Principals of Certain Semantic Processes

UNEDITED DRAFT

by David McGoveran, 2014

Many natural systems can be represented as two or more independent, yet interacting systems. Unlike most systems modeled by physical science, these systems are typically open rather than closed or isolated – they permit the transfer of something (e.g., mass, energy or information[1]). One cannot provide an initial enumeration of the elements of these systems in such a way that will continue to suffice descriptively despite the system's evolution.

The model of such systems provided in this paper was first introduced by the author in 1975 in the context of modeling psycholinguistic processes and, more generally, human communication. In that paper, the concept of primitive semantic quanta was introduced. Process interactions mediated by such semantic quanta are represented in a semantic topology. Under suitable restriction, portions of this semantic topology (certain sub-topologies) provide the necessary foundation of a mathematical Hilbert space.

An Illustrative Model

Consider two (somewhat idealized) open and independent but interacting Systems A and B in a Universe $\boldsymbol{U}$, all evolving with time-like parameter t. Each system has a boundary δ, and a partially ordered set of finite inputs and finite outputs that move across the system's boundary. Each input element and each output element may have any number of properties or attributes associated with it.

Each system has a memory comprising at least two partitions. The first partition PR (for "partition raw") consists of memory of more or less raw inputs – i.e., that is not organized according to any interpretive operation. For technical reasons we will permit it to be a subset of all "actual" inputs – i.e., raw inputs may have been filtered.[2] The second partition PI (for "partition interpreted") consists of some portion of those elements stored in the first partition, but organized according to some principle or *operator* O such that relationships among them are established. In particular, elements represented in memory may be organized by establishing ordering relationships (i.e., some combination of partial or total orderings), collections (i.e,. partitioning by common attribute – a kind of ordering), or hierarchical (e.g., simple or multiple) containment relationships (again, an ordering relationship) among selected elements.[3] The operator O serves to provide a cumulative interpretation of the inputs over evolution of the systems. Note that I use the term evolution to indicate parameterized change with a time-like

[1] Of course, these are all equivalent in some sense.

[2] In a generalization, the "filter" operation may be extended so that it is permitted to augment the "actual" inputs.

[3] In a further generalization, there may be an arbitrary number of PR partitions each with its own filters, and PI partitions, each with its own operator.

parameter t, but not necessarily parameterized by physical time. For convenience of exposition, I will refer to this parameter simply as time t.

The output of an operator O is consistent with some invariant I, such that the states of PI (both initial and final) under O satisfy I up to some subset of PI which we will refer to as R. Conceptually, R is the remainder of PI after update by O and is that portion of PI that, given the current input, O cannot organize in such a manner that it is consistent with I.[4] Furthermore, we assume that O operates under a heuristic optimization principle that seeks to minimize R. In other words, if O can organize 'PI UNION *input'* in two or more ways consistent with I, it will select the one that minimizes R. Call this "the Principle of Least Semantic Disruption".[5]

In an important further specialization of these systems, assume that whenever R is non-empty, the system responds by obtaining (or accepting, inferring, generating, etc.) additional input. Mathematically, O is like a function that recurses whenever R is non-empty and otherwise halts. Call this "the Accretion Principle."[6]

Now consider an interaction between A and B. We will use some suggestive terminology, but with the caveat that it is neither rigorous nor constraining.

System A in state S_A0 sends output o as a message m1 to B in an attempt to resolve some ambiguity R_A in its state. The message m consists of elements of PI_A that have been derived from PR_A via O_A. (In an alternative generalization, we might consider this communication as writing asynchronously to and reading asynchronously from a shared memory.) Such messages are not instantaneous, but requires a time Δt comprising both a transit time from A to B (or B to A) and a processing time for assimilation by B (or A).[7] B, on receipt of m1, captures it as input in PR_B and then assimilates it in PI_B, resulting in state S_B1 with *ambiguity* R_B1. R_B1 is an ambiguity in the sense that its interpretation is unresolved in the context provided by PI_A and has the potential for multiple, equally valid interpretations (given what is knowable to System A).

