Journal of Logic, Language and Information (2005) 14: 235–251

What Is Context *For*?
Syntax in a Non-Abstract World

TOM SGOUROS
15 Boston Neck Road, Wickford, RI 02852, U.S.A.
E-mail: tom@sgouros.com

(Received 30 March 2004; in final form 21 July 2004)

Abstract. An explanation for the uncertain progress of formalist linguistics is sought in an examination of the concept of syntax. The idea of analyzing language formally was made possible by developments in 20th century logic. It has been pointed out by many that the analogy between natural language and a formal system may be imperfect, but the objection made here is that the very concept of syntax, when applied to any non-abstract system of communication, is flawed as it is commonly used. Syntax is properly defined with respect to an individual transformation rule that might be applied to some message. Collections of syntax rules, however, are inevitably due to categories imposed by an observer, and do not correspond to functional features found in non-abstract systems. As such, these categories should not be relied upon as aids to understanding any natural system.

Key words: information, semantics, Shannon, syntax

1. Introduction

The formal approach to linguistic analysis has a long history. Since the nineteenth century speculations of Gottlob Frege, successive generations of logicians and linguists have grappled — and continue to grapple — with the formal aspects of language. But what are the results? Some retain high hopes for the research program, but to say that the progress has been desultory and tentative to date would be only to agree with, for example, Chomsky:

> This is, of course, a program, and it is far from a finished product. The conclusions tentatively reached are unlikely to stand in their present form; and, needless to say, one can have no certainty that the whole approach is on the right track. (Chomsky, 2000, p. 8)

Chomsky, of course, remains encouraged about the prospects for further research and proud of the research this approach has inspired (Chomsky, 2000). But these are two distinct things. It is not inconsistent to applaud the latter and harbor misgivings about the former. Niels Bohr's model of the atom provided a similarly inspiring research program to many physicists. But the theory itself did not survive the advances it triggered.

What if, despite the confidence expressed by formalists, it turns out that human language is not modeled well by the tools of any formalist school of linguistics? After all, after decades of effort, formalists cannot produce the one argument that makes an existence proof irrefutable. That is, they can show data to explain why a universal grammar must exist, and what it must look like, and even intuit rigorously formal rules such a grammar must follow, but they cannot produce a formal grammar with any convincing claim to completeness for any human language. It is sometimes argued that this is not the goal. But there seems to be no logical obstacle to that goal within the formalist framework. This article is simply a speculation about what follows if we treat the shortcomings of formal grammar as data points rather than as obstacles to be overcome by yet more subtle formulations of the original ideas.

2. Separating Syntax and Semantics

The formal separation of syntax and semantics is largely a product of 20th century investigations of formal systems. The basic idea, of course, is ancient. Nonetheless, Gottlob Frege made an important step in sharpening the distinction in his work on proof theory, by demanding that the rules of inference in an axiomatic system be declared as explicitly as the axioms themselves. Following him, Jan Lukasiewicz, in a paper published in 1920, pointed out that in a formal system a proof need never refer to the meanings of the axioms, but can proceed by manipulation of the symbols alone. It is adequate then, in such a system, to ignore completely the semantics of the statements in question, attending only to their syntax (Bocheński, 1961, pp. 38–43).

In the 1930s, Rudolf Carnap attempted to apply rigorously what had been learned about formal logical systems to the syntax of natural language, as opposed to the meaning of natural language, which is where the earlier linguistic queries of logicians had led (Carnap, 1937). Carnap's influence leaked into the world of linguistics, at least in part via the work of Yehoshua Bar-Hillel (1953, 1954). By this point, the formalist approach to linguistics was already established by the contributions of Leonard Bloomfield, but Carnap's idea that syntactic rules could account for a great deal of what had been considered unarguably "meaning" was a powerful push toward the generative grammar project.*

By now, the formal approaches to the analysis of natural languages have split into several further categories. The Bloomfield analysis of grammatical structures gave way to Chomsky's formulation that the underlying structure of language is the proper subject of formal analysis. Chomsky's original theory has undergone several refinements, and some wholesale revisions (Chomsky,

*The introduction to Bar-Hillel (1964) contains an interesting historical account of some of these influences, as they were manifested in Bar-Hillel's own career.

1995), and has spawned several other variants, including model-theoretic syntax (Rogers, 1999), Lexical-Functional Grammar (Bresnan, 1982), Generalized Phrase Structure Grammar (Pollard and Sag, 1994), and several more. In addition to these, Montague reached back to Carnap for inspiration, creating a separate strain of formal logical grammars (Montague, 1970a, 1970b). (See Graffi (2001), for an aerial view.)

In 1948, apparently independently of all these linguistic and meta-logical currents, Claude Shannon presented his information theory, proposing to consider only the form of a message — ignoring its meaning — in coming up with a measure of its information content (Shannon, 1948).* He presented no justification for doing so, identifying it as a premise of the theory in his opening paragraphs. Whatever the source of the idea, it was an idea that fit well with the formalization of syntax in linguistics, and it became one of the intellectual cornerstones of computational studies of language and mind (Wiener, 1961; Cherry, 1978; Dupuy, 2001).

