Mathematical Textual Analysis

Daniel Nikpayuk

February 26, 2014

Here, I propose a mathematical definition to model the Humanities' *critical approach* to textual analysis. This set-theoretic definition is multilevel (complicated), and so I will build up to it—explaining intuitively along the way the motivation.

In the Humanities, as a template approach to critical analysis we would begin by analyzing the *context*, then its *semiotics*, finally the *media*[1]. It is right here—at the beginning—worth pointing out the difference in philosophical background taken by humanists and mathematicians: **phenomenology**.

I do not claim expertise as a philosopher, but to be so bold, I would say we need to look at the question of existentialism. Descartes, who was a mathematician, logician, philosopher, is famous for the statement: "I think, therefore I am". As it turns out, even if you have (like Descartes) philosophically convinced yourself that you exist, how do you prove the existence of the outside world? For example, how do you prove this computer in front of you exists?

There are differing philosophical approaches, but phenomenology basically says: It's of no consequence whether or not *it* actually exists; it is enough that you have in your own mind a representation of it; and since you exist, the representation in your head of it exists. Basically, phenomenology sidesteps the issue altogether.

I bring this up because mathematicians are phenomenologists. We study language in its own right— mechanical language. We have no interest in the actual context to which our studied languages are applied; we simply study the languages themselves. We study systems of representations of things. As such, although it is not a formal part of the following definitions, I will state my assumption and viewpoint that whatever the context of discourse is interpreted to be, there is a distinction between the *intuitive* understanding of context, and the *referential* understanding of context.

With the intuitive understanding of context, we are making a *declaration*; and as with any meaningful but intuitive declaration, one can pull out inexhaustible referential meaning from it. As example, I declare the word "music" but leave it intuitive. If you are familiar with this word, I'm sure you can think of many referential song names, artists, labels, descriptions, connotations, associations, etc. having to do with this word; but even if you put together a list of every representation used to interact with this intuitive idea, you will not have captured it in its entirety. The same goes for—significant—words such as "love" and "death" and. . . and of course there are many more other such declarations. . .

This same stance—distinguishing between intuitive and referential understanding—is taken when looking at semiotics and media as well. This is to say, there are intuitive versions, and there are formal versions. As we are currently in the realm of math, we are—strictly speaking—only working with the semiotics of it all. A question might arise to the reader: "If we only have semiotics to work with, how are we using those signifiers to represent signifiers?" Or you might say: "All signifiers exist to talk about the context!" But as

---

[1]I'm sure by now I'm already making some humanists cringe with my problematic generalizations. My apologies, if that is the case.

it stands, when such a system becomes complex, when we build up enough signifiers that they overwhelm us, we then need user-friendly means of interacting with all those signifiers. We end up with signifiers that are used to navigate and talk about those simpler context oriented signifiers.

The point being, if we were to classify these signs, signifiers, signifieds—that represent the whole of our mathematical universe—by means of their intended use, we could still identify some as meant to be used strictly for the critical context; others strictly as critical semiotics; and others still to represent the critical media. This is the type of viewpoint taken in this article.

As such, we will start by formalizing our intuitive idea of a *critique*, which is a specialization of a slightly broader concept:

Technology Space: A **technology space** is defined axiomatically (recursively) as follows:

1. A context space is a technology space;

2. A semiotic space is a technology space;

3. A media space is a technology space;

4. Any technology space consists entirely of a context space, a semiotic space, and a media space.

The interpretation and implications of this definition need to be drawn out. Regarding the interpretation, if you're not use to this style of definition, do not fear, it is not circular: First of all, as it is an axiomatic definition—it declares the concepts of context, semiotics, and media, but it otherwise leaves them undefined. Secondly, anything we declare to be a technology, we know by this definition, that it can be broken down into its context, semiotics and media. This definition is recursive because each of these spaces are by definition their own technologies, and so each will have their own context, semiotics, and media. In turn, each of these may also be broken down into their respective spaces; and so on; and so on. It never ends.

You might say: "What's the point of having a never ending definition?" It's not that it's never ending; rather it is well defined and modular—though certainly vague at this point. Think of it like pi: $\pi \sim 3.14159265358979\ldots$

Pi is never ending, but people make use of it all the time; no one complains (well...some still do...). We make use of approximations of π, and if a certain approximation isn't good enough for a particular application, we take a more refined approximation. Simple.

So, wrap your head around that for a little bit, but once you do, it's time to move on. In our theory here, a "critique" is nothing more than an actual instance of a technology space: An application of a technology space with all the missing details filled in—or as many filled in as are sufficient.

Regardless of the clarity, simplicity, expressivity of this definition, it is I admit: Still Very Vague (SVV). There are a lot of missing steps from turning this definition into any actual critique. So what are those missing steps? Or at least what are they in an abstract template sort of sense? That is the content of the remainder of this article. Without going into the details just yet though, the basic idea is that we're looking for a more refined template definition, but we still want it to be broad enough that it's applicable to as many real life texts as possible. Now, as we are working within the realm of formal mechanical language—mathematics—we should thus use a template that provides the greatest chance for creating future possible critiques.

It would make sense then to take a branch of math that can be used as a model of all other branches of math. Why? Because any foundational branch that can model all other known branches of math would maintain that greatest entropy for possible future outcomes we're looking for. We are after all using a technology space to model the design of a more specific critique. We're just looking for a few extra middleway details to ease the translation. With that said, as I have mentioned at the beginning, I am using *set theory* as the framework to provide this refined but still abstract definition of a technology space.[2]

We begin now by looking more closely at the deeper meaning of context:

---

[2]Category theory among others is a perfectly good candidate as well (as you will see), only I have chosen set theory because the most natural application of a technology space I can think of is "subsetting" and so set theory is a natural fit.

