Symbol Grounding and the Origin of Language: From Show to Tell

Alexandre Blondin Massé[1], Stevan Harnad[1,2], and Bernard St-Louis[1]
[1] Institut des sciences cognitives
Université du Québec à Montréal
http://www.crsc.uqam.ca/
[2] Department of Electronics and Computer Science
University of Southampton
Highfield, Southampton, UK
http://www.ecs.soton.ac.uk/~harnad/

Abstract: Organisms' adaptive success depends on being able to do the right thing with the right *kind* of thing. This is categorization. Most species can learn categories (1) by direct experience ("induction"). Only human beings can learn categories (2) by word of mouth ("instruction"). Artificial-life simulations have shown the evolutionary advantage of instruction over induction and human electrophysiology experiments have shown that these two radically different ways of acquiring categories still share some common features in our brains today. Graph-theoretic analyses reveal that dictionaries consist of a core of more concrete words that are learned earlier, by direct experience (induction); the meanings of the rest of the dictionary can be learned by definition (instruction) alone, by combining the inductively grounded core words into subject/predicate propositions with truth values. We conjecture that language began when attempts to communicate through miming became conventionalized into arbitrary sequences of shared, increasingly arbitrary category names that made it possible for members of our species to transmit new categories to one another by defining and describing them via propositions (instruction).

1. Show vs. tell

Like many papers, this one began as a PowerPoint presentation although that form of communication leaves a lot to be desired, from both the hearer's and the speaker's perspective. As a hearer, one usually finds that the PowerPoints, far from clarifying, conflict with or distract from what the speaker is saying. And as a speaker, one finds that they cut down on the flexibility and spontaneity of one's presentation (though of course writing it all out in advance cuts it down even more!).

But this paper concerns the origins of language, and language began before PowerPoints. Moreover, in some ways language also supersedes PowerPoints. PowerPoints are "show and tell," but more show than tell. In this chapter we will try to remind you of the power of telling alone, the oral tradition. In many ways, the origin of language amounts to *the transition from show to tell* (as was portrayed in the pantomime performance enacting the advent of language at the opening of the Summer Institute on which this volume is based) (Harnad 2000, 2003)

2. Before orality

The oral tradition could not have begun, of course, with formal oratory, before a large audience, as this paper did. It must have started with something much more informal, local and interactive: something closer to conversation. And writing began much, much later. So reading aloud a talk like this is a bit of a cheat. Nor is it likely that language began with word of mouth, for reasons we will describe shortly.

So the origins of language, we suggest, do not correspond to the origins of vocal language; nor do they correspond to the origins of vocal communication. Language itself began earlier than vocal language (though not earlier than vocal communication, obviously). Once it had made its power and adaptive benefits felt, however, language simply migrated (behaviorally, neurally, and genetically) to the vocal modality, with all of its obvious advantages (Steklis & Harnad 1976). Before it could migrate, however, language itself had to be born.

How was language born? And why? And what was the adaptive advantage that it conferred? That advantage must have been enormous, to have become encoded in our genotypes and encephalized in our language-prepared brains as it did. Was language "invented," the way writing and print and the Internet were invented (Harnad 1991)? Or did it somehow differentiate gradually (rather like a tadpole turning onto a frog) out of a long, perhaps endless, succession of precursors (perhaps "protolanguages" of the kind that were often evoked during the Summer Institute)? And when did the evolutionary hard-coding of language into our genes and brains end and the historical soft-coding through learning and experience take over?

3. What is language?

Before we can try to answer any of these questions we are going to have to take a stab at what it is that we are trying to explain the origins *of*: *What is language*? None of the contributors to the Summer Institute ventured to define language explicitly, but there have been some implicit definitions. Some of these have been very broad—what Fitch (this volume), with Chomsky, calls the "Faculty of Language, Broad (FLB)" (Hauser et al 2002). This would encompass not just phonology, grammar, semantics, and pragmatics, but all the other structures and skills involved in or dependent on language, from speaking and hearing to perceiving and acting, including socializing and hunting, tool-making and using, communication, pointing, imitation, miming, mind-reading, and their many neural and genetic substrates and precursors, as rooted in the brain's dorsal and ventral pathways, canonical neurons, and mirror neurons, in ourselves, in our cousin species the apes, and in our ancestor species, the hominins, hominids, and perhaps even katydids.

If all of this is language, or a precursor part or product of language, we're aiming for such a huge target that we can hardly miss it. But having managed to hit it, it's not obvious what it is that we've shot down.

In contrast to this broad notion of language, there are the narrower notions (Fitch's "Faculty of Language, Narrow" (FLN)), such as speech. Or, even more minimal, the "minimalism" mentioned repeatedly (e.g., by Jackendoff, this volume), but not described. We won't try to describe minimalism either, because we can't. But in this view, although it is not the whole of language, the quintessential property of language is *Universal Grammar* (UG), a complex set of rules that cannot be acquired explicitly from the child's experience because the child does not produce or receive anywhere near enough information or corrective feedback to learn the rules of UG inductively, by trial and error. So UG must be inborn (Harnad 2008, 2014).

If UG is central to language and is inborn rather than learned, then that would turn a big part of the question of the origin of language into the question of the origin of UG. That question has already been raised (for example, by one of us!) 35 years ago, at the first language-origins conference in New York (Harnad 1976).[i] And Noam Chomsky's answer then was more or less that there is no need to ask or answer that question (Chomsky 1976).

Since then, some people have tried to answer the question anyway. But explaining the evolutionary origin of UG faces some problems, because UG is not the sort of thing—like wings, limbs, hearts, or eyes—for which a straightforward evolutionary story can be told without an awful lot of ad hoc hemming, hawing, and vagueness. Other options are to deny that UG exists at all, or to argue that it *can* be learned after all, and hence that there is no need to worry about how it could have evolved.

