Science Progress (2006), 89(1):61-70.
www.scilet.com

The nature of language

WILLIAM L. ABLER
Department of Geology
The Field Museum
1400 South Lake Shore Drive
Chicago, Illinois 60605, USA
E-mail: wabler@fieldmuseum.org

ABSTRACT

Sentences share with equations properties of discrete, linear infinities; distinct symbol types; alternation of symbol types in the signal; deep-structure with a main verb and recursion; associative, commutative, and distributive properties; autonomous levels of organization; paraphrase, ellipsis, and ambiguity; powers of assertion, truth and falsity; modification of an entire statement; and a blank proxy element. Sentences are derived from the symmetrical equation by applying a single asymmetry transformation. The self-regulated, autonomous nature of the symmetrical equation means that the idea of language evolution presents numerous obstacles. For the most part, ideas contained in this paper are adapted from the author's book, Structure of Matter, Structure of Mind[1].

Keywords: *language, mathematics, foundations, origin, evolution*

Introduction

When the origin of language is understood as part of the debate between evolution and Creation, responsible scientists must choose evolution. Within a few years after the publication of Darwin's *Origin* (1859), the possibility was considered, but never examined systematically, that the core structures of language may have their source in first principles of matter. That deficiency will be remedied here. By comparing language to other, related systems, rather than simply looking at language, we can gain a stable perspective on its place in nature. As a working definition, by 'language' I will mean 'sentences', because where there are sentences, there is language. Ultimately, it is hoped that we can reach a better definition.

In addition to generating discrete, linear infinities, a property shared with the gene, language shares nearly all its higher-level universal properties with arithmetic. Structures and operations shared between the two systems are so exhaustive that, in themselves, they would bring either system to a minimum functioning standard. In spite of the complex changability of every language at its fullest development, shared properties permit us to examine both systems in the stable simplicity of the universal structural core. The present paper begins with an examination of shared structures, operations, powers, and properties. In everyday conversation, the word 'language' is sometimes used in a metaphorical sense to mean 'information' or 'communication', as in 'the language of life', or 'mathematics is a language'. Here, the word 'language' is restricted to its primary meaning of 'naturally-occurring human language'.

Comparison

Distinct symbol types

Language and arithmetic have an unlimited number of symbols, i.e., the numbers and vocabulary words, whose relationships are expressed by a small, fixed set of operations. In arithmetic, the operators include 'plus', 'minus', 'times'; in language they include such words as 'so', 'or', 'and', 'of', 'in spite of'.

Symbol alternation

The symbol types alternate, and must alternate, in the linear signal. This rule applies to inflected languages as much as to word-order languages, i.e., the inflections are essentially bound grammar words. Thus, we find:

"John and Tom worked on the book for over a year", but not
*John Tom work book year and ed on the for over a

"9×11=99", but not
*91199×=

"Arma virumque cano", but not
*Arm vir can a um que o

Linear ordering

Not just sentences of language, but also equations, can be expressed in linear form. For example,

