
David Mumford

Archive for Reprints,
Notes, Talks, and Blog

Professor Emeritus
Brown and Harvard Universities
David_Mumford@brown.edu

[HOME](#) [ALGEBRAIC
GEOMETRY](#) [VISION](#) [BEYOND
RESEARCH](#) [ABOUT](#) [BLOG](#)

Grammar isn't merely part of language

October 12, 2016

This post is inspired by reading the latest Tom Wolfe diatribe, "The Kingdom of Speech". While the book sets off to discuss the issues of what were the origins and evolution of speech in early man, the largest part of this book is devoted to a juicy recounting of the feud between Noam Chomsky and Daniel Everett over whether recursion and other grammatical structures must be present in all languages. Chomsky famously holds that some mutation endowed early man with a "language organ" that forces all languages to share some form of its built-in "universal grammar". Everett, on the other hand, was the first to thoroughly learn the vastly simplified language spoken by the Amazonian Piraha (pronounced *peedahan*) that possesses very little of Chomsky's grammar and, in particular, appears to lack any recursive constructions (aka embedded clauses). What I want to claim in this blog is that both are wrong and that grammar in language is merely a recent extension of much

The Dismal Science and the future of work

July 20, 2015

Grammar isn't merely part of language

October 12, 2016

Hard to be an optimist

January 1, 2017

Let the mystery be

April 13, 2018

Can an artificial intelligence machine be conscious?

April 11, 2019

Can an artificial intelligence machine be conscious, part

older grammars that are built into the brains of all intelligent animals to analyze sensory input, to structure their actions and even formulate their thoughts. All of these abilities, beyond the simplest level, are structured in hierarchical patterns built up from interchangeable units but obeying constraints, just as speech is.

I first encountered this idea in reading my colleague Phil Lieberman's excellent 1984 book "The Biology and Evolution of Language". Most of this book is devoted to the still controversial idea that *Homo Sapiens* carries a mutation lacking in *Homo Neanderthalensis* by which its airway above the larynx was lengthened and straightened allowing the posterior side of the tongue to form the vowel sounds "ee", "ah", "oo" (i,a,u in standard IPA notation) and thus increase hugely the potential bit-rate of speech. If true, this suggests a clear story for the origin of language, consistent with evidence from the development of the rest of our culture. However, the part of his book that concerns the origin of syntax -- and in particular Chomsky's language organ hypothesis -- is the beginning, esp. chapter 3. His thesis here is:

"The hypothesis I shall develop is that the neural mechanisms that evolved to facilitate the automatization of motor control were preadapted for rule-governed behavior, in particular for the syntax of human language."

He proceeds to give what he calls "Grammars for Motor Activity", making clear how parse

II?

July 12, 2019

Letter to my Grandchildren

March 1, 2020

Ridiculous Math Problems

April 1, 2020

Reading Spinoza

April 19, 2020

Nothing is simple in the real world

July 8, 2020

The Astonishing Convergence of AI and the Human Brain

October 1, 2020

Ruminations on cosmology and time

March 1, 2021

The Shape of Rogue Waves

June 27, 2021

Draft of a Blog Book

August 16, 2022

Black hats and white hats

December 12, 2023

Consciousness,

trees almost identical to those of language arise when decomposing actions into smaller and smaller parts. It is curious that these ideas are nowhere referenced in the paper of Hauser, Chomsky et al (Frontiers of Psychology, vol. 5, May 2014) that generated Wolfe's diatribe.