Shannon's work, however, contains more relevance to linguistics than this. After all, he was a communication engineer, professionally concerned not so much with transmitting abstract information but with real messages in the real world. Any linguist concerned with the psychological reality of their theorizing is as well. After all, a spoken utterance is as much a message between a transmitter and a receiver as any radio signal travelling through the air. The fact that Shannon pointed out the horizons that open up when you consider information as an abstraction should not blind us to the fact that his work contains other important ideas about constructing systems to be used in the real world. If understood the right way, these ideas can shed some light on the plausibility — and the intractability — of formal grammar.

3. Correcting Errors

A subject of great importance to engineers designing real-world implementations of communication systems is error correction. The original purpose of Shannon's paper, after all, was to establish criteria for evaluating the possible accuracy of a given communication system. Since no communication system is perfect, a certain amount of redundancy is the way to ensure that a message is delivered accurately. Shannon created a theoretical framework for estimating how much redundancy is necessary for perfect transmission of a message given some imperfect communication channel.

Among communication engineers, there are some standard ways to add redundancy to a message. The common, though by now slightly archaic, RS-232 standard includes its own error-correction system, where each 7- or 8-bit byte can contain an extra "parity" bit to indicate whether there are an even or odd number of ones in the

*Hartley (1928) had earlier presented a similar measure of information, but had been somewhat less explicit about the distinction between form and content. Hartley also cited no antecedents for the idea.

given byte. If you receive a byte with the parity bit set incorrectly, you can infer that one or more of its bits has been misread, and ask for that byte to be retransmitted.

Another standard way to add redundancy to a message is to include a "checksum" for each block of bytes transmitted. A checksum is the result of some calculation (often just a sum, but occasionally something more complex) using all the bytes that need checking. When a block of bytes is received (the size is prearranged), you duplicate the original calculation and compare the result with the included checksum. If they don't match, you know there has been a transmission error somewhere. You may not know exactly which byte is wrong, but knowing that it is one of the bytes contributing to the checksum is often enough. And if you're lucky enough to determine which byte is wrong by some other means, the checksum, along with all the other received bytes, can be used to recreate the missing data.

But look what is happening. An error — say a timing error that caused some bits to be misread — must be corrected using some data in addition to the received message. The parity-checker needs the parity bits, and also needs to know whether it's expecting even or odd parity, and what that means. The checksum-checker needs the checksum, as well as instructions for recalculating it. That is, both of them need the correction data and need to know what to do with it. The correction can only happen when three elements are present: the original data, the correction data, and the correction instructions.

Shannon wrote about information like checksums as travelling over a "correction channel," an independent communication channel carrying information redundant to the message and used to reduce errors (Shannon, 1948, p. 398ff). The receiver receives a message, then uses the correction information to correct it. Shannon further points out that in practice the checksum might be transmitted alongside the message, or it might not. Either way, it's a distinct bit of data, related to the message, but independent of it.

Departing only slightly from Shannon's presentation, we can think of this as a two-level operation. The first level takes some input (fluctuating voltages, say) and produces some output (a series of bytes). The second level takes the output of the first level, and the information from the correction channel, and produces some other output (a corrected series of bytes or an error signal if the correction is not possible). This is an appealing abstraction, since we can then imagine a layering of more levels, which would be useful in a case where any level may not have enough information to correct all the errors detected.* Maybe there is a third level, whose purpose is to take the output from the second level, and use "metadata" — descriptions of the message data — to unpack the message into characters or numbers. If the data are

*The levels in our example may lend themselves to further decomposition to distinct levels. For example, the RS-232 standard itself may be viewed as two nested syntaxes: one specifying the voltage levels for ones and zeros and another specifying the timing. That these two are fully independent can be readily seen by the fact that other serial communication standards, such as RS-485 and RS-422 use the timing syntax of RS-232, but have a different definition of what constitutes a one or a zero. (In fact, those two standards don't use voltages at all.)

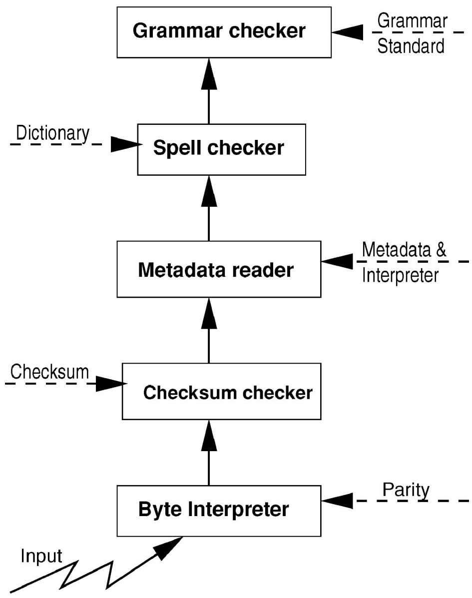

Figure 1. A sketch of a possible arrangement of processing levels.

characters, perhaps the next level up uses a spell-checker, whose input is the original text (and maybe the error signal from the second level) and a dictionary. There is a diagram of an arrangement like this in Figure 1.