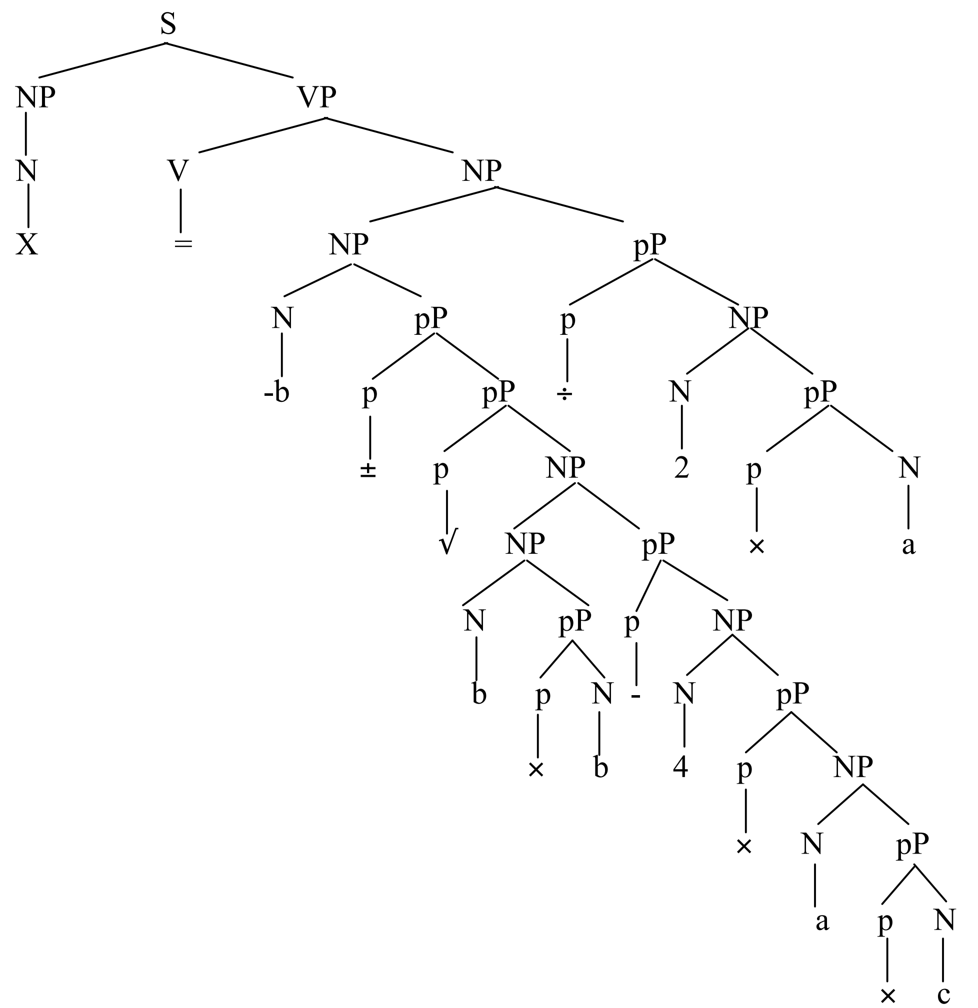
$$X= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

can be read out in linear form, "X equals negative b, plus-or-minus the square-root of b-squared minus four a-c, all over two-a"; and can be written in linear form, thus.

$$X = [-b \pm \surd (b2 - 4ac)] \div 2a$$

Deep-structure with main verb and recursion

Not only sentences, but all equations, have a linguistic deep-structure with a main verb. Only the deep-structure of an equation is shown here. Numbers are represented as nouns (N); operators are represented as prepositions (p); and the main verb (V) is ‘equals’. The equation is the quadratic solution used above.



Paraphrase, ellipsis, and ambiguity

Paraphrase, ellipsis, and ambiguity hold a special place among language universals, because they appear to be largely semantic, rather than formal, in origin. But appearances can be misleading; and paraphrase, ellipsis, and ambiguity can provide clues to the structural basis of sentences. Paraphrase can be defined as conveying the same meaning with different words. For example, “Glass is transparent” is a paraphrase of, “You can see right though glass”. Ellipsis can be defined as conveying the same meaning with fewer words. For example, “John won the hundred-meter race, but not the five-hundred” is an elliptical form of, “John won the hundred-meter race, but [John did] not

[win] the five-hundred [meter race]". Ambiguity, or double-meaning, can be defined as conveying different meanings with the same words. For example, "The diplomat knew the value of caviar" means both, "The diplomat knew the price of an ounce of caviar", and "The diplomat understood the use of caviar in diplomacy".

But paraphrase, ellipsis, and ambiguity also exist in arithmetic. Thus, 'six', 'square-root of thirty-six', and 'thirty-six raised to the one-half power' paraphrase each other because they convey the same meaning, but have different words and syntax. Further, 'six times three-plus-three' is an example of ellipsis because it means 'six-times-three plus [six times] three', where the phrase 'six times' is spoken once but used twice. Last, the expression, 'six times three plus three' means both 'six times three-plus-three equals thirty-six' and 'six-times-three, plus three, equals twenty-one', i.e., it is an example of ambiguity in arithmetic, as is 'square-root of one'.

Under arithmetic, then, paraphrase, ellipsis and ambiguity are not a loosely organized system of superficially similar devices, but theorems that follow inevitably from the symmetry property of equations. Thus,

$$6 = \sqrt{36} = 36^{1/2} \qquad \text{(paraphrase)}$$

$$6(3 + 3) = 6 \times 3 + [6 \times]3 \qquad \text{(ellipsis)}$$

$$6 \times 3 + 3 = 36; \text{ and } 6 \times 3 + 3 = 21 \qquad \text{(ambiguity)}$$

are all true because each side of each equation is symmetrical with the opposite side of that equation.