We will try to avoid the pitfalls of both the over-wide and the over-narrow view of language by proposing not a definition of language but a test to determine whether or not something is a language at all. First, however, it has to be understood that our test pertains to *natural* language, such as English, French, American Sign Language, or Proto-Indo-European. It is not about artificial formal languages such as mathematics, logic, or computer programming languages. Those artificial languages do have formal definitions, but we suggest that in reality all of them are merely parts or subsets of natural language.

3.1. The power to say anything that can be said

So what is our test of whether something is a natural language? It is based on a criterion independently proposed by both Jerrold Katz (1976) and Steklis & Harnad (1976), likewise 35 years ago at the *New York Academy of Sciences* conference. It is a criterion whose implications it has taken its author 35 years to understand.

Katz's version was the "glossability" thesis: *A natural language is a symbol system in which one can express any and every proposition*. A proposition is any statement, subject plus predicate, that has a truth value, true or false. "*2+2=4*" is a proposition; so is "*the cat is on the mat.*" And so is every sequence of words in this text that ends with a period. (Questions and imperatives are slight variants, but much the same sort of thing.) A very general fact also needs to be pointed out here that will be important for what we say later: *Every proposition is also a category inclusion statement*. In other words, saying that *A is B* is the same as saying that the subject category, A, is a subset, member, or part of the predicate category, B.

Our version of the very same criterion as Katz's Glossability Thesis was the "translatability" thesis: Anything you can say in any natural language can also be said in any other natural language—though, it is important to add, *not necessarily said in the same number of words*. Word-for-word translation is often not possible, as you have discovered if you have used *Google Translate*. One language may say some things less economically than another. But it can still say them all.

For those readers who have doubts, and think there is something one can say in another language that one cannot say in English: please tell us —in English, at any desired length—what that something is, and why one cannot say it. You will find that you have generated your own counterexample.

3.2. What would "protolanguage" be?

Now, with this test for whether or not something is a language—i.e., whether it is a system in which you can say anything that can be said—let us immediately consider one of the notions that is repeatedly invoked in this volume: "protolanguages." Remember that the theme of the Summer Institute was the origin of language. If a natural language is a symbol system in which you can express any and every proposition, what about a protolanguage? If you cannot express all propositions in a protolanguage, which ones can you *not* express, and why not? (We will leave this, too, as an exercise for readers to ponder on their own, but we predict that the exercise will not be successful, and the reason for its failure will be significant for our attempt to get at the heart of what is and is not a language.)

Alternatively, if one *can* express all propositions in a protolanguage, then why call it "protolanguage" at all? For then the question we are addressing would simply become: What is the origin of protolanguage? Because, as they say, it can't be turtles (or tadpoles) all the way down.

So, if you accept our criterion for what constitutes a natural language, we have now reduced the question of what is the origin of language to the question of what is the origin of a symbol system in which you can say anything that can be said in any natural language. When did that happen? Where did it happen? Why did it happen? How did it happen? And what was the tremendous adaptive advantage that it conferred—an advantage great enough to have resulted (within a relatively short period of evolutionary time) in the fact that that new capacity and propensity became encoded in the genotypes and encephalized in the brains of every normal member of our species?

We won't be able to tell you when and where it began happening; but we see no reason not to accept the evidence of the archeologists who point out that, whereas there has been longstanding evidence for a number of uniquely human practices such as the making of weapons, tools, and other artifacts for hundreds of thousands of years (all of which may or may not have already been coexisting alongside language), it is likely that by the time these uniquely human practices began to multiply rapidly (between 50,000 and 250,000 years ago), human beings already had language -- and that the rapid growth of knowledge and culture was in fact one of the symptoms of the adaptive benefits language itself was conferring on us.

3.3. What is a symbol system?

Now tools and weapons—and ways of making them—are clearly inventions. Was language itself an invention? Did we invent a symbol system in which you can express any and every proposition? Well, in a sense we must have, but not quite in the same sense that we (much later) invented writing or print or the Web. To see this, we have to look more carefully into what a symbol system is. We mentioned earlier that by natural language we do not mean artificial formal languages such as mathematics, logic, or computer programming languages: those formal languages are already parts of natural language.

The formal definition of a symbol, however, can be taken from these artificial languages (Harnad 1990): a symbol is an object—an object of arbitrary physical "shape." But a formal symbol cannot exist or be defined or described in isolation. A symbol is essentially part of a *symbol system*, a set of such objects, along with a set of rules for manipulating those objects—combining and recombining them into composite shapes, *based on rules that operate only on the shapes of the arbitrary symbols, not on their meanings*. Some of the combinations of symbols will be "well formed" according to the rules, and some will not be well formed. The rules for combining them, and for determining which combinations are well formed, are called "formal" or "syntactic" rules, because they are based only on shape, and, as we noted, the shape of the symbols is arbitrary.

What does that mean? *Arbitrary with respect to what?* Well, so far, syntactic rules sound like they are merely the rules of some formal game, a game of manipulating arbitrary symbols by combining and recombining them. But why would we play such a formal game? We are not heading here toward the notion that the precursor or driver of natural language was just game-playing. Besides, we are still discussing *formal* languages, not yet the full power of *natural* languages. But even formal languages consist of more than just syntax. Not only are some symbol combinations syntactically well formed and some not. But the syntactically well-formed ones also have a *semantics*: they can be systematically interpreted as *meaning* something.