Of course spell-checkers aren't perfect, generally unable to tell the difference between a properly spelled but improperly used word and a properly spelled, properly used one. It is not impossible for there to be errors that pass both the checksum-checker and the spell-checker, such as substituting the word "debt" for "daft." So we could imagine another level that would check on the previous level's performance by checking the grammar of the corrected message, and as many more layers of transformation and error-checking as we need.

It isn't too hard to see where this approach leads: to a sequence of processing levels, where one level is defined to be "higher" than another if the output from the latter is available to the former. A higher level gets its input from a lower level, but also requires another source of information. This other source corresponds to Shannon's correction channel, and it may have come with the message, or from some other source. Also note that there is no principled difference between the error-checking levels, such as the checksum-checker, and other sorts of transformations, such as the metadata interpreter. Some error-checking is done by transformations, such as calculation of checksums, or grammar-checking, while other forms only transform the message when an error is signalled.

In Shannon's early discussion of the correction channel, he implicitly acknowledges the different levels of syntax and the relation between the level of analysis and the level of the error. However, he did not make the relation explicit and then lost the distinction between levels in his discussion when he pointed out that the original channel can carry the correction information along with the message. Still, the point was made that, though the message and the correction information cannot be completely independent, they are distinct entities.

What would happen if we chose to call the information in the correction channel the "context" in which the message is to be interpreted? Similarly, the process of successful transformation could be called "understanding" the message, and the transformed message its "meaning." Using these words, we might make some interesting statements about our example. First, as we go up the levels, the "context" required for "understanding" tends toward the richer. At the lowest level, the context is just an answer to the question, "Even or odd?" Moving up, the context becomes a number, then a metadata declaration, then a dictionary and a grammar standard. Second, when the transmission is accurate, the additional levels of context don't add much. But when the transmission is not good — or the receiver unreliable — it may be that all the levels are necessary, and that the message isn't correctly understood until after the grammar-checking layer. Third, you could say that a message can't be understood correctly without its context. Fourth, an unmistakable sign of having understood something is to be able to say whether its meaning is correct or not. And so on. This use of these three words in the case of machine communication mirrors our use of the words with respect to communication between people.

4. The Real World

Considerations of syntactic levels are germane only to real communication in a real world. In an abstract world, we can assume away all communication errors. Less facetiously, in a formal world, symbols need not be instantiated by anything in particular, they are just symbols. But in a computer, for example, a bit is represented by a voltage, measured in specific places, a byte by a particular arrangement of several bits, a number or letter by a specific encoding implemented in bytes, and so on. Ignoring these levels of representation in turn allows us to disregard the various levels at which a message exists, and to imagine a message appearing *ex nihilo* at the receiver as a sequence of letters or numbers and some error-correction information. In the examples in his paper, Shannon used just such abstract messages of letters to consider higher levels, looking at the structure of words and sentences and paragraphs. His purpose was to estimate possible data rates for perfect message transmission, not to consider cases of imperfect correction, and so in his treatment of communication, he neglected the levels beneath the letters, an oversight that others have perpetuated (Brillouin, 1962; Wiener, 1961).*

*Processing levels like those described here are an important part of modern computer design. Important internet design documents, for example, specify a number of processing levels on which programs are supposed to work. The idea is that a program operating at level 4 can safely ignore all the details of levels 1–3. Useful discussions of these ISO/OSI networking model layers can be found in

Perhaps the single biggest difference between an abstract formal language and a natural language is this: that analysis of an abstract language can begin at some arbitrarily chosen level, while analysis of a natural language, used in a real world, cannot. Words and letters are not transmitted from one person to another. An unbroken stream of sense data is what we get; words and letters are what we make of that stream. This is as true of computers as it is of people, even if the sense data is not nearly as rich. Here, then, is a way to accept a materialist view of how the brain works while rejecting the facile comparision between real-world human brains and abstract symbol-manipulating computers: it turns out that computers don't really work that way, either.

In a view such as the one presented here, what constitutes "the" syntax of a message from one computer to another? The answer is that the question as stated makes no sense. There are only the individual rules used to transform a message from one state to another. These transformations can be subtle or profound, ranging from an error-correcting system that might correct a single byte or none at all, to an RS-232 module transforming a modulated series of voltages into binary numbers between 0 and 255. Of the message itself, one can say that there are aspects of its form that allow successful transformation by some particular rule. But before and after the operation of that rule, the relevant definitions of the structure and meaning of the message will be different, depending on what kind of transformation is under consideration.

Because they vary so wildly, trying to shoehorn all these different forms of syntax into a single category is not the best route to understanding the system. Obviously there is a category we can call "the" syntax of a message, containing the collection of rules we've named. But we can also create a category of animals whose names have an 'R' in them. The important consideration isn't whether the category exists, but whether this is a useful category that implies something real about the world in which we observe its members. The argument here is that, given the real-world need for context to correct the errors of each individual operation, it is not necessarily useful to consider categories of syntactic operations to be a single functional whole.

In our example, the syntax of the message, as seen from the fourth (spell-checking) level, is that it comes in letters combined to form words. From that same point of view, the semantic content of the message is whatever is left after any wrong words have been corrected. But seen from the first level, the syntax has to do with voltages and timing, and the meaning is the series of bytes passed to the second level.

