a summary of our argument but by highlighting some of its implications.

8.1 The Linguistic Foundations of Language?

There is a paradox at the heart of our argument. We have argued, in essence, for the *linguistic foundations of language*. But how can language be grounded in language? There are in fact two puzzles here. First, if language is a system whose elements derive their sense, in part at least, from contrast with other elements, how can the system exist in advance of the elements of which it is composed? Second, if a definitive property of language is the presence of metalanguage, how could there be language about language if there were not already language? We want to offer a solution to this paradox, one that solves both its constitutive puzzles: no elements without system and no language without metalanguage.

Concerning the second of these puzzles, many contemporary thinkers assume that “reflexive discourse is a superficial supplement to language itself, one which could be removed without seriously affecting language” (Taylor 2000, 486). Searle (1995, 73), for instance, suggests, that language is “self-identifying.” As Taylor (2000, 487) puts it:

[Searle] argues that institutional facts such as money, property, marriage, etc., require language—indeed “are constituted by” language. However, he claims that language is the exception to this. For, although also institutional, linguistic facts—such as the fact that *comely* is a word of English or that it means “pleasant to look at”—do not require language. Language does not require (meta)language in the same way that other institutional facts do. Instead, language is precisely designed to be a self-identifying category of institutional facts.

But Taylor asks, “How far could human language have evolved without the use of reflexive discourse?”

Naturally, this question sounds somewhat paradoxical—as if the suggestion were that early humans could not have developed “primary” language (language itself) until they had first developed a higher order language, a metalanguage, for talking about primary language. Of course, one might then want to ask: “Well, if that’s true, then how could they have developed metalanguage until they had first developed meta-metalanguage for talking about metalanguage?” The absurdity of the regress is obvious. (489)

Paradox or not, it is undeniable that a language in which one could not perform the elementary metalinguistic operation of agreeing with what someone else has said, or asking for clarification, or indeed asking what a particular gesture or word was intended to convey (let alone “mean”), would hardly be worthy of the name.

We think that a solution to the riddle can be found in the universal organization of *repair* (section 4.5), and in particular in the discoveries that (1) *Huh?* is a word and, (2), it is found in every known language (Dingemanse, et al. 2013). The universal distribution of this word is explained by convergent evolution. We speculate that *Huh?* is “as close as you can get to the core of the human faculty for language” (Enfield 2017, 207). Could *Huh?* have been the first word, and the fount of language?

Imagine the scenario with which we began this book: a time in which protohumans communicated using voluntary, purposive gestures like those of contemporary nonhuman primates such as bonobos and orangutans (Rosano 2013; Tomasello 2008). These gestures could be used to make requests, to initiate play, to get another’s attention, to beckon, and to offer. These gestures were, as researchers have shown for other primates, used in sequences of social interaction. For instance, a beckoning gesture addressed to a particular other could establish the relevance of a specific form of response, for instance compliance by moving toward the beckoning individual. Where a relevant response was not produced, the beckoner might produce a second attempt or might move to a location that was more visible to the addressee and thereby attempt to secure their attention.

Such gestural communication would have provided the requisite conditions for the use of *Huh?* Imagine an early human is addressed by a gesture, but fails to understand its intended import and instead utters an interjection of puzzlement in the sequential slot where their behavioral response should have occurred. The interjection is metacommunicative insofar as it is *about* the communicative action just made by another agent. This is the wedge that introduces metalanguage into a not-yet-linguistic system. Then imagine that this response promotes the redoing of the gesture. In turn, the link between the interjection of puzzlement and the redoing of the not-yet-understood gesture becomes ritualized and so may be used again to achieve the same end—a repetition of the previous communicative move—either by this protohuman or by another who happened to witness the events unfold. We can imagine here a shift from learning via ontogenetic ritualization to imitation, that is, conventional transmission, the beating heart of language and culture.

Would this constitute language? No, but it would explain how you can have metalanguage prior to the establishment of language as we know it. This is because, as we explained in chapter 5, *Huh?* is a metalinguistic vocal point or deictic. It orients to an immediately prior communicative act.