System B's PI_B and R_B1 are now in an inconsistent state with respect to I_B. Metaphorically, O_A attempts to resolve this by sending a message m2 to A, consisting of selected elements of PI_B and R_B1. By the time m2 is received by System A, A has evolved beyond state S_A0, having had the sum of times Δt1 and Δt2 (associated with transition and processing m1 and m2, respectively) to receive other inputs and modify PI_A accordingly. In consequence, it is very likely that no content

[4] In a further generalization, R may include some part of the previous state of PI, subject to imposition of a higher level invariant that minimizes that unresolved part of the previous state of PI. For now, we assume that O simply integrates some or all of the input into PI and that R represents the portion of the input it cannot integrate consistent with I.
[5] Various types of failures of this heuristic may be seen as adverse mental conditions and possibly even of useful mental processes (e.g., non judgmental states).
[6] Of course, both the Principle of Least Semantic Disruption and the Accretion Principle can be modeled as operators.
[7] Note that these times might be, for example, relativistic times corresponding to causal interactions.

of message m2 from B will suffice to resolve the ambiguity R_A1. It follows that PIA evolves to a new state SA1 (typically not the immediate successor of SA0) with new ambiguity R_A2.

In this manner A and B are now in a position to engage in a potentially endless cycle of communication, driven by their individual invariants.

Observations on State Evolution

Now let's consider the sequence of states $<A_i>$ of PI_A and R_A as System A evolves given arbitrary inputs. (Note that the analysis which follows would apply equally well to System B or some other similar system.) Classify these states as either satisfying invariant I_A or not. If I_A is non-trivial, any arbitrarily selected state will be unlikely to satisfy it. We will see A transition from an initial state consistent with I_A through some number of inconsistent states until, whether by accident or design, after some number of inputs have been assimilated, consistency is obtained again. When consistency is obtained, there is no pending ambiguity, no multiplicity of possible interpretations. Not the similarity, at least conceptual if not indeed formal, to the quantum mechanical concept of collapse of the wave function. This quantum mechanical model (i.e., application) of the theory of interacting semantic processes becomes even more poignant with respect to systems with multiple invariants (see below).

Across the sequence of consistent states, it is possible to apply deductive reasoning, including predictive inference. These ordered states are mutually-consistent (in terms of the invariant I_A) and so may be understood as derived expressions (theorems) in a single axiomatic system. In other words, these states now represent *knowledge*.

Multiple Invariants

In realistic systems, there will be multiple invariants at play. For simplicity, let's examine what happens with two distinct (i.e., non-equivalent) invariants I1 and I2, corresponding to independent operators O1 and O2. Broadly speaking, there are two possibilities: I1 and I2 are mutually compatible or mutually incompatible. In other words, either there exists some state that will simultaneously satisfy both I1 and I2 or else no such state exists.

Suppose I1 and I2 are incompatible. In this case there will never be a state for which there does not exist some non-empty R with respect to either I1 or I2. The system never obtains a globally consistent state. Following the Accretion Principle, it is perpetually driven to assimilate more input in an effort to obtain the unobtainable.

Suppose I1 and I2 are compatible. Then, as the system evolves, there will be states that satisfy I1 and not I2, states that satisfy I2 and not I1, states that satisfy neither, and states that satisfy both. Only in the latter case could the system potentially halt. In the other cases, at least one if not both of the operators corresponding to I1 and I2 identify an ambiguity (R1, R2, or both). The system's response to this ambiguous state must generally be understood as resulting in separate[8] new inputs.

In general, a system may be more easily understood as interacting semantic processes having many invariants. It is often illuminating to identify a relationship among the invariants. For example, invariants may be given a relative priority – that is, invariants with lower priority are not examined unless all invariants of higher priority fail to be satisfied.