IEC (1994) or at the beginning of Stevens (1999). But as with the other domains under discussion here, the standard outlined by this scheme is almost always honored in the breach. That is, virtually no real computer programs rigorously respect the boundaries between the layers. Common transgressions of the layer schema can be readily found in both web page and email transport protocols, for example. (See Fielding et al. (1999) on HTTP content negotiation or authentication, or Resnick (2001) on uninterpreted SMTP extension header fields.) Again, a close look at computers in the real world shows them not to be the idealized abstract symbol manipulators of song and story.

A particular syntactic rule could be applied to an entire message, to one or more portions of a message, or to multiple copies of a message. It could be applied to the same message under different transformations, or it could be applied recursively. In other words, there is a tremendous variety of potential ways for simple rules to be applied. In the small-scale view of each rule, the definition of syntax might be clear, but with a minimally intricate arrangement of rules, there might be no useful way to define "the" syntax of a message. If we regard the receiver of a message as unitary, perhaps this is a distinction without a difference: it's all syntax. But if we regard the receiver as an assemblage of more-or-less independent sub-systems interacting with one another, then the fact that the view from each one is different may be very significant to the change in state created in the receiver by the message.

As with the concept of syntax generally, so it will be with the large categories of syntactic rules. It is tempting to look at a drawing of the different transformations involved in processing a message and assign large groups of them to distinct conceptual classifications. The layer model of internet communication is a perfect example of such an attempt (see footnote on p. 240). Categorization is useful, and often a route to further understanding, but it is important to remember both that these classifications may not represent a functional reality, and that in a real communication program, a set of high-level transformations are unlikely to get *all* their input from the next set down the ladder. That is, it is potentially quite misleading to imagine that these broad categories represent sequential processors, with each operating on the output of the next one down the hierarchy. Each member of some category may behave like that, but the categories themselves do not. They are only classifications, externally imposed by observers on large sets of vaguely similar operations, and cannot be understood to be functional units themselves.*

5. Interpreting Speech

Jackendoff (2002) provides a persuasive theoretical framework for a hierarchical arrangement of syntactic levels in our internal processing of human speech. According to his analysis, which he dubs the "parallel architecture," a (spoken) message is analyzed first at the phonological level, which is split into the (separate,

*As this article turns from machine syntax to a discussion of natural language, it's worth an aside here to note that the "rules" and "transformations" discussed here need not be strictly identified as transformational grammar rules. Adherents to transformational grammars may view them as grammatical transformations. Linguists who cleave to the view that grammar sets up constraints to be satisfied, or that grammatical rules are satisfied in a model-theoretic sense, may view them as whatever transformations are necessary to turn the acoustic properties of some word into an understanding of that word. The argument here concerns the character of the rules and how context is an essential feature to their operation. I make no claim as to the exact nature of what the rules do, only about how they do it. That is, the rules described in this paper are mechanical in nature, but they may be more basic than noun phrase construction rules, for example, or less basic than the sorts of operations envisioned by garden-variety connectionist schemes, most of which employ a dramatically simplified neuron model.

though closely interdependent) prosodic, syllabic, segmental, and morphophonological tiers, and then at the syntactic level, which can be considered another set of tiers, comparable to — and independent of — the first four. His analysis continues upward, to include semantic rules as yet another level.

This is an attractive argument, and not very distant from the view outlined here, with important exceptions. For one thing, it is not clear why one should consider the levels Jackendoff identifies to be as distinct from one another as he does. Each tier he identifies represents a set of operations, many of which operate in parallel with one another, and some of which operate in parallel with operations in the tiers above and below. But with each tier representing many different operations happening at more or less the same time, but overlapping in time with operations at the tiers above and below, it is not clear how much physical reality is invested in the definition of these tiers.

The relevant unit — if relevance is to be defined with respect to the actual functioning of a brain listening to speech — would seem to be the individual operations that instantiate some specific rule or part of a rule, rather than the groups of such rules that make up the tiers or levels defined by Jackendoff. As in the computer communication example, these tiers seem to be the product of categories imposed by the observer rather than functional groupings inherent in the processing of speech. Consider how one might define a functional group. Operations that belong in the same group might be ones conducted at the same time, at the same neural location, who receive the same input, or whose output is combined. By Jackendoff's own description of how his parallel architecture envisions the operation of a brain, none of these conditions is necessarily true for two rules occupying the same tier.

Jackendoff does allow that the phenomena of great interest occur at the interfaces of the tiers he defines. But the phenomena in question, such as a rule in the morphological tier requiring additional input (or context) from the phonological tier, are as likely to pertain to rules within the categories as to ones on the margin. The question remains: what explanatory function does the classification into tiers serve?

This is not to say that it isn't possible to preserve Jackendoff's essential idea of a hierarchy of rules in the processing of speech. As in our analysis of machine communication, a rule is at a higher level than another if the output from the second is available as input to the first. But clear relationships like this do not always lend themselves to clear classifications. It is quite clear, for example, how to rank people by their annual income. But once you get them all lined up in order by income, it is not clear how to classify their economic standing — are they rich? poor? middle class? — even when their place in line gives you an excellent clue. Such classifications unavoidably reflect the biases of the observer.