Arithmetic is a familiar example of a formal symbol system. You have symbols like 0, 1, +, –, =. The combination "*0 1 – +*" is ill formed, whereas the combination "*0 +1 =1*" is well formed. "*0+1=1*" also happens to be *true*, whereas "*1+1=1*" is well formed but *false*. So we need two more symbols, T and F, to formulate yet another set of syntactic rules, this time not to determine which combinations are well formed, but to determine which well-formed combinations are True and which False.

That's still all just syntax too and consists of some axioms (which are strings of symbols that are assigned the symbol True), and syntactic rules for combining strings of well-formed symbols into longer strings called derivations or proofs. The proofs themselves are all just syntactic too: you can't say that *2+2=4* is true just because you "know" it, or just because you can "see" it. You have to give the proof, which is just a string of rule-governed symbols. But nevertheless, if you have picked the right syntactic rules and the right axioms, then all and only the true propositions of arithmetic will be computable using just the syntactic rules.

(Actually, not all of them, on account of Goedel's incompleteness theorem, but let's not get into that here, because it is not really relevant to the point we are illustrating. If everyone were as familiar with propositional calculus or predicate calculus as they are with arithmetic, we could instead have used either of those for our formal example of a symbol system, and that too would have been all syntactic, as well as true, except that it would also be complete, and hence not vulnerable to Goedel's incompleteness theorem.)

So arithmetic is a set of symbols, manipulated only syntactically—that is, only on the basis of their shapes, not their meanings. Yet the well-formed symbols and symbol combinations all do have meanings, namely, all the true propositions of arithmetic.

3.4. Arbitrariness of shape, autonomy of syntax, and meaning

We can now see the sense in which the "shapes" of the symbols themselves are arbitrary: *0* is not "shaped" like nothingness, *3* is not shaped like threeness, the equals sign is not shaped like equality, and so on. Nor is the proof of *2+2=4* shaped like truth! The symbols and symbol combinations simply have the remarkable property that if the syntactic rules are well chosen they can be systematically interpreted as the true propositions of arithmetic. We could have chosen objects with other shapes to serve as our symbols. Our choice is just our arbitrary notational system, on which we all jointly agree, by convention. But the syntactic rules, if they were right, would give the same result in any notational system.

Notice that in a formal symbol system we have true "autonomy of syntax," in the sense that the rules governing the symbol manipulations are completely *independent* of meaning and purely *formal*: they have no need for or recourse to our semantic interpretations. Yet they *are* systematically interpretable as meaningful—interpretable by us, the users of the formal symbol system, who know what "2" and "equals" and "true" actually mean. But when we use *2* or *equals* as words in natural language, it is no longer just the manipulation of symbols based on syntactic rules operating on their arbitrary shapes, as in a formal language. In natural language, we also manipulate words based on their *meanings*.

Natural language, too, has arbitrarily shaped symbols, agreed upon by convention, as well as syntactic rules that determine what is and is not well formed. And its symbols are likewise systematically interpretable as meaning something. (Some people argue that natural language's syntax is autonomous from meaning too (Koster 1986). There are perhaps more grounds for reservations about that, but let's assume that's true too.)

What is certainly true is that, when we manipulate natural language symbols (in other words: *words*) in order to say all the things we mean to say, our symbol manipulations are governed by a further constraint, over and above the syntax that determines whether or not they are well-formed: that further constraint is the "shape" of the words' meanings.

4. Categorization

With the exception of a few *function* words, such as *the* and *if* and *not,* most of the words in a natural language dictionary are *content* words, and as such they are the names of *categories*. Categories are *kinds*: kinds of objects, events, actions, properties, states. To be able to categorize is to be able to do the right thing with the right *kind* of thing (Harnad 2005). And things have shapes. And so do we. So to be able to do the right thing with the right kind of thing, our brains have to be able to distinguish the shapes of the members from the nonmembers of each of the categories we possess.

Categorization is so general that it covers just about all of our behavioral and cognitive skills—all those that depend on a discrete or categorical decision. This excludes only continuous skills, such as walking, swimming, basketball playing; we can categorize some of the discrete choice-points of their dynamic actions, but the skills themselves are *continuous* rather than categorical, and their dynamic "shapes" are certainly not arbitrary but congruent with the real-world objects and states on which they operate—congruent with what Gibson (1979) has dubbed their "affordances," or the shape of what the objects of our actions afford *doing* with them, with bodies that are shaped like our own.

Most of our categories are not inborn. We have to learn what things are contained in the categories *animals, plants, owls, butterflies, rocks, mountains, trees, tables, earthquakes, quarrels, walking, running, swimming, flying, basketballs, basketball playing, green, big, equal, tall, inside, under, risky, goodness, truth,* and *beauty*.

Even proper names (such as *Ferdinand de Saussure* or *Noam Chomsky*) name categories, in the sense that every time you see the individual who has that proper name, he does not look exactly identical to the last time you saw him: position changes, clothes change, time changes—all these instances have to be recognized as being members of one and the same category: instances of the individual person named (otherwise we would remain lost and powerless in a succession of individual, unique snapshots, like Borges's (1964) *Funes the Memorious*.).

4.1. The symbol grounding problem

All those content-word categories listed in our dictionaries have members. If we know what a given word means, we know what are and are not the members of the category it denotes (although there is no resemblance between the shape of the word and the shape of the members of the category it denotes.) How do we learn the meaning of a word? Well, we can look it up in a dictionary. The dictionary will *define* it for us—on condition that we already know the meaning of the words used in the word's definition. If not, no problem, we can look up those words too, and so on. But, as with so-called "protolanguages," it can't be definitions all the way down.