Even more useful are hierarchies of invariants, including simple and multi-hierarchies. These come into play when the primary semantic organization has (or is known or desired to have) containment relationships. In cases of a simple hierarchy, the lowest level of the hierarchy corresponds to inputs which are then organized into units at the next highest level, each of which is then classified (i.e., recognized or interpreted) at that level. Until the invariant at that next highest level is satisfied, the identity of the unit's class at that level is ambiguous – it may be any one of some subset of all the possible classes. Only when the invariant is satisfied at the higher level can the class be more positively identified. This relationship of among the levels of the hierarchy serves to provide coherence across the levels and absolute resolution only when all invariants in the hierarchy are satisfied. We will see an example of this in natural language recognition and generation at the syntactic level.[9]

Departures from the Ideal Process

The model of interacting systems must not treat them as infallible[10]. Various possible "errors" or variations need to be considered if the model is to be robust. In particular, we need to consider ways in which the output of A and input of B may be altered. These include:

- A may interpret, transform, edit, redact, or enhance what is sent. In other words, it may treat output as derived from PI and R.
- B may be interpret, transform, edit, redact, or enhance what is received.
- Either A or B or both may make errors.
- The transmission channel may introduce errors or transformations of the message.

Each of these is best modeled as intentional and perhaps arbitrary on the part of the systems (rather than accidental and even random). In particular, we ultimately treat all processes,

[8] They are theoretically asynchronously obtained, for example though separate channels or modalities. Synchronization should be modeled as an operator in its own right, possibly being discovered as a common time-like attribute of events under the various invariants in play.

[9] The semantic model would provide natural language understanding, but this is too complex to describe in detail herein.

[10] This is an unfortunate word, as it suggests there exists some correct way for these systems to interact. In fact, we are merely trying to consider all the ways in which such systems might depart from our initial approach, it being constructed in a particular manner for pedagogical reasons alone.

regardless of how poorly defined, as belonging to one of the interacting systems. These issues have not been addressed directly herein.

Comments on the Dynamical Model

At best, a message m1 from A as received by B will be assimilated only partially by B into its PI_B. In its effort to resolve the resulting ambiguity, B sends message m2 to A.[11] Note that this dynamical model is possible only as a partitioning of a closed Universe $\cup$ comprised of $A \cap B$ and NOT ($A \cap B$). It is an *artifact* of this analytical reduction.

The overall dynamics now iterates. A attempts to assimilate message m2, an ambiguity or mismatch is obtained, and message m3 is generated.

As further detail, consider that A and B are evolving separately. Thus, while m1 is being assimilated by B, A and in particular its PI_A are changing. The representation of PI_B + m1 will not, in general, be consistent with the new state of PIA as long as A continues organizing inputs from PR_A and modifying PI_A. Likewise, while A is assimilating m2 and generating m3, B is organizing inputs from PR_B and modifying $PI_{\mathbf{B}}$.

Invariants can be very complex with multiple sub-constraints. They can change over time. Even a simple invariant or "rule" can trigger interaction (i.e., satisfy the conditions for recursion).

Consider this in terms of a simplistic, artificial scenario involving human communication. Suppose that a cultural rule is that, whatever someone says to you, you must respond with one of the following:

- "I understand."
- "I don't understand."
- "Thank you."

It is easy to see that attempting to satisfy this cultural constraint can lead to endless, albeit informational impoverished, discourse. It is important now to generalize the notion of communication beyond speech inputs and outputs to general sensory inputs and motor activities, respectively.

Then we can generalize these to pure semantics – an information theoretic approach to interacting processes.

Semantics

In "Quantum Logic and the Semantics of Natural Languages" (1974) and "Relativistic Quantum Logic and the Dyanmics of Natural Languages" (1975), I explored the syntactic and semantic

[11] Alternatively, think of m2 as a general action on the (local) environment of B.

structure of natural languages in a radically new way. At the time, Noam Chomsky's theories regarding natural language were widely accepted.