So if the boundary between categories of syntax is somewhat unclear, what about the boundary at the "top" of them all, the one that separates syntactic concerns from pragmatic and semantic ones? The question betrays a fundamental misconstrual of the problem. That is, we have no *a priori* reason to assume that there is any boundary at all. Why should we assume that brains process "semantic" transformations any differently from the ways they process "syntactic" transformations? To us, the kinds of information seem quite different, but there is no reason to conclude from this that there exists a difference in kind between these sorts of transformations.* The traditional separation between syntax, semantics and pragmatics, is an assumption about how things must work, not an observation about how they actually do.

Jackendoff himself points out that semantic rules for the construction of words and phrases are as productive as syntactic ones (p. 428). By a parallel argument, I can demonstrate that my understanding of folk psychology constitutes a set of productive rules that I use to infer meaning from people and situations I observe. That is, instead of applying them to sentences, I apply these rules to people I know or know of, and can suggest possible explanations for past behavior and predictions of future behavior.** Of course these meanings are more complex and flexible than the ones associated with phrasal semantics, but what evidence do we have that they are manipulated within our brains in a profoundly different way than the ones at the lower level? Until we learn that the brain processes "folk-psychological" production rules in a fundamentally different way than it processes noun-phrase construction rules, would it not be better to assume that both are processed in similar ways? (Similar, but not identical. One difference of course is that — in excellent analogy with the computer example above — the context necessary to understand folk-psychology propositions is much richer than that needed to understand most noun phrases.)

Granting, at least for the moment, that groupings like "morphophonological tier syntactic rules" might not have functional correlates, what can we do? For one thing, we can revisit the error correction techniques described in section 3. Though the identification of levels is contested here, the hierarchical structure of the rules themselves is not. Building a processor out of an assemblage of many simple operators may not only be justifiable on conceptual grounds, but on engineering principles, too. It's a sound way to build an accurate machine out of imprecise parts, since error corrections in the higher levels take care of mistakes or ambiguities at the lower level.

Note that, as in our machine communication example, the corrections need not happen in immediately successive levels. Whatever it is that distinguishes an [m] from an [n] might be missed at the low level. Context at the next level can help if we're talking about distinguishing the words 'not' and 'mop,' but more context than that will be required to differentiate between 'nearly' and 'merely,' which are

*Of course, only Occam leads one to assume otherwise. Fodor, for one, makes explicit the claim that all brain operations are essentially syntactic in nature (Fodor, 1994). To him, the question is how these syntactic operations could come to have meaning at all. But he, too, subscribes to an oversimplified idea of what it means to be a computer in a real world.

**For example, a rule that when someone is angry at you, it is not a good idea to ask them for a favor, can be applied to the universe of people I know to make similar statements about each one of them.

the same part of speech, and still more between 'Nat' and 'Matt.'* The context necessary — that is, the data in the correction channel — may be embedded in the sentences to be understood, in the words or the tone of voice, or it may be supplied by some entirely distinct source: another of our senses, perhaps, or some piece of information acquired earlier. But this context is an essential part of making some transformation of the message into the message's meaning. That is, it is essential to understanding that message.

In the context of attempting to analyze a working brain, categories such as the many subdivisions of syntax catalogued by Jackendoff cannot be regarded as more useful than a functional classification. Until someone can demonstrate that they actually do reflect a functional reality, there will always be a chance these categories will hinder understanding more than they help.

6. Evolving Partway There

The ideas about language presented here are not of the sort that will admit deductive proof. These are claims about how abstractions must be realized, not directly testable propositions about the world. But these ideas do lend themselves to predictions through which we can construct two arguments about their plausibility. The first such argument concerns the evolution of the structure of brains.

Since the original statement of what has come to be known as the Baldwin effect, attention has been given to the evolution of learning and the evolution of learning species. The Baldwin effect states that a population of animals that acquire a trait or learn a behavior during their lifetimes often evolve to have that trait innately. This is not a Lamarckian stance, but a simple statement that by learning, a population of organisms alters the fitness space it occupies, creating new selection pressures that may favor animals who have acquired this learning innately (Baldwin, 1896; Waddington, 1961).

A surprising feature of evolution under these conditions is that we can seldom expect the newly-evolved innate trait to be as effective as the learned trait was. Why exactly is disputed. It might be because increasing fitness past a certain point involves subtle costs that elude our sums. Or it might simply be that once a population has evolved to 90% of some fitness maximum, there is no selective advantage to an organism that is born at 95% of the maximum, especially if the ones born at 90% can quickly learn enough to be the equal of anyone born with a head start. In fact, the degree of flexibility involved in the learning apparatus can work *against* approaching the fitness maximum (Ancel, 2000), especially in a variable environment (Anderson, 1995). When anyone in the population can learn what is needed quickly, what selective advantage can be gained by being born with a given trait?

*Following this presentation of the levels of analysis, a figure of speech or a pun might be just a deliberate way to force your audience's analysis to jump up a level or two, where the meanings are richer and the connotations more fraught.