This is the *symbol grounding problem* (Harnad 1990). In a natural language symbol system, some, at least, of the meanings of some words have to be acquired by some means other than verbal definition. The natural candidate, of course, is direct sensorimotor experience: since words are the names of categories, we can learn which are and are not their members through trial and error *induction*. An example would be learning which mushrooms are edible and which are toxic. You sample a little bit of a mushroom, wait to see whether you get sick or you get nourished, and you keep sampling different mushrooms often enough till your brain has learned to reliably tell apart the edible and inedible ones (if you have not poisoned yourself or starved first). When we successfully learn a new category by sensorimotor induction, our brains learn to detect the sensorimotor *shape* of the feature(s) that reliably distinguish the members from the nonmembers.

4.2. Simulating the origin of language

Cangelosi and Harnad (2001; Cangelosi et al. 2002) did an artificial life simulation of creatures acquiring categories through sensorimotor induction in a "toy" world with three different kinds of mushrooms: A, B, and C. We will not describe the details except to say that two of the categories, A and B, could only be learned directly, by trial and error induction, with neural nets learning to detect the features that distinguished the members from the nonmembers for each of the first two categories of mushrooms (A/non-A; B/non-B). For the third category, C, there were two ways it could be learned, because the rule for membership in that category was simply the *conjunction* of the rule for A and for B: A mushroom was a C if it was both A and B; otherwise, it was non-C.

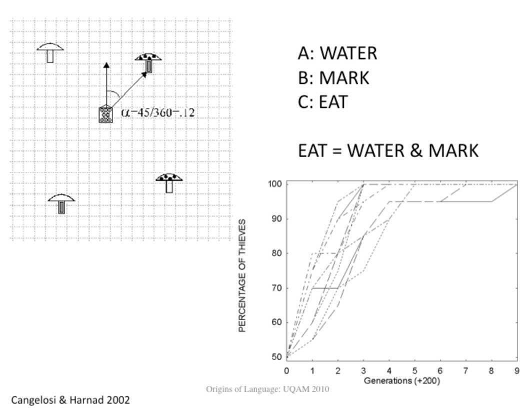


Figure 1. Artificial life simulation of the evolutionary advantages of instruction over induction. In a simulated mushroom world the category "A" mushrooms and "B" mushrooms could only be learned through trial and error induction, but the category "C" mushrooms could be learned either through induction or through instruction (C = A + B). Within a few generations, the virtual creatures that tended to learn C by instruction out-survived and out-reproduced those that tended to learn C by induction, demonstrating the

adaptive advantages of instruction over induction in the acquisition of new categories (Cangelosi & Harnad 2001).

We put two kinds of creatures into competition with one another in our toy world: creatures that could only learn by induction, through corrective feedback from lengthy trial-and-error sampling to detect the distinguishing features, and creatures that could learn by either induction or instruction, by "overhearing" creatures who already knew categories A, B, and C, "vocalizing" when they picked C mushrooms, because they vocalized the names *A, B,* and *C.*

This is a bit of a cheat, because we wired the creatures to vocalize the category names; but -- since we do not believe language started vocally anyway -- think of the category "names" not as vocalizations but as *observable actions*. As noted earlier, categorization is "doing the right thing with the right kind of thing." And of course naming is just one kind of "doing"—a late, arbitrary kind. In the simulation, the creatures do something different with the two kinds of mushrooms; they perform a different sensorimotor act. Let's say they WATER the A mushrooms, they MARK the location of the B mushrooms, and they EAT the C mushrooms. So let us call the three kinds of mushroom the WATER/non-WATER mushrooms, the MARK/non-MARK mushrooms, and the EAT/non-EAT mushrooms, with the "names" enacted bodily, by performing the right nonarbitrary actions on the members of the category, rather than vocally.

As we said, the WATER and MARK categories could only be learned by induction, so all the creatures learned them that way -- both the induction-only learners and the instruction learners. (That's the symbol grounding.)

But then when it came to learning the EAT category, that too could be learned in the long, hard, risky way—by trying to detect the distinguishing features through trial-and-error sensorimotor induction—or it could be learned in a fast, one-trial way: by observing that others, who already knew the category, performed WATER only with WATER mushrooms, MARK only with MARK mushrooms, and WATER+MARK+EAT with EAT mushrooms. This is learning by symbolic instruction.[ii]

When we put the induction-only learners in competition with the instruction learners, within a few generations the instruction learners had out-survived and out-reproduced the induction learners, of which there were no longer any survivors left. (We called the instruction learners "thieves" for reasons left to the reader. Here is a hint: Unlike the theft of food, the "theft" of *categories*—avoiding having to do induction by taking instruction instead—is a victimless crime; indeed, the benefits of instruction ("hearsay") depend essentially on mutualism; Poynder & Harnad 2007.)

4.3 Pantomime to proposition: The transition from *show* to *tell*

This, then, is our hypothesis about how and why language began. As many others in this volume report, all the essential cognitive components already seem to have been in place in other hominins 250,000 years ago—just as many of them are also present in our contemporary ape cousins as well as in other intelligent social species today: the capacity to learn the features distinguishing categories by sensorimotor induction; the capacity to learn by observation and imitation; the capacity for pointing, shared attention, and mind-reading; a strong tendency for kin cooperation in the rearing of late-developing young; and the possession of both "canonical" or "affordance neurons" (which recognize what you can do with what) and "mirror neurons" (which recognize when another organism is doing the same kind of thing you are doing).

Many of these cognitive components (and probably others too) were already available before language began (though some of them may also be effects rather than causes of language). It could well be that our ancestors had the power of communication by pantomime before the advent of language; but *you can't convey new categories by pantomime alone.* All you can do is mime objects, events, and actions, and, to a certain extent, properties and states (including requests and threats).