About 1955[12], Chomsky had proposed that all human beings are born with an inherent predisposition toward language and have a genetically inherited universal grammar or syntax. The idea is that there are many natural linguistic expressions that mean roughly the same thing, i.e., they have the same semantic content though not the same syntactic structure. Chomsky proposed that a common syntax or "deep structure" underlies these different "surface structures" and that transformations generate a particular surface structure from the deep structure. Chomsky's posited deep structure, and variants suggested by others, reflected a least common denominator approach. Most often, their grammatical structure represented Boolean logic.

In the 1974 paper I demonstrated that this was not only inadequate but required an impossible type of transformation. In 1936, Garrett Birkhoff and John von Neumann[13] proved that it was not possible to derive a non-distributive lattice from distributive sublattices. On the other hand, it is possible to derive a distributive lattice from non-distributive sublattices. A similar result is easy to prove regarding deriving non-commutative lattices from commutative sublattices.

So, by finding examples of non-distributive and non-commutative structures in the surface grammar of natural languages, it is demonstrated that any deep grammar cannot be either distributive or commutative. In other words, any deep structure must, like the surface grammar, have characteristics of quantum logic. This result suggests that there is a much more intimate relationship between syntax and semantics than previously anticipated. It also suggested that there is semantics has a quanta-like character. By modeling meaning as quanta that are related in lattices having appropriate properties (i.e., non-distributive, non-commutative lattices), grammar then becomes the transformations among such lattices.

In the 1975 paper I made this proposal more explicit and described a theory of the dynamics of natural language – the generation and recognition of natural language. A key insight is that semantics pertains to many cognitive inputs (both internal and external percepts). It became clear that attempting to separate these into modalities according to either distinct senses or distinct affectors was artificial and folly. Studies in cognition had shown, and continue to show, that language, gestures, and thought are intimately bound together as cognitive processes. Furthermore, this coupling is bidirectional; what we see, hear, or think can cause involuntary neuromuscular activity and neuromuscular activity can influence what we see, hear, or think.

The resulting model is one in which the semantic quanta are organized into a discrete semantic space, inter-related by partial orderings. The organizing process is such that it is effectively reversible as a process (though individual applications may not be in a particular context) and so

[12] Noam Chomsky, "Syntactic Structures," © 1955 MIT Press.
[13] The Logic of Quantum Mechanics. Garrett Birkhoff; John Von Neumann. The Annals of Mathematics, 2nd Ser., Vol. 37, No. 4. (Oct., 1936). I don't recall if this particular paper – or a related one – contain the referenced proof.

may operate in a generative mode. This was yet another aspect of the theory that was contrary to the received view at the time: language recognition and language generation were thought to be distinct cognitive processes. This view was particular dominant in the AI community and computer programs attempting natural language recognition.

A Linguistic Example

From 1983-1988 a computer program called SENTAX was developed in collaboration with Prof. William H. Miller to test and refine aspects of the theory, in particular generation of natural language expressions in English. The program implemented a dictionary of English words classified into types and a multi-level hierarchy of production rules. Types were hierarchical.

This is perhaps best illustrated via a simplified example with three levels. At the lowest level words are classified as articles, nouns, verbs, adjectives, adverbs, etc. The next level contains phrases such as noun phrases and verb phrases. The top level contains sentences such as declaratives and interrogatives. Level one production rules specify possibly binary word sequences according which word type is permitted to follow a given word type. A sequence of word types, compliant with these rules, will belong to some type at the phrase level. Once a phrase is complete, production rules at the phrase level take over. At such a juncture, the number of permissible word level transitions (governed by production rules) tends to be large. Although constrained by phrase level production rules, there may still be multiple phrase transitions possible. Once the next word (i.e., the first word of the next phrase) is identified, the number of possible successor phrase tends to reduce. Similar, certain key words in the language determine whether or not the sentence will be a declarative or an interrogatory, and this identification at the sentence level constrains the permissible phrase level transitions. The number of permissible transitions at any point in the generation can be understood as a measure of ambiguity. The ambiguity tends to expand and contract as a sentence is generated, finally reaching a minimum when the high level of the production rule system fires a terminal rule.