Whatever the reason, the findings suggest that, so long as the organism retains its ability to learn, we can expect any subsystem that can be augmented by that ability *not* to evolve to innately perfect performance. Conversely, traits acquired by organisms or systems without much flexibility are likely to evolve to become completely innate. For example, the number of useful features an ostrich's feet can develop during its lifetime is quite small compared to the number of behaviors a lion's brain can learn. Consequently, ostriches evolved callouses that develop in the egg, but young lions must be taught how to kill a water buffalo.*

Now consider some behavior in some learning animal that involves making some discrimination among sensory data, such as distinguishing between benign and poisonous mushrooms, or sensing kin. According to the Baldwin effect, over time this animal might (depending on how flexible and fast a learner it is) evolve an innate ability to make this discrimination successfully say 60% of the time. After the generations have flowed past, the infant descendants of these animals are now faced with a *different* learning problem than their ancestors. They no longer have to learn how to handle the pure discrimination problem. When faced with this same problem, they now have to learn both how to interpret the result of their now-innate discriminator (which might now be called their "intuition" or "instinct"), and how heavily to rely on it. They will learn, that is, how to correct its errors.

In other words, innate instincts are not perfect, and therefore require error correction. Vervet monkeys, for example, use their innate instinct to decide which of three different alarm calls to make (depending on whether the predator they detect is a snake, a leopard, or an eagle), but they must learn to avoid the false alarms so common among the younger monkeys (Cheney and Seyfarth, 1990, p. 129ff).

This new error correction behavior is potentially subject to assimilation by the genotype in precisely the same way as the first. So after more generations pass, the descendants of our learning animal may develop so that the output of the 60% accurate structure flows into the input of some other structure that can correct at least some of its mistakes. In order to correct mistakes, this second structure needs either some kind of access to the original stimuli of the first, or some other form of access to the context in which the earlier operation was completed. Thus, applied to a brain many times in succession, the Baldwin effect would tend to produce exactly the situation suggested in the above discussion of the organization of syntax rules: a hierarchy of imperfect processors, each set up to use context to correct others' errors, resulting in a processor greater than the sum of its parts.**

*Papaj, 1993 points out that flexible learners in the insect world are more likely to assimilate new behaviors into new instincts if only because they are more likely to propose new behaviors. The more behavioral possibilities proposed, the more likely it is that one will evolve into an instinct. The more flexible animals will have more instincts than their less flexible cousins, but those instincts may each be less refined.

**Note that no claim is made here about how the language analysis "circuitry" humans use actually did evolve. Some authors (e.g., Turkel (2002)) claim that the Baldwin effect can explain many aspects of the evolution of language, while others Deacon (1997) point out that linguistic variability among

7. Skrymir's Horn

In addition to the evolutionary argument just cited, there is a second plausibility argument for this conception of syntax. The situation described here, where categories of syntactic rules are essentially subjective, would, if it were true, produce a situation where the dividing line between different sets of rules is unclear and a ripe subject for argument. The dividing lines between syntax, semantics and pragmatics would be as unclear as the line between phonological rules and lexical ones. It would create a situation where, depending on how you define the levels, you could demonstrate that people contain specialized syntax engines for very specific micro-domains of language, but you could never be able to specify an equally complete grammar for the rest of any language. Nor, for that matter, would you likely be able to find a micro-domain without exceptions to the rules. It would create a situation where close examination of any such micro-domain could lead one to conclude there is enough regularity in the system to derive grand rules to explain all, but where the interactions between these domains are complex and context-dependent enough that it would be impossible in practice to derive these rules. In other words, it would explain a great deal about the state of modern linguistics.

There is structure to language, and it is worth studying — since it appears to be an important part of what makes us human — but the idea that it is a thoroughly formal structure, with neatly separated syntax, semantics and pragmatics, is a leftover from an idealized version of communication that never existed, but was simply postulated to be analogous to the formal systems of Russell and Frege.*

What's more, the brains that create and understand language are dependent on the circumstances of an utterance to provide context for correcting errors in its reception and interpretation. Without that context, it is unlikely that sentences can be understood by any actual human. Of course, sample sentences are regularly presented without context in linguistics texts. But consider a sentence like:

Lydia threw the ball.

Without knowing who Lydia is, how far she can throw, what kind of ball she threw (baseball? football? fancy-dress ball?), and to whom she threw it, what

---

human populations and the short time scales of language changes make this unlikely, except for the most fundamental issues. The point attempted here is a more general one: whether the structures we use for language evolved because of language or have been adapted to that use from some other purpose, the Baldwin effect and its consequences suggest they will likely be of a similar structure, consisting of many hierarchically arranged processors whose reliability depends heavily on error correction.