5. Combining and communicating categories

Was the power of language invented or discovered? The power of language (according to our hypothesis) was, in the first instance, *the power of acquiring a new composite category from other people who already knew the old categories out of which the new one was composed*. The simplest combination is conjunction: a member of C is a member of both A and B. The first combinatory categories were probably "extensional" categories, which means that members of C are learned to be those members of A that are also members of B. But since the members of A are detected by our brain's learned feature-detectors for A and the members of B are detected by our feature-detectors for B, the members of C can be detected by simply combining the feature-detectors for A and for B. This could be detected passively, through observational learning, as in our simulations, with the learner simply noting that it is faster to learn what a C is from watching the actions of those who already know what a C is—noting that they are doing what you need to do with a C with whatever happens to be both an A and a B. (This only works, of course, if you already know what an A and a B are, and if the rule for C is just conjunction.)

But since hominins were far more social creatures than our simulated mushroom foragers, the learning need not have been so passive. Kin who already knew A, B, and C—for example the learner's mother—would be motivated to *teach* the learner C. And since A, B, and C are overt actions (unlike in our simulation, where they were just vocalizations), the mother would be motivated to draw the category combinations to the learner's attention not only by performing the three actions on the objects themselves but also by miming them in the absence of the objects.

We suggest that this is how the *proposition* was born. Learners may have begun picking up new composite categories through passive observational learning. But once the knower—motivated to help kin or collaborators learn by sharing categories—became actively involved in the communication of the composite categories, it became intentional rather than incidental instruction, with the teacher actively miming the new category descriptions to the learner. (This motivation and strategy would be a natural extension of existing strategies to execute skill so that others can imitate it.) The category "names" would also become shorter and faster—less iconic and more arbitrary—once their propositional intent (the "propositional attitude") was being construed and adopted mutually by both the teacher and the learner.

It is interesting to watch contemporary apes when researchers are trying to teach them language. It seems indisputable that they never learn language; —otherwise, despite the excuses she gave for them, Sue Savage-Rumbaugh's chimps would have been able to attend the Summer Institute to discuss the origin of language with the rest of us, just as any newly discovered contemporary hunter-gatherer from the depths of the Amazon jungle could join the "conversation" (that trendy and overused word, but apt enough here!).

Chimps have categories. We keep training them to "name" their categories (whether with gestures, symbolic objects, computer keyboards, or words) — even to combine those names into quasi-propositional strings. And the chimps oblige us if they are hungry or they feel like it. But what is striking is that they never really pick up the linguistic ball and run with it. They just don't seem to be *motivated* to do so, even if they sometimes seem to "get it," locally, for individual cases.

5.1. The disposition to propose: Intelligence or motivation?

We are not sure whether chimps really do get it. They get the associations and the contingencies; but do they get them as *propositions*, with truth values? It's hard to say. What's most puzzling is why they don't seem to pick up on the power that is within their hands when they start systematically combining symbols to define new categories. What seems missing is not the intelligence but the motivation, indeed the compulsion, to name and describe.

(Perhaps something like these trained chimps' unsystematic, local, instrumental signaling is what the proponents of "protolanguage" have in mind—but if so, it fails our test of being able to say anything and everything precisely because *it is not really "saying" anything at all*: We (with our language-biased brains) are merely projecting that propositional gloss onto it, like compulsively adding narrative subtitles to a silent movie.)

To construe *A is B* as a proposition—as asserting something true about the world, rather than just producing a string of associations for instrumental purposes—may have required a new cognitive capacity, perhaps even one with a genetic basis. But given how smart hominins were already, and how all the pieces were already in play, such a "mutation" to adopt the "propositional attitude"—if a downright mutation was indeed needed at all—is not far-fetched to imagine. (It does not call for an additional mutation for UG, for example.)

So maybe that's what evolution did with our species. Because we were more social, more cooperative and collaborative, more kin-dependent—and not necessarily because we were that much smarter—some of us discovered the power of acquiring categories by instruction instead of just induction, first passively, by chance, without the help of any genetic predisposition. But then, as those who happened to be more motivated to learn to acquire and share categories in this all-powerful new way began to profit from the considerable advantages it conferred—the benefits of acquiring new categories without the time, energy, uncertainty, and risk of acquiring them through direct sensorimotor induction—Baldwinian evolution (Simpson 1953) began to favor this *disposition to learn* and to use symbols to name categories and to recombine their names in order to predicate and propositionalize about further categories, because of the adaptive benefits of category description and sharing. The tendency to acquire and convey categories by instruction thus grew stronger and stronger in the genomes and brains of the offspring of those who were more motivated and disposed to learn to do so. And that became our species' "language-biased" brain.

As noted, the capacity or disposition to construe *an A is a B* as a proposition, with a truth value, may either have been a mutation, or primates may already have had the capacity and only needed to have it amplified motivationally and behaviorally by Baldwinian evolution. The point is that *the full power of language already comes with the propositional territory*, once you begin systematically naming your categories and recombining their names into propositions describing or defining new composite categories. The voyage begins with the benefits of acquiring the immediate subsistence and survival categories of our Paleolithic forebears through instruction instead of induction, but it leads seamlessly to the contents of our contemporary (and future) mental lexicon of categories, as reflected in our oral tradition, and eventually our written tradition, with its glossaries, dictionaries, encyclopedias, manuals, and textbooks, and now the Web (Dror & Harnad 2009).