Context Space

As a context space is its own technology space, it has its own context, semiotics, and media. What exactly—within the framework of set theory—does this mean?

Context Space - Context: Let $\mathcal{L}$ be some initial set of interest. By convention, we call it a **lens**, and the following set

$$\mathbb{C}(\mathcal{L}) \quad := \quad \bigcup_{k \geq 1} \mathbb{P}^k(\mathcal{L})$$

will be called the **context** of $\mathcal{L}$. Any given subset of the context will be known as a **subtext**.

If it is clear what is meant by $\mathcal{L}$ in some application, we will simply refer to its context space as $\mathbb{C}$.[3] Furthermore, when discussing the origins of a given subtext, it might be useful to refer to a given $\mathbb{P}^k$ directly; as such we will refer to them as *pillars* of (contextual) infrastructure—in particular a given $\mathbb{P}^k$ will be referred to as the kth pillar.

So we have what we mean by the context of the context space. What then would be the semiotics of this context space? Signifiers are used to represent the context (the signified), and if our diversity of context equals all the possible subtexts, we could represent each of these directly, but often one will want to modularize upward: We will in practice want to represent more complicated subtexts by means of combining and abstracting simpler subtexts. We do this by means of filters:

Context Space - Semiotics: Let $\mathbb{PC}(\mathcal{L}) := \mathbb{P}(\mathbb{C}(\mathcal{L}))$ be the powerset of the context (the collection of all subtexts of the context) with respect to its lens. We define the **filter** set with respect to a given lens as follows:

$$\mathbb{F}(\mathcal{L}) \quad := \quad \{ \quad f : A \to B, \quad A, B \subseteq \mathbb{PC}(\mathcal{L}) \quad | \quad \text{for all } a \in A, \quad a \supseteq f(a) \quad \}$$

Any element within this filter set is called a **context filter**. In particular, any filter $f : \emptyset \to \{\mathbb{P}^k\}$ such that $k \geq 1$, is called a **power filter**.

There's a lot to take in regarding this definition, so it's worth a bit of pondering.

Basically, we have collected together all mappings that take a subtext as input (this could also be the context itself [as subtext]), and returns a subtext of itself as output. We are calling these *filters* because we are filtering out subtexts. What's more, each such filter shouldn't have to shoulder-the-burden of being able to take all possible subtexts as input—it might be specialized. This consideration additionally informs the above definition: This is the reason A, B are subsets of the powerset rather than being defined as the whole powerset themselves.

Next, by logical technicality (use of the qualifier "for all") this filter set admits membership to filters of the form $f : \emptyset \to B$ which is why I have given some of them their own special designation. If you think about it, this also fits the philosophical distinction between intuitive and formal, the one I began this article with. Initially, we would start by pulling out a formal context from the more ambiguous intuitive context. This intuitive context, as it is not representable in any tangible way, is naturally represented here as the empty set (the set with the greatest entropy; the most ambiguity).

So we have now identified the basic mechanism for furthering our discourse of subtexting. The next step in our contextual analysis is in mediating our subtext representations. Since our ability to represent the context is by means of subtexting—informed by filtering—we look now to mediate our filters. The first thing to note is that our filters are nothing more than functions in this framework, and so it would not be unnatural to consider taking compositions of filters ← and to that end, in taking chains of compositions of filters. This leads to our next definition:

Context Space - Media: The **hermeneutic** set, respective of a given lens and its context set, is defined as follows:[4]

---

[3]This of course will not be the case if using this theory to talk about the system of complex numbers.

[4]You might protest that this definition is additionally dependent on a filter set, but for any given lens its corresponding filter set is as a consequence of its definition unique.

$$\mathbb{H}(\mathcal{L}) \quad := \quad \{ \quad <f_k>_{k=0}^n, \quad f_k \in \mathbb{F}(\mathcal{L}), n \in \mathbb{N} \quad | \quad f_k \circ f_{k-1} \text{ exists for each } 1 \le k \le n \quad \}$$

where each sequence as element is termed a **hermeneutic**.

The first consideration worth noting is that we are—in the general sense—more interested in the *record* of subtexting than in the final subtext composition itself. These sequences of filters are in their formal way meant to represent the intuitive idea of "an interpretation". Hence the name *hermeneutic*. Beyond that, the general intuitive idea of media is to mediate conditions—conditionals?—and constraints upon the distribution of messages created by the semiotics (all with respect to the context of course). As we are working within a set-theoretic framework, and our semiotics have already been restricted to functions, our options toward the nature of the media itself is already constrained. By keeping the record rather than only the final composition, we have not *un*necessarily constrained this media template more than we've had to.

Now that we have introduced the context, its filters, and its hermeneutics, it's time to put them together:

Context Space: A **context space** $\mathbb{CS}(\mathcal{L})$ is a triple of a context $\mathbb{C}(\mathcal{L})$, its filters $\mathbb{F}(\mathcal{L})$, and its hermeneutics $\mathbb{H}(\mathcal{L})$.

The final note worth mentioning in this section is the combinatorial nature of the context space. The filter set is as large as it can be, it has the highest entropy in regards to future possible outcomes of filters defined by it. The same is the case for the hermeneutics; it has the restriction that its chains of filters must be composable, but it otherwise still allows for all possible combinations.

In practice a hermeneutic set ends up being quite "full"; it ends up being too full and unwieldy actually, to be of advantage to the design of actual textual analyses—hence, applied critiques will