*As an aside, devotées of this brand of linguistic formalism often do not display a convincing understanding of the 20th-century developments in how we understand formal systems. From the heirs of Russell and Frege — Hilbert, Gödel, Turing, among others — we have learned that formal systems are somewhat more informal than we'd thought, and that there is nothing blessed about any particular choice of axioms and inference rules. In other words, were it the case that the rules and axioms of grammar were fundamental, universal, and immutable, it would perhaps be the only such formal system known (see, e.g., Chaitin (1982).)

meaning is understood by people who read that sentence isolated on a page of example sentences? If this sentence were transmitted to us, and received as "Lydia thr[unintelligible] the ball," could we fill in the blank without knowing these things? The formal attributes of the sentence — that is, the context of American English in the 21st century — help us guess that the incomplete word is a verb, and possibly even to guess which verb (though she could thrust, threaten or thrill a ball, too). But without knowing more about the situation, we cannot choose the correct tense. People who claim to understand such sample sentences really mean that they understand them in the impoverished sense that allows them to guess a verb, but remain unable to confirm the guess or guess the verb's tense. This is a degree of understanding, after all, but a fairly low one. Linguists rely on their readers understanding the sentences introducing these samples in a much more profound way than they understand the samples themselves.

That understanding comes in degrees is an essential corollary of the view of syntax presented here. Consider how much more we understand from our sample sentence (and how much more of it we could fill in if transmission errors obscured more parts) as we learn each of the following facts:

1. Lydia played on a little league baseball team;
2. She was the only girl on the team, elevated to pitcher by the unfortunate beaning of her teammate;
3. Her brother, the opposing pitcher, was facing her as batter;
4. He is now recovering nicely in the local hospital.

These facts have resolved uncertainty about Lydia, about the ball, about the verb "to throw," about her target, and about the verb tense.* As Shannon points out, in the context of receiving a message, error correction *is* the elimination of uncertainty. Context, then, is not an optional accessory to some utterance, it's crucial to the error correction necessary for analysis of that utterance. The richer the available context, the more profound the understanding, but even the impoverished kind is correctly called understanding.

The syntactic view of linguistic structure has separate processing domains for the syntax, the semantics, and the pragmatics of a sentence: its form, its meaning and the situation in which it is used. It is commonly allowed that interesting phenomena occur at the interfaces of these domains, but the domains are separate (Jackendoff, 2002, p. 425). But, contrary to this common assumption, these clear distinctions are true only for an individual transformation. The syntax of some communication and the context in which it occurs provide the input to some specific transformation of

*They haven't reolved all uncertainty. For example, we don't know what to make of Lydia's aim or intent. More context might reveal this, while raising other questions that might be answered by still more context. This situation, where increasing amounts of context create confusion in some matters while resolving uncertainty in others, is hardly unfamiliar to most people. (See, e.g., Warren (1964).)

that message, while the transformed message *is* the message's meaning. Although this is true in the small scale analysis of a single transformation, it is not possible to extend these concepts to describe a system containing many transformations. In such a system, syntax, semantics and pragmatics become inextricably interleaved by the hierarchical nature of the transformations' arrangement, even while their definitions may seem clear at any chosen point in the process.

To speak of the meaning of a sentence distinct from the context in which it was uttered is pointless. Without the context, you can't get the meaning out. The British information theorist Colin Cherry wrote:

> The meaning and the utterance form a unit: a 'meaningful utterance,' as made by a specific person, in a specific environment and time. A 'meaning' is not a label tied round the neck of a spoken word or phrase. It is more like the beauty of a complexion, which lies 'altogether in the eye of its beholder' (but changes with the light!) (Cherry, 1978, p. 171).

A useful analogy may be made to the study of DNA. The information content of a series of DNA bases is of some interest to those attempting to create databases of base sequences, or to those designing compression schemes for storing or transmitting the genetic data, but all attempts to analyze the function of genes by inferring some formal structure to its arrangements of bases have, to date, failed. (See, for example, (Shapiro, 1999, p. 24), or (IHGSC, 2001).) It is undeniable that there is much interesting structure in the details of how we create and understand language. But using that structure to infer the physical existence of a "phonological tier" of syntactic transformations, for example, is not justified by the data we have available today.

It is a gross distortion, and quite unfair, to say that the project of generative grammar has brought us nothing. We can compare Chomsky to Thor at the giant Skrymir's castle. Challenged to empty Skrymir's huge drinking horn in one draught, and not realizing the end of it was connected to the sea, Thor took a huge gulp. Of course he could not empty it, but the tide did go out.

Several decades of attempting to uncover the structure of language have had two salutary outcomes. The first is that the attempt spurred legions of linguists and psychologists to gather data which, while not supporting the original project very well, now provide a wealth of raw material for the pioneers of cognitive science.

The second outcome is to render conceivable suggestions like the one here, made plausible by those decades of hard effort by those same legions of researchers: an argument from rigor of investigation.

Acknowledgements

I am grateful to Frederic Evans for provoking much of this argument, and to David Gow for early and insightful comments on this paper. Also, the opportunity to

observe the members of the DODS/OPeNDAP project in their debates about how to define science metadata provided many insights into the subjectivity of categorizing syntactic rules (Cornillon et al., 2003).

References

Ancel, L.W., 2000, "Undermining the Baldwin expediting effect: Does phenotypic plasticity accelerate evolution?" *Theoretical Population Biology* **58**, 307–319.

Anderson, R.W., 1995, "Learning and evolution: A quantitative genetics approach," *Journal of Theoretical Biology* **175**, 89–101.

Baldwin, J.M., 1896, "A new factor in evolution," *American Naturalist* **30**, 441–451.