This leads to the other of the two component puzzles: how system-like relations can arise from such item-based beginnings (see Enfield 2014 on the “item-system problem”). The key is to understand that once a slot is established for an expected action in a recurrent communicative sequence, alternatives to that action—as long as they occur in the same slot—derive meaning purely from the fact that they are alternatives to a known or expected point of reference (cf. Darwin 1872, 50ff on the principle of antithesis in the expression of emotion). In this way, once a protoword like *Huh?* was established, it created a slot into which other interjections might meaningfully occur. The possibility of distinct elements occurring in the same syntagmatic slot creates the potential for one of the fundamental features of a linguistic system: paradigmatic alternation (de Saussure 1916; Zipf 1949). We have known of such grammar in the structure of interaction (i.e., syntagmatic-paradigmatic organizations of interactional moves) at least since Schegloff described the case of “uh-huh” as an alternative to the initiation of repair (Schegloff 1982; Sidnell 2009b). So, we can imagine that once *Huh?* was established, an opposing form could easily occur in that slot, indicating the “antithesis” of *Huh?*—that is, conveying the meaning that the prior move has been understood and that compliance is to follow.²

This sort of simple paradigm of opposed, minimally semantic forms would constitute a basis for language. This basis is not in the logical concepts of reference and predication or in the individual cognitive capacity to represent events and situations. Instead, it emerges from universal structures of semiotic action in the enchronic frame of move-by-move accountable interaction, of the kind illustrated, for example, by the two men sawing timber in chapter 1. A simple paradigm of responsive interjections might thus serve as initial steps toward grammar (see Hurford 2012 for an account of the evolution of grammar beginning with interjections). Subject to laws of semiotic conventionalization, over time these interjections would take on more semantic nuance and content. And these forms would provide jumping-off points for expansions in the form of additional slots and paradigms, kicking off open-ended processes of grammaticalization that underpin the complex grammatical systems seen in the world’s languages today.

8.2 Implications and Contributions of Our Argument

Before concluding, let us draw connections to some parallel lines of work where we feel that our account can contribute value.

8.2.1 The Intelligence for Language

What happened in our evolutionary past such that we possessed the cognitive prerequisites to make the steps to language that we have described? We have presented arguments and evidence that the human capacity for language is grounded in a rich social-interactional infrastructure, comprised of diverse elements.³ Recall the first of our two main claims in this book: that language presupposes primary human intersubjectivity, which is inherent even in our most basic forms of joint activity, such as the two-person timber sawing discussed in chapter 1. This claim is consistent with an influential line of thinking in research on the evolution of primate intelligence, which has argued that what sets humans apart from our closest relatives—who, like every other species, *lack* language—is not our intelligence but our *social* intelligence.

Ethologists in the 1960s were trying to understand the origins of higher primates' intelligence. They realized that the evolutionary pressure for ever-better brains was not coming from any special new challenges presented by the physical world. Alison Jolly introduced the idea that the evolutionary context for increasing primate intelligence was not in the physical world but in “primate social life” (1966, 506). Similarly, neuropsychologist Nick Humphrey asked why humans are “much cleverer than they need be” and answered that “the higher intellectual faculties of primates have evolved as an adaptation to the complexities of social living” (1976, 316).⁴ The ensuing *social intelligence hypothesis* (Byrne and Whiten 1978) explored these connections. When primatologist Robin Dunbar saw a correlation between increased primate brain size and increased social group size, he hypothesized that language emerged as part of a social-cognitive revolution in our species. Riding on increased psychological capacities for tracking and understanding multiple embedded social relationships and perspectives, language provided a way to scale up the individual's number of social relationships beyond the overly time-demanding limits set by physical grooming (Dunbar 1993, 2003). Work in this vein has reframed the function of language similarly to the way Jolly and Humphrey reframed the

function of primate intelligence. The idea is that we will only make progress when we understand that the function of language/intelligence is primarily social, not primarily informational/logical.