Autonomous levels

Paraphrase, and especially ambiguity, in arithmetic and language show that the levels of organization are separate and autonomous in both systems. Thus, if a string of symbols specified a precise meaning, then the same symbols arranged in the same order would specify the same meaning. But ambiguity both in arithmetic and language shows that a string of symbols does not specify a precise meaning in either system. To a lesser extent, if meaning were specified by a string of symbols, different strings of different symbols would specify different meanings. But paraphrase shows that this is not inevitable in either arithmetic or language. Thus, meaning is not spelled out precisely by symbols in either system; and meaning remains at least partially independent from the symbols that indicate it in both systems. Paraphrase and ambiguity are unique to sentences and equations.

Associative, commutative, distributive

While associative, commutative, and distributive properties are considered fundamental in arithmetic, they are equally a property of language.

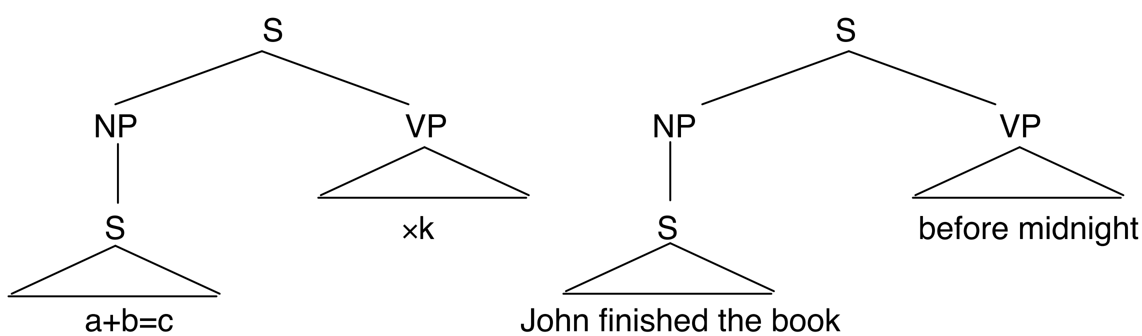
Thus, in arithmetic:

Associative: $a + b = b + a$
Commutative: $(a + b) + (c + d) = (d + b) + (c + a)$
Distributive: $k(a + b) = ka + kb$

And in language:

Associative: John and Tom opened the package vs. Tom and John opened the package.
Commutative: John and Tom, and Ben and Pete opened the package vs. Pete and Tom, and Ben and John opened the package.
Distributive: John bought coffee and tea vs. John bought coffee, and John bought tea.

Multiplying by a constant; modifying by an adverb clause
The standard arithmetical operation of multiplying an equation by a constant is the formal equivalent of modifying a sentence by an adverb clause. Thus:



$(a + b = c)\,k$ John finished the book before midnight.

Power of assertion; truth and falsity
Equations uniquely share with sentences the power of assertion. Thus, when I say that "$a + b = c$", I assert that the material on one side of the 'equals' is somehow the same as the material on the other side. If the sentence in equation form is, say, "$7 + 7 = 14$", then it is true. If the sentence is, say, "$7 + 8 = 14$", then it is false, i.e., since equations are self-referential, they are true when they are symmetrical, and false when they are not. But formulas in equation form, e.g., "$a=b$", are always assertions and capable of being true or false. Sentences of ordinary language, such as, "John can recite The Tempest from memory" also make an assertion, but one whose truth or falsity must be checked outside

language. Human culture depends upon the power of assertion, since all laws, both natural, such as the Pythagorean theorem, or human, such as The Ten Commandments, have to be assertions.

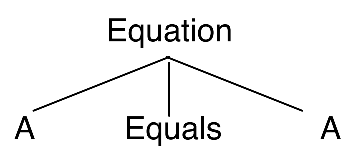
Blank proxy

The first defining property of algebra is the introduction of the variable "x". X represents a number whose identity we do not know, so that it might, in principle, assume any value, depending on the equation where it finds itself. In this respect, x is a proxy that takes its identity from its context; and that is precisely the function of the pronoun in language. While pronouns in many languages carry the beginnings of identity, "I/me; he/him, she/her, we/us, they/them", in English, for example, pronouns are still blank proxies in taking from context the precise identity of some particular he, or she, or it, or they, or them. The Chinese pronoun "ta", which does not specify gender or number or case, is closer to the x of algebra.