5.2. From hand to mouth

Now the likelihood of first discovering this boundless propositional power is incomparably higher in the gestural/pantomime modality, which already includes all the nonarbitrary things that we *do* with (concrete) categories; those acts can then be short-circuited to serve as the categories' increasingly arbitrary names, ready for recombination to define new categories.

But once a species picked up the linguistic ball in the gestural modality and began to run with it (linguistically, propositionally), the advantages of freeing its hands to *carry* something other than the ball, so to speak (rather than having to gesticulate while trying to do everything else)—along with the advantages of naming when the teacher was out of the learner's line of sight, or at a distance, or in the dark—would quickly supersede the advantages of gesture as the start-up modality for language (Steklis & Harnad 1976). (One could almost hear the grunts of frustration if the intended learner failed to see, or the teacher's hands were otherwise occupied: the "yo-he-ho" and "pooh-pooh" theories of the origin of language parodied by Max Mueller come to mind.)

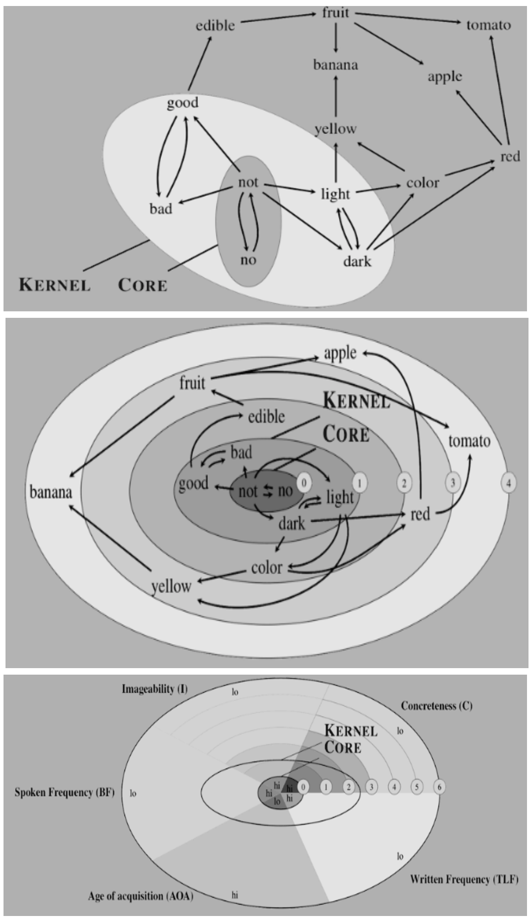
Besides, once the name has become arbitrary, its "shape"—and hence its sensory modality—no longer matters. So as the power of language in category acquisition became encephalized, the tendency to assign arbitrary names to categories—and to combine their names into propositions defining new categories—migrated to the auditory modality, which was already so admirably prepared for the task in other species as well as our own (though no doubt it too had to undergo a period of intensive further encephalization, under selective pressure from the adaptive advantages of increasingly efficient speech).[iii]

Notice that there is nothing that is really like a "protolanguage" in any of this—unless you consider that every time a language lexicalizes a new category with a word, or its syntax is modified, it has thereby become a new protolanguage. A language is just a set of symbols with which we can say anything and everything, whether in gestures or in speech, whether quickly or slowly, and whether with a vocabulary of many symbols or few.

5.3 Symbol grounding

Now a question: *How much induction do you have to do before language can kick in, in the form of instruction*? (Clearly our toy simulation with the three mushroom categories does not answer this question.) Surprisingly, a hint at an answer may be latent in our contemporary dictionaries, and graph theory may be able to help us dig it out. By first eliminating all the words that are only defined, but that are not used to define any further words, and then also going on to eliminate successively all the other words that can be "reached" by definition out of the remaining words, we have reduced dictionaries to their "grounding kernel" (Blondin Massé & Harnad 2010) (Figure 2). The kernel words are those out of which all the rest of the words in the dictionary can be defined, but within which all further definition is circular, because the kernel words interdefine among themselves (rather like *good* = *not bad* and *bad* = *not good*). The kernel turns out to amount to about 10% of the dictionary. It also turns out that the words in the kernel are learned significantly earlier than the rest of the words in the dictionary, as well as being more concrete, imageable, and frequent, both orally and in writing (Vincent-Lamarre et al. 2016). But the kernel is not yet the smallest number of words out of which all the other words in the dictionary can be defined.

Word	**Definition**	**Word**	**Definition**
apple	red fruit	*good*	not bad
bad	not good	*light*	not dark
banana	yellow fruit	*no*	not
color	light dark	*not*	no
dark	not light	*red*	dark color
edible	good	*tomato*	red fruit
fruit	edible	*yellow*	light color



(a)

(b)

(c)

Figure 2. Extracting a Dictionary's Grounding Kernel and Core. The table is a 14-word toy dictionary (*not a real dictionary*) artificially constructed to illustrate how a simple recursive algorithm can reduce any dictionary to its grounding *Kernel* (larger ellipse in Figure 2a) from which all the rest of the words can be reached by definition alone (arrows). The smaller ellipse is the *Core* (trivial in this toy example), a large, strongly connected component within the Kernel that turns out to emerge empirically in real dictionaries. Figure 2b shows how words are at increasing definitional distance from the Core in the toy dictionary. In real dictionaries (bottom left figure), words in the Kernel turn out to have been learned at a younger age, and are more concrete, imageable

> and frequent than the rest of the dictionary. The correlation is all-or-none between the Core and the rest of the dictionary for age and frequency, whereas it is graded with the degree of definitional distance from the Core for concreteness and imageability. This means there are words inside the Kernel but outside the Core that are more abstract and are learnt throughout the life cycle. The Kernel is unique, but *it is not the smallest number of words from which all the rest can be reached by definition alone*: That smallest set of words is the *minimal grounding set* (MinSet), which is not unique, and we are beginning to extract it in real dictionaries (Vincent-Lamarre et al 2016). The strongly connected Core is *not a MinSet*. Finding the MinSet is equivalent to the graph-theoretic problem of finding a *minimum feedback vertex set*, which is an NP-complete problem for graphs in general, but there are ways to solve it in the special case of dictionaries, because of their small size and their structure.