So far so good. But the line of research on language opened up by Dunbar has not progressed, for two reasons. The first is methodological. This line of work has assumed that the social functions of human language can be studied in the same ways as the social functions of other forms of animal communication. It has been concerned with questions like what people talk about, who people talk to, and for how long. To take one example, Dunbar, Duncan, and Marriot (1997) studied human conversation using the methods that Dunbar was trained to use in his pioneering work on the social behavior of gelada baboons in the Ethiopian highlands (Dunbar 1985). The researchers sat near people in cafeterias, bars, and trains, taking field notes on their conversations through “direct auditory monitoring” rather than audio- or videorecording the interactions:

To assess the content of conversations, the general topic of the speaker’s utterances was scored as an instantaneous scan sample taken at 30-second intervals: the topic of conversation at the moment of the sample was judged by the immediately preceding period of conversation. (1997, 234)

Leaving aside the difficulties of determining the “topic of conversation” at any given moment, this method has no way of measuring or analyzing the workings of the enchronic infrastructure for human interaction that are central to the management of intersubjectivity, as we have shown in this book. The ethological approach to language has, perhaps paradoxically, never appreciated that human social interaction is conducted in radically different ways from that of other primates precisely because language as we know it is the medium for interaction.⁵ Language has been treated as a scaled-up tool for performing just the same functions as physical grooming: managing and tracking social affiliation and reputation. The ethological approach has not looked at the far richer measures of language’s social functions that have been explored over the last half century in sociology: turn-taking, sequence organization, repair, and enchronic accountability. In this book, we have placed those things front and center.

The second limitation of Dunbar’s line of research on the evolutionary foundations of language is that it doesn’t explain why the grammatical structures of human languages are as complex and diverse as they are (Hurlford 1999, 2012). Michael Tomasello, another evolutionary psychologist,

developed a proposal that addressed this by distinguishing between three levels of complexity in language structure and correlating these with three broad classes of social function that language can have (Tomasello 2008):

Simple grammar ↔ requesting

Complex grammar ↔ informing

Fancy grammar ↔ sharing

But this theoretical proposal is highly abstract and needs to be fleshed out conceptually and studied empirically. In this book we have made some steps in that direction. In chapter 7, we outlined some elements of what might constitute “fancy grammar” in more specific terms than those offered by Tomasello (2008), though we are still far from providing the needed empirical demonstration of Tomasello’s conjectures.

In sum, while research grounded in ethology and the study of primate social behavior and intelligence has made extraordinary progress, we have here outlined paths that this work is yet to take. We believe that these paths will lead to considerable advances in our understanding of the foundations of human language through an account of its social-interactive underpinnings that adequately addresses the structural complexity of the phenomena.

8.2.2 The Social Logic of Meaning

The idea that language is a system for carrying out logical operations on information has been around as long as philosophy itself. But language in real life doesn’t tend to conform with textbook principles of logic. Consider the following exchange:

Professor: Where is your assignment?

Student: I’ve been ill this week.

In purely logical terms, the student’s response is a non sequitur, but in natural conversation it is perfectly apt. In 1975, the philosopher Paul Grice sought to reconcile the mismatch between the known principles of formal logic and the principles that people apply in understanding utterances in everyday conversation. Grice (1957) had earlier established that meaning in language involves not just the semantics of words but also a higher-order consideration of the intentions of the person speaking. His work took the question of the understanding of language—which was widely regarded as a matter of logical calculation by rational individual agents (in the En

domain)—and reframed it as an operation by which interpreters must take into account the presumed intentions of speakers (in the *Es* domain), and by which, in turn, speakers must design their utterances assuming such accounting by interpreters. Grice established that linguistic meaning arises from a dialogic interplay of minds, along lines not dissimilar to the game-theoretic models of interpersonal strategies of communication and manipulation developed by economist Thomas Schelling (1960; see also Lewis 1969).

This breakthrough in the philosophical conception of how language works led to much subsequent research on the interpersonal logic of language use, which in turn drew attention to the kind of cognition that was required for language to be learned and used (see previous section). If language use requires us to be able to model the knowledge and intentions of our interlocutors in real time, then this means that the psychological wherewithal for language must include not only the capacity for symbol manipulation, but more importantly the higher-order social-cognitive capacities needed for “mindreading” (Sperber and Wilson 1986; Levinson 2000, 2006). In an example of how this line of work has been taken up in research on language evolution, Scott-Phillips (2015) emphasizes the role of the uniquely human cognitive advance of “ostensive-inferential communication,” the essential insight of Grice (1957). Scott-Phillips states that humans are “adapted for ostensive communication, that is, to express and recognize communicative and informative intentions, and in doing so better navigate the social environment” (2015, 139).