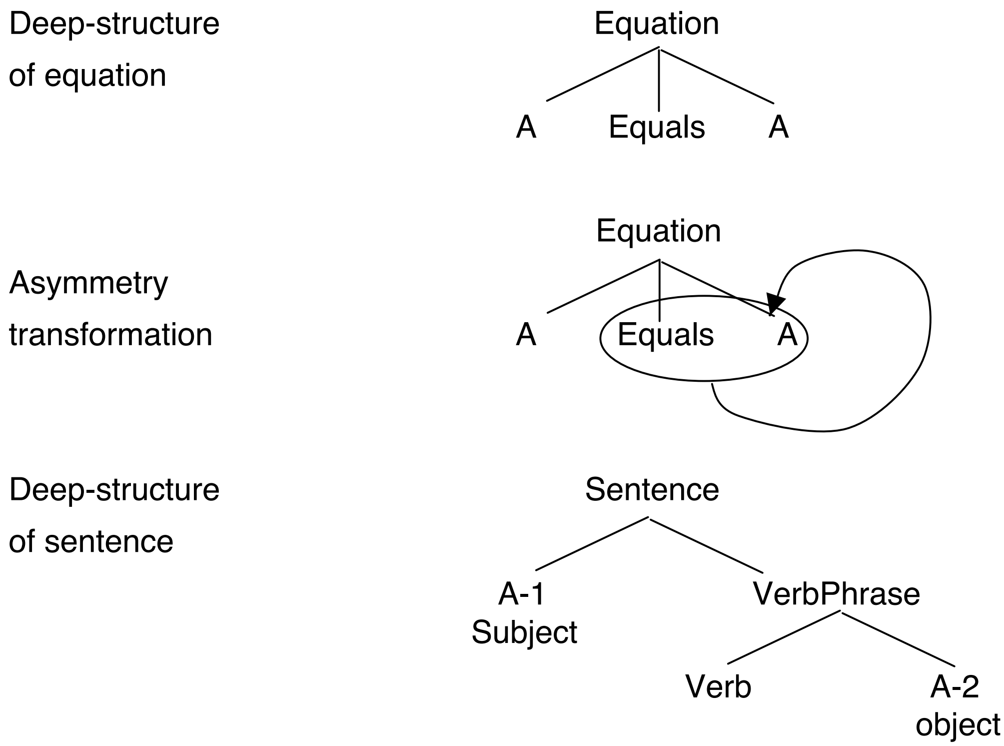
Symmetry versus asymmetry

The single critical property that differentiates equations from ordinary sentences is symmetry, i.e., equations are symmetrical, while ordinary sentences are not. The symmetry of equations means that the two sides can be interchanged to produce the same equation. For example "$\sqrt{25} = 5$" is the same equation as "$5 = \sqrt{25}$". Sentences are not symmetrical. For example, "John ate the apple" is not the same sentence as, "The apple ate John". Thus, to generate ordinary sentences, a small but fundamental revision is needed in the deep-structure of the equation, i.e., it must be made asymmetrical.

Symmetrical deep-structure of the basic equation



The deep-structure of the sentence can be derived from that of the equation by moving two of the branches out to the end of one of them; and the deep-structure of the equation can be derived from that of the sentence by the reverse transformation.



Since each can be derived from the other by applying the same transformation in the opposite direction, equations and sentences share a common source. Convergence, by contrast, would have highlighted one or two key features, not the system as a whole.

The asymmetry transformation converts the equation deep-structure directly to the Subject-Verb-Object SVO form of the sentence, as in English or Chinese, as shown in the figure. The application of one more transformation generates the SOV deep-structure, as in Japanese or Hindi. The VSO sentence type is formally complicated, but all VSO languages have an alternate SVO sentence type. There are no VOS, OSV, or OVS languages, i.e., no Object-Subject languages (Greenberg 1963:61).

The equation deep-structure itself is symmetrical, and in that property is unlike any natural language, since natural languages are asymmetrical in the Subject-Object relation. The equation is instead the shared source, or common ancestor, of the sentence formula. The biological system that exploits the sentence formula does so by stretching it beyond its ancestral shape.