We have computed the smallest defining set or MinSet for four dictionaries (Vincent-Lamarre et al 2016). Unlike the kernel, the MinSet is not unique. There are many alternative MinSets of the same smallest size. For the four dictionaries analyzed that size varied from about 400 words to about 1400 words. This means that with fewer than 1500 content words (category names) plus a few function words, such as *and, not, if,* etc., *plus the all-important power of predication*[iv] we can define all the other words contained in our contemporary dictionaries (as well as those in all our encyclopedias and all textbook glossaries) and hence all the words in our mental lexicon.

This has some of the flavor of our "test" for whether something is a natural language: that you can express any and every proposition in it. Whatever that turns out to be, it seems as if fewer than 1500 category names plus a few functors and syntactic rules is as much as you need in order to get language's full expressive power.

6. Keeping our feet on the ground

Despite its dramatic benefits, however, learning categories through language, by symbolic instruction, is not completely equivalent to learning them by sensorimotor induction. We also have some behavioral and electrophysiological evidence that, although the underlying brain processes of both ways of acquiring categories may have much in common, the results of category learning by induction vs. instruction are not identical (St-Louis & Harnad 2010). Instruction requires far fewer training trials (in fact, a good instruction should allow correct categorization from the very first attempt), but it does take longer to apply the instruction to each instance encountered, at least initially; the reaction time for categorization when it has been learned by induction is faster than when it has been learned by instruction. And of course it cannot be instruction all the way down: instruction presupposes already having all the categories used to *define* the new category. Even if, at bottom, and in principle, only 1500 categories would need to be learned directly through sensorimotor induction, in practice it is clear that we continue learning new categories in both these ways throughout our life cycle. Moreover, whereas a categorization rule can be learned purely symbolically, applying it to actual instances is still sensorimotor, except when the instances are only named or described verbally.

So "tell" has superseded "show" in human evolution but (as with occasional PowerPoints) it still helps to support, illustrate, supplement, or refresh the telling with some showing. After all, direct sensorimotor experience -- rather than just indirect verbal hearsay -- is, at bottom, still what living is all about, even for *Homo loquens*.

References

Blondin Massé, A., Chicoisne, G., Gargouri, Y., Harnad, S., Picard, O., & Marcotte, O. (2008). How is meaning grounded in dictionary definitions? *TextGraphs-3 Workshop at the 22nd International Conference on Computational Linguistics.* http://www.archipel.uqam.ca/657/

Blondin Massé, A., & Harnad, S. (2010). *Symbol grounding and language origins* (supp.). Cognitive Sciences Institute, Université du Québec à Montréal, June 2010. http://users.ecs.soton.ac.uk/harnad/Temp/blondin-langconf.pdf

Borges, J. L. (1964). Funes the memorious. Trans. James E. Irby. Labyrinths: Selected Stories and Other Writings. Ed. Donald A. Yates and James E. Irby. New York: New Directions, 59-66. http://users.clas.ufl.edu/burt/spaceshotsairheads/borges-funes.pdf

Cangelosi, A., & Harnad, S. (2001). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication, 4,* 117–142. http://cogprints.org/2036/

Cangelosi, A., Greco, A., & Harnad, S. (2002). Symbol grounding and the symbolic theft hypothesis. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 191–210). London: Springer. http://cogprints.org/2132/

Chomsky, N. (1976) On the nature of language. In S. R. Harnad and H. D. Steklis and J. Lancaster (Eds.), Origins and Evolution of Language and Speech. *Annals of the New York Academy of Sciences*, Volume 280, pages 46--57.

Dror, I. and Harnad, S. (2009) Offloading Cognition onto Cognitive Technology. In Dror, I. and Harnad, S. (Eds) (2009): *Cognition Distributed: How Cognitive Technology Extends Our Minds*. Amsterdam: John Benjamins http://eprints.ecs.soton.ac.uk/16602/

Gibson, J. J. (1979). *An ecological approach to visual perception.* Boston, MA: Houghton Mifflin.

Harnad, S. (1976). Induction, evolution and accountability. In S. Harnad, H. D. Steklis, & J. B. Lancaster (Eds.), *Origins and evolution of language and speech* (pp. 58–60). New York: New York Academy of Sciences. http://cogprints.org/0863

Harnad, S. (1990). The symbol grounding problem. *Physica D, 42,* 335–346. http://cogprints.org/0615/

Harnad, S. (1991). Post-Gutenberg galaxy: The fourth revolution in the means of production of knowledge. *Public-Access Computer Systems Review, 2,* 39–53. http://cogprints.org/1580/

Harnad, S. (2000). From sensorimotor praxis and pantomime to symbolic representations. *The Evolution of Language. Proceedings of 3rd International Conference,* Paris April 3–6, 2000 (pp. 118–125). http://cogprints.org/1619/

Harnad, S. (2003) Back to the oral tradition through skywriting at the speed of thought. *Interdisciplines*. http://eprints.ecs.soton.ac.uk/7723/

Harnad, S. (2005). To cognize is to categorize: Cognition is categorization. In C. Lefebvre & H. Cohen (Eds.), *Handbook of categorization* (pp. 19–43). Amsterdam: Elsevier. http://eprints.ecs.soton.ac.uk/11725/