But questions remain open. First, what does “navigating the social environment” mean? Scott-Phillips emphasizes the marked increase in group sizes that humans show, and notes that “large group sizes make for a very politicized existence” (2015, 140). But every group-living animal has a politicized existence. Are we talking about an existence that is merely “more” politicized than that of other primates, or was there a qualitative break? The answer, as we’ve demonstrated in this book, is that there was a qualitative break. Language-mediated micro-politics are more than mere matters of manipulation or influence over others. Language doesn’t just allow us to navigate a social environment, it transforms and indeed constitutes the very social environments we must navigate. This is one of language’s most momentous consequences.

Second, is the neo-Gricean emphasis on “mindreading” (something that listeners do) and “manipulation” (something that speakers do, exploiting

what listeners do), perhaps paradoxically, overly individual-based? The account preserves a view of humans more as calculators than as agents. Scott-Phillips (2015, 141) is explicit about this: “communication is not a single trait, but rather the product of two other traits, namely mechanisms for signal production, and mechanisms for signal reception.” That is, there are two individual-based mechanisms: “signallers try to mentally manipulate their audience, and audiences try to mindread their signallers” (ibid.). This notion of “social navigation” may require sophisticated cognitive abilities but it overly simplifies what interactants really do with language in interaction—under a tyranny of accountability—as we’ve endeavored to show in this book.

8.3 Consequences of Language

We have argued that language is grounded in an enchronic infrastructure for human interaction, and that this infrastructure contains a uniquely human form of primary intersubjectivity. In turn, we have argued that the infrastructure is a condition for a type of enhanced intersubjectivity that emerges through the operation and development of language within that infrastructure. Other species have complex interactions but they do not have culturally transmitted systems of signs akin to human language, with duality of patterning (in which semantic symbols are composed of distinctive nonsemantic units, allowing for unlimited growth of the lexico-semantic component), generative grammar (in which meaningful signs are combinable in patterns that are themselves meaningful), and the capacity for self-reference. Lacking these capacities has significant consequences for intersubjectivity.

Perhaps the most important feature of interaction sequencing in humans is that the structures it produces are accountable. If a recipient does not produce the conditionally relevant next action, this is an accountable fact, and people orient to it. Recipients may account for not responding (e.g., by indicating that their reason for not answering is that they do not know) and first-speakers may pursue such accounts if they are not offered. Another aspect of the accountability involved in interaction sequencing is seen in reports and complaints about third-party failures (e.g., “He didn’t answer me”). Such reports radically transform the nature of human social life: “Humans alone among the social vertebrates can know one another substantially by repute” (Emler 1990, 177–178). Reputation has untold consequences for human

sociality (see Boissevain 1974; Emler 1990; Enquist and Leimar 1993; Cohen 2012; Sperber and Baumard 2012; Origgì 2018). Other apes, not having available to them a flexible semiotic code like a human language, are accountable in the domain of brute forces, not social norms.

Philosophers have argued that human institutions would not exist without the reality-creating power of language as a device for enacting status-function declarations (Searle 2010). We argue that this power alone cannot account for how human institutional reality is created and maintained. Intersubjective understanding is necessary—as both condition and consequence—for any form of human activity. As we saw in the examples of sawing and weaving, it is not limited to overtly communicative activities. And this generalizes beyond face-to-face engagement.

All cooperative work involves an underlying enchronic architecture and so it makes little sense to isolate symbolic and communicative life from economic and material life. The enchronic infrastructure is a condition for all institutional reality, from markets to psychiatry, news interviews to court proceedings, medical consultation to healing ritual, crime to punishment, marriage to gift exchange. But these social realities are not consequences solely of the enchronic to-and-fro of intersubjective action. They have been made possible within that to-and-fro by the propositional, self-referential, and accountable systems of reference, semantics, and grammar. They are consequences of language.

This is a section of [doi:10.7551/mitpress/14795.001.0001](https://doi.org/10.7551/mitpress/14795.001.0001)

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](https://doi.org/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

© 2022 N. J. Enfield and Jack Sidnell

This work is subject to a Creative Commons CC-BY-NC-ND license.
Subject to such license, all rights are reserved.



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://lcn.loc.gov/2022006408>

LC ebook record available at <https://lcn.loc.gov/2022006409>