The idea of language evolution

As a developmental matter, either (a) equations are an off-shoot of sentences; or (b) sentences are an off-shoot of equations. Sentences were known first in human experience, of course, but equations can not be dismissed on that basis. Equations are true when they are symmetrical, i.e., equations are self-referential, self-regulated, and independent. Their property of truth or falsity is universal, and independent of external conditions.

In their independence, equations are the antithesis of natural selection, which is local in its action, and contingent upon local conditions. Equations, which are independent of external conditions, are not a product of evolution in any sense. We can interpret them and exploit them; but the universal and independent nature equations means that they are found, not made; so it make sense to ask whether equations are first.

If equations are first, then the idea of language evolution has no referent in the natural world, since ordinary sentences are then derived from equations by the action of a single transformation. In this case, the contribution of a biological system is to find and apply that single transformation.

On the other hand, the idea that language evolved by well-understood mechanisms of natural selection, and that arithmetic is an off-shoot of language, presents numerous obstacles. One possibility is that arithmetic did not evolve because it is an off-shoot of language. That idea is the linguistic-arithmetical equivalent of believing that teeth did not evolve because they are an off-shoot of scales. It corresponds to nothing in the natural world.

Another obstacle is presented by differentiation. Arithmetic, with its simpler, less-differentiated structure, might be seen as a vestigial system, derived from language through disuse, like the eyes of cave-dwelling fishes. Or arithmetic might be viewed as a specialized system, derived from language through further selection. In either case, we must ask through what stages language might have passed during its evolutionary development. Since a deep-structure simpler than that of the basic sentence could only be that of the basic equation, to a first approximation, language is derived from arithmetic.

A language-first theory presents the puzzle of a system of structures, properties, powers, and operations, all of which emerged by random chance in a primary system, and all of which would have been different under different environmental conditions, becoming inevitable theorems in a derived system that was not being selected for, and is not a product of selection. Further, many properties of language are of doubtful adaptive value in the first place. Ambiguity, or double-meaning, for example, reduces by half the number of sentences that a speaker must be able to generate; but half of infinity is still infinity, removing any advantage. Ellipsis reduces the length of discourses by making sentences shorter, but shortness is not obviously an advantage. French labels on bilingual packaging are always longer than English, but that does not mean that English is more advanced than French. If shortness were an advantage, it would be a trend in language change; but Latin texts, and Latin is a precursor of French, are shorter than their English translations. Paraphrase, saying the same thing in different ways, improves speakers' ability to explain, but at a cost of increasing the length of discourses. On the other hand, if we take arithmetic as being first, we find that its properties are simply retained in the daughter system, language.

Another objection to the idea of language evolution has its basis in evolutionary theory. In principle, selective advantage is expressed by an equation. If selective advantage was involved in the evolution of sentences, and if equations are an off-shoot of sentences, we have the contradiction of using an equation to explain the existence of equations, i.e., of assuming the proof. Using sentences to explain the existence of sentences is not a contradiction, because sentences are not self-referential, i.e., their truth or falsity lies outside language, and assumes nothing.

Further, the kind of symmetry that evolves in biology, such as that of a butterfly, is never perfect; but the symmetry of equations is perfect. The perfect symmetry of equations, then, poses a genuine obstacle to the idea first of arithmetic evolution, then of language evolution.

Special pleading apart, the idea of language evolution has no referent in nature. Sentences are equations that have been converted into an asymmetrical form, i.e., language is an asymmetrical form of arithmetic, and may be defined on that basis. The underlying structures of language have their source in fundamental physical properties of discreteness, symmetry, embedding (recursion), and linear ordering. The specific contribution of biology to this basic formula in physics is to impose asymmetry upon it, creating the basic formula of the sentence. Through the mediation of the asymmetry transformation, then, language exploits foundations of physics directly.

Conclusion

Language scientists, and others concerned with language, may find value in this view of language. No longer an accidental and isolated by-product of properties that belong to other sciences, language emerges as a direct reflection of first principles of nature, and capable of casting light upon them.

References

Abler, W.L. (2005) *Structure of Matter, Structure of Mind*, Sofia: Pensoft; Philadelphia: Bainbridge.

Greenberg, J.H. (1963) *Universals of Language*, Cambridge: MIT.