Bar-Hillel, Y., 1953, "A quasi-arithmetical notation for syntactic description," *Language* **29**, 47–58.

Bar-Hillel, Y., 1954, "Logical syntax and semantics," *Language* **30**, 230–237.

Bar-Hillel, Y., 1964, *Language and Information: Selected Essays on their Theory and Application*, Reading, MA: Addison-Wesley. Contains (Bar-Hillel, 1953) and (Bar-Hillel, 1954).

Bocheński, I.M., 1961, *A History of Formal Logic*, University of Notre Dame Press, Translated from the German edition of 1956 by Thomas Ivo.

Bresnan, J. (ed.), 1982, *The Mental Representation of Grammatical Relations*, Cambridge, MA: MIT Press.

Brillouin, L., 1962, *Science and Information Theory*, 2nd edition, New York: Academic Press.

Carnap, R., 1937, *The Logical Syntax of Language*, London: Routledge & Kegan Paul.

Chaitin, G.J., 1982, "Gödel's theorem and information," *International Journal of Theoretical Physics* **22**, 941–954.

Cheney, D.L. and Seyfarth, R.M., 1990, *How Monkeys See The World: Inside the Mind of Another Species*, Chicago, IL: University of Chicago Press.

Cherry, C., 1978, *On Human Communication*, 3rd edition, Cambridge, MA: MIT Press.

Chomsky, N., 1995, *The Minimalist Program*, Cambridge, MA: MIT Press.

Chomsky, N., 2000, *New Horizons in the Study of Language and Mind*, Cambridge: Cambridge University Press.

Cornillon, P., Gallagher, J., and Sgouros, T., 2003, "OPeNDAP: Accessing data in a distributed, heterogeneous environment," *CODATA Data Science Journal* **2**, 164–174. Online 5 November, 2003: http://journals.eecs.qub.ac.uk/codata/Journal/contents/2_03/ 2_03pdfs/DS247.pdf

Deacon, T.W., 1997, *The Symbolic Species: The Co-Evolution of Language and the Brain*, New York: W. W. Norton.

Dupuy, J.-P., 2001, *The Mechanization of the Mind: On the Origins of Cognitive Science*, Princeton, NJ: Princeton University Press. Translated from French (1994) by M.B. DeBevoise.

Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and Berners-Lee, T., 1999, Hypertext Transfer Protocol – HTTP/1.1, RFC 2616, Network Working Group.

Fodor, J.A., 1994, *The Elm and the Expert: Mentalese and its Semantics*, Cambridge, MA: MIT Press.

Graffi, G., 2001, *200 Years of Syntax: A Critical Survey*, Vol. 98 of *Amsterdam Studies in the Theory and History of Linguistic Science*, Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hartley, R.V.L., 1928, "Transmittion of information," *Bell System Technical Journal* **7**, 535–563.

IEC, 1994, "Information technology – Open Systems Interconnection – Basic Reference Model: The Basic Model – ISO/IEC 7498," Technical report, International Electrotechnical Commission.

IHGSC, 2001, "Initial sequencing and analysis of the human genome," *Nature* **409**, 860–921.

Jackendoff, R., 2002, *Foundations of Language*, Oxford: Oxford University Press.

Montague, R., 1970a, "English as a formal language," in *Linguaggi nella società e nella tecnica*, B. Visentini ed., Milan: Edizioni di Comunità.

Montague, R., 1970b, "Universal grammar," *Theoria* **36**, 373–398.

Papaj, D.R., 1993, "Automatic behavior and the evolution of instinct: Lessons from learning in Parasitoids," in *Insect Learning: Ecological and Evolutionary Perspectives*, D.R. Papaj and A.C. Lewis eds., New York: Chapman & Hall, Chapter 10, pp. 243–272.

Pollard, C. and Sag, I.A., 1994, *Head-Driven Phrase Structure Grammar*, Chicago, IL: University of Chicago Press.

Resnick, P., 2001, Internet Message Format, RFC 2822, Network Working Group.

Rogers, J., 1999, *A Descriptive Approach to Language-Theoretic Complexity*, Stanford, CA: CSLI Publications.

Shannon, C.E., 1948, "A mathematical theory of communication," *Bell System Technical Journal* **27**, 379–423, 623–656.

Shapiro, J.A., 1999, "Genome system architecture and natural genetic engineering in evolution," in *Molecular Strategies in Biological Evolution*, L. Caporale ed., *Annals of the New York Academy of Sciences* **870**, 23–35.

Stevens, W.R., 1999, *UNIX Network Programming*, 2nd edition, New York: Prentice-Hall, Inc..

Turkel, W.J., 2002, "The learning guided evolution of language," Chapter 8, pp. 235–254 in *Linguistic Evolution Through Language Acquisition*, T. Briscoe ed., Cambridge: Cambridge University Press.

Waddington, C., 1961, "Genetic assimilation," *Advances in Genetics* **10**, 257–293.

Warren, E., 1964, *Report of the President's Commission on the Assassination of President Kennedy*, Washington, DC: US Government Printing Office.

Wiener, N., 1961, *Cybernetics*, 2nd edition, Cambridge, MA: MIT Press.