In certain circumstances, such systems can predict with high certainty the exact transition that must next occur. In all cases, the number of permissible transitions is finite and so each such transition has a certain probability. These probabilities can be ranked and probabilistic predictions made.

SENTAX™ worked quite well even though it was implemented on a IBM PC under MSDOS which limited its complexity. Unlike other generative programs at that time, it was capable of generating recursive grammatical sentences. It could classify input, detect poorly formed sentences, and select from among random inputs to generate understandable sentences. It could avoid sentences like Chomsky's "Colorless green ideas sleep furiously."

SENTAX™ was designed to learn from its inputs inductively, although the research program was ended before this could be implemented and tested. For example, if a new word or phrase appeared consistently in a particular grammatical position, it would be classified appropriately.

Similarly, if a new sequence appeared consistently, the appropriate production rule(s) would be created. More interesting, although it may not be immediately apparent, the design could also partition its dictionary and rules into internally consistent schemes. For example, if the context were to transition to a different dialect, SENTAX™ was designed to detect that and partitioned the dictionary and rules accordingly (i.e., each partition being a class).

Other Applications

The theory of interacting semantic processes has many potential applications. These include:

- artificial intelligence – Systems that learn (organize) and anticipate (predict) transition in complex ways between coherent and ambiguous states, each of which may be context dependent.

- mental breakdown – Consider a model of a cognitive system S in which knowledge is assimilated as described above. Augment the model with a hierarchical super-invariant and corresponding operator that governs the Principle of Least Semantic Disruption. In particular, the purpose of this operator is to reorganize PI whenever R becomes excessive. Such a condition would suggest that the current organization of PI is not globally optimal, nor robust to the inputs now needing to be assimilated (i.e., organized). In logic terms, contradictions or inconsistencies begin to appear.

  This reorganization will consume an amount of time during which the "behavior" of system S will be indefinite. New inputs will not be readily assimilated and previously assimilated inputs will no longer have the same meaning – the same relationship. In other words, the state of system S will be fragmented and disassociated until the reorganization completes. Depending on the amount of ongoing "disruptive" (i.e., incommensurate) semantic input, this state may continue for long periods.

- diplomacy and negotiation - Interpretation under a particular set of invariants (a context) results in a perspective, allowing multiple perspectives to be modeled. Simultaneous satisfaction of the invariants corresponding to multiple perspectives subject to the Principle of Least Semantic Disruption is then the goal.

- computerized transaction management – Logically, a transaction is a transition between data states that are consistent according to some set of constraints or conditions, i.e., an invariant. Conceptually, the inputs and outputs of a transaction may be understood as communication with another system. This approach enables a flexible approach to transaction management without pre-determining all constraints, in effect permitting transactions to emerge where necessary.

  - A set of conditions satisfied by a group of operations define a transaction.

  - The class of transactions that satisfy the same set of conditions define a system.

  - The inputs and outputs of those transactions establish interaction with other systems.

- economics – Economic systems that exchange assets (e.g., goods and services) may be understood as interacting processes. Exchanges are not zero-sum, resulting in aspects of the exchange that must be resolved through further exchanges.

- physics – Partially ordered events (i.e., collections of observables) in physics may be understood as transitions between states that satisfy invariants (conservation laws or symmetries). Causal structure and therefore 4-space is inferred from this semantic, topological, pre-geometry rather than being an *a priori* physical construct.

- business – Businesses may be modeled as process sub-systems that interact with each other, and with a variety of external systems. This leads to an understanding of business in terms of business processes, each process having an objective or goal (invariant), and composed of business activities having sub-goals (sub-invariants). Each business activity is then a process in its own right, interacting with other processes. This hierarchy of business processes may have many levels. Inputs are resources (e.g., materials, machines, people, information, etc.) and requirements (e.g., new orders, or quality mandates) and outputs are measurements, goods, services, and "downstream" requirements.

Many of these systems require quantized or at least discrete mathematical models. Otherwise, continuum mathematics is an approximation that can introduce cumulative errors. It was for this purpose that the Ordering Operator Calculus was, and continues to be, developed.