My research connected to the nature of syntax came from studying vision and taking admittedly somewhat controversial positions on the algorithms needed, especially those used for visual object recognition, both in computers and animals. In particular, I believe grammars are needed in parsing images into the patches where different objects are visible and that moreover, just as faces are made up of eyes, nose and mouth, almost all objects are made up of a structured group of component smaller objects. The set of all objects identified in an image then forms a parse tree similar to those of language grammars. Likewise almost any completed action is made up of smaller actions, compatibly sequenced and grouped into sub-actions. The idea in all cases is that the complete utterance resp. complete image resp. complete action carries many parts, some parts being part of other parts. Taking inclusion as a basic relation, we get a *tree* of parts with the whole thing at the *root* of the tree and the smallest constituents at its *leaves* (computer scientists prefer to visualize their "trees" upside-down with the root at the top, leaves at the bottom, as is usual also for "parse trees"). But at the same time, each part might have been a constituent of a different tree making a different whole and any part can be replaced by others making a possible new whole -- i.e. parts are interchangeable within limits set by

Robots, and DNA

January 3, 2024

STATEMENT

April 1, 2024

Letter to my Grandchildren-2

August 30, 2024

Black hats and white hats-II

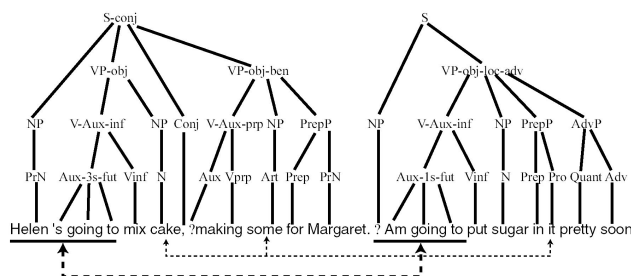
January 20, 2025

constraints which apply to all trees with these parts. There is a very large set of potential parts and each whole utterance resp. image resp. action is built up like leggos of small parts put together respecting various rules into larger ones and continuing up to the whole. Summarizing, all these data structures are *hierarchical* and made up of *interchangeable* parts and subject to *constraints* of varying complexity. I believe that any structure of this type should be called a grammar.

Here are some examples taken from **my talk at a vision workshop in Miami in 2009**. Let me start with examples from languages.

Remember from your school lessons that an English sentence is made up of a subject, verb and object and that there are modifying adjectives, adverbs, clauses, etc. Here is the parse of an utterance of a very verbal toddler (from the classic paper "What a two-and-a-half-year-old child said in one day", L. Haggerty, J. Genetic Psychology, 1929, p.75):

Grammar in the parsed speech of Helen,
a 2 ½ year old



Here we have two classical parse trees plus a question mark for the implied but not spoken subject of the second sentence plus two links between non-adjacent words that are also

syntactically connected. The idea of interchangeability is illustrated by the words "for Margaret", a part that can be put in infinitely many other sentences, a part of type "prepositional phrase". The top dotted line is there because the word "cake" must agree in number with the word "it". For instance, if Margaret had said she wanted to make cookies, she would need to say "them" in the second sentence (although such grammatical precision may not have been available to Margaret at that age). A classic example of distant agreement, here between words in one sentence with three embedded clauses is "Which problem/problems did you say your professor said she thought was/were unsolvable?" This has been used to argue for the transformational grammars by Chomsky. This is not unreasonable but we will argue that identical issues occur in vision, so neural skills for obeying these constraints must be more primitive and cortically widespread.

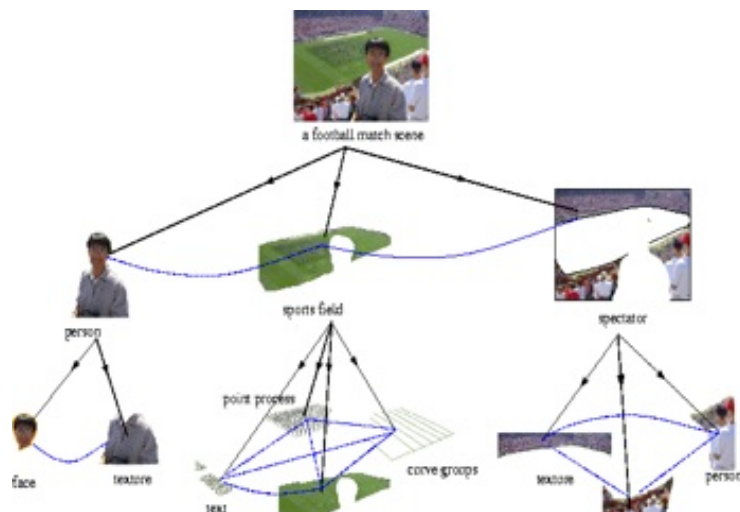
In other languages, the parts that are grouped almost never need to be adjacent and agreement is typically between distant parts, e.g. in Virgil we find the latin sentence

Ultima Cumaei venit iam carminis aetas

which translates word-for-word as "last of-Cumaea has-arrived now of-song age" or, rearranging the order as dictated by the disambiguating suffixes: "The last age of the Cumaean song has now arrived". Thus the noun phrase "last age" is made up of the first and last words and genitive clause "of the Cumaean song" is the second and fifth words,

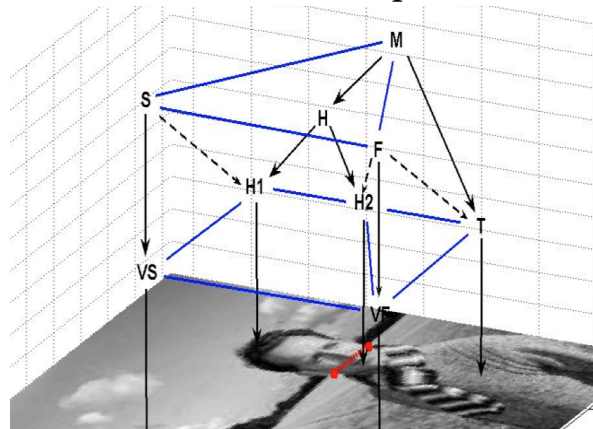
while the verb phrase "now arrived" is in the very middle. The subject is made up of four words, numbers 1,2,5 and 6. So word order is a superficial aspect but there is still an underlying set of parts, distinguished by case and gender, that are interchangeable with other possible parts, these parts altogether forming a tree (the "deep" structure). In other languages, e.g. sanskrit, words themselves are compound groups, made by fusing simpler words with elaborate rules that systematically change phonemes, as detailed in Panini's famous c.300 BCE grammar. Here the parse tree leaves can be syllables of the compound words.

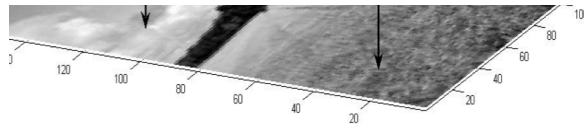
It was a real eye-opener to me when it became evident that images, just like sentences, are naturally described by parse trees. For a full development of this theory, see [my paper "A Stochastic Grammar of Images"](#) with Song-Chun Zhu. The biggest difference with language grammars is that in images there is no linear order between parts. and even, when one object partly occludes another, two non-adjacent patches of an image may be parts of one object and a hidden patch is inferred. Here is an example, courtesy of Song-Chun Zhu, of the sort of parse tree that a simple image leads to:



The football match image is at the top, the root. Below this, it is broken into three main objects -- the foreground person, the field and the stadium. These in turn are made up of parts and this would go on to smaller pieces except that the tree has been truncated. The ultimate leaves, the visual analogs of phonemes, are the tiny patches (e.g. 3 by 3 or somewhat bigger sets of pixels) which, it turns out, are overwhelmingly either uniform, show edges, show bars or show "blobs". This emerges both from statistical analysis and from the neurophysiology of primary visual cortex (V1).

Grammatical constraints are present whenever objects break up into parts whose relative position and size are almost always constrained so as to follow a "template". The early twentieth century school known as *gestalt psychology* worked out more complex rules of the grammar of images (although not, of course, using this terminology). They showed the way, for example, that symmetry and consistent orientation of lines and curves creates groupings of non-adjacent patches and demonstrated how powerfully hidden patches, hidden by occlusion, were inferred by subjects. Here is a simple image in which the occluded parts have been added to the parse tree:





The blue lines indicate adjacency, solid black arrows are inclusion of one part in another and dotted arrows point to a hidden part. Thus H₁ and H₂ are the head, separated into the part occluding sky and the part occluding the field, and joined into the larger part H. S is the sky while VS is the visible part of the sky and H₁ conceals an invisible part, Similarly for the field F and VF. The man M is made up of the head H and torso T. One can even make an example analogous to the above sentence concerning the professor's unsolvable problem, in which a chain of partially occluded objects acts similarly to the chain of embedded clauses:

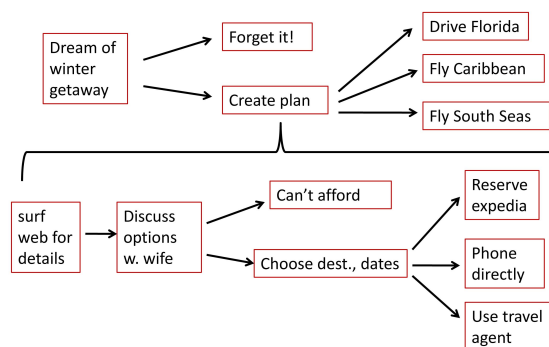


Here we have trees to the left of the chief, whose left arm occludes the teepee that occludes the reappearing trees -- whose color must match that of the first trees.

Returning to motor actions and formation of plans of action, it is evident that actions and plans are hierarchical. Just take the elementary school exercise -- write down the steps required to make a peanut butter sandwich. No matter what the child writes, you can subdivide the action further, e.g. not "walk to the refrigerator" but first locate it, then estimate its distance, then take a set of steps checking for obstacles to be avoided, then

reach for handle etc. The student can't win because there is so much detail that we take for granted! Clearly actions are made up of interchangeable parts and clearly they must be assembled so as to satisfy many constraints, some simple like the next action beginning where the previous left off and some subtler.

The grammars of actions are complicated, however, by two extra factors: *causality* and *multiple agents*. Some actions cause other things to happen, a twist not present in the parse trees of speech and images. Judea Pearl has written extensively on the mathematics of the relation of causality and correlation and on a different sort of graph, his *Bayesian networks* and *causal trees*. Moreover, many actions involve or require more than one person. A key example for human evolution is that of hunting. It is quite remarkable that Everett describes how the Piraha use a very reduced form of their language based on whistling when hunting. From the standpoint of the mental representation of the grammar of actions, a third complication is the use of these grammars in making plans for future actions. An example where some of the many expansions of one plan are shown is:



To summarize, I believe that any animal that can use its eyes to develop mental

representations of the world around it or can carry out complex actions involving multiple steps must develop cortical mechanisms for using grammars. This includes all mammals and certain other species, e.g. octopuses and many birds. These grammars involve a mental representation of trees built from interchangeable parts and satisfying large numbers of constraints. Language and sophisticated planning may well be unique to humans but grammar is a much more widely shared skill. How this is realized e.g. in mammalian cortex, is a major question, one of the most fundamental in the still early unraveling of how our brains work.

The blog above has been translated into Estonian by Sonja Kulmala in Tartu: [here's the link](#) . It's certainly interesting to contrast the grammars in synthetic agglutinative languages like Estonian and Turkish with analytic languages like Chinese. Thank you Sonja.

After reading the above, Prof. Shiva Shankar drew my attention to the following that appears on cover of Frits Staal's book *Ritual and Mantras, Rules without Meaning*:

An original study of ritual and mantras which shows that rites lead a life of their own, unaffected by religion or society. In its analysis of Vedic ritual, it uses methods inspired by logic, linguistics, anthropology and Asian studies. New insights are offered into various topics including music, bird song and the origin of language. The discussion culminates in a proposal for a new human science that challenges the current dogma of 'the two cultures' of

sciences and humanities.

He seems be saying that rituals and mantras are an embodiment of a purely abstract grammar, expressed in both action and speech with a minimum of semantic baggage.

The blog above has been translated into German by Philipp Egger: [click here](#). Thank you Philipp. And it has been translated into Russian by Michael Taylor: [click here](#). Thank you Michael.

David Mumford's content on this site is available under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#). [Click here for sitemap](#)