Harnad, S. (2008). Why and how the problem of the evolution of Universal Grammar (UG) is

hard. *Behavioral and Brain Sciences, 31,* 524–525. http://eprints.ecs.soton.ac.uk/15618/
Hauser, M.D., Chomsky, N & Fitch, W.T. (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569-79
Harnad, S (2014) Chomsky's Universe. L'Univers de Chomsky. *À babord: Revue sociale es politique* 52. https://www.ababord.org/L-univers-de-Chomsky
Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology, 14,* 5–24.
Hockett, CF & Ascher, R (1964) The Human Revolution. *Current Anthropology* 5, 135-168 http://www.jstor.org/stable/27838917
Katz, J. J. (1976). The effability hypothesis. In S. Harnad, H. D. Steklis, & J. B. Lancaster (Eds.), *Origins and evolution of language and speech* (pp. 445–455). New York: New York Academy of Sciences.
Koster, Jan (1986) *Domains and dynasties: the radical autonomy of syntax*. Dordrecht: Foris
Picard, O., Blondin Massé, A., & Harnad, S. (2010). *Learning word meaning from dictionary definitions: Sensorimotor induction precedes verbal instruction.* Summer Institute on the Origins of Language, Cognitive Sciences Institute, Université du Québec à Montréal, June 2010. http://users.ecs.soton.ac.uk/harnad/Temp/picard-langconf.pdf
Poynder, R., & Harnad, S. (2007). From glottogenesis to the category commons. *The Basement Interviews.* http://ia361300.us.archive.org/13/items/The_Basement_Interviews/Stevan_Harnad_Interview.pdf
Simpson, G. G. (1953). The baldwin effect. *Evolution*, 7(2), 110-117.
Steklis, H. D., & Harnad, S (1976). From hand to mouth: Some critical stages in the evolution of language In S. Harnad, H. D. Steklis, & J. B. Lancaster (Eds.), *Origins and evolution of language and speech* (pp. 445–455). New York: New York Academy of Sciences. http://cogprints.org/0866/
St-Louis, B., & Harnad, S. (2010). *Learning categories by sensorimotor induction and verbal instruction.* Paper presented at the Summer Institute on the Origins of Language, Cognitive Sciences Institute, Université du Québec à Montréal, June 2010. http://users.ecs.soton.ac.uk/harnad/Temp/stlouis-langconf.pdf
Vincent-Lamarre, Philippe., Blondin Massé, Alexandre, Lopes, Marcus, Lord, Mèlanie, Marcotte, Odile, & Harnad, Stevan (2016). *The Latent Structure of Dictionaries* TopiCS in Cognitive Science (in press) http://eprints.soton.ac.uk/370845/

Notes

---

[i] We learned from several presentations at the Summer Institute that the widespread impression is *incorrect* that the question of language origins was neglected for a century because the *Société de linguistique de Paris* had banned the topic as too speculative. The Société apparently banned the topic for more parochial reasons, having to do with territorial rivalry with the anthropology society. The century of neglect, however, was real, and largely attributed to the speculative nature of the question. The articles by Hockett (1964) and Hewes (1973) in *Current Anthropology*, followed by the NYAS Conference in 1976 probably marked the global reopening of the topic.

[ii] Note that this is not merely “observational learning” of sensorimotor skills; it is *category learning*. Nontrivial categories cannot be learned by mere imitation, because even if you try imitating what those who know the category are doing, when they are no longer there, you cannot do what they did unless you have learned the features distinguishing the members from the nonmembers. To learn that, your only options are (1) to acquire the rule directly, through direct sensorimotor induction, guided by error-corrective feedback; or (2) to acquire the rule verbally, through symbolic instruction—*on condition that you already know the categories that describe the compositional rule*. Since the mushroom foragers already knew A and B by induction, they could learn C = A + B by passive observation (“hearsay”), rather than having to learn it bottom-up, as they had learned A and B.

[iii] We think Fitch (this volume) is wrong about animal vocalization being a direct precursor of language. (This throwaway hypothesis of Darwin’s arose from the absence of a genetic mechanism for evolution at the time, as well as from the superficial similarity between species evolution and linguistic “evolution”—the same superficial similarity that has led some of the other contributors to this volume to conflate historical change in language, which is purely learned and cultural, with the real biological problem of language origins, which is evolutionary change in the genetic and neural substrates of language.) The fact that birdsong or whale-song has what looks like “syntactic” structure does not make it linguistic; and, as we have tried to argue, what is missing is not syntax (nor even UG), but *propositionality*. Animal vocalizations are largely courtship rituals, and as such they are more of a piece with visual courtship displays than with anything to do with language; birds and whales and insects court vocally instead of just visually, because it is hard to see in the trees or the ocean depths or the dark. And it is largely the males who do the courting, making it something of a one-way street, communicatively speaking. Animal vocalization may have been a precursor of music, but not of language. The “meaningfulness” and expressivity of music is neither linguistic meaningfulness nor linguistic expressivity—otherwise, we could have written this paper in F# minor instead of English. Vocal expressiveness survives today, and is even coupled with language, but in the form of prosody and intonation, which, like onomatopoeia, are merely the nonarbitrary vocal accompaniment, not the arbitrary amodal carrier, of the symbols and their propositional content (whether gestured, spoken, or written)—its background music, not its message. It is not that nonlinguistic signals and communication are not still important for our species, just that they are indeed not linguistic.

[iv] In some languages predication is marked by “is”, as in “apple is red,” but in some languages “is” is omitted, as in “apple red” (subject/predicate), stated while adopting the “propositional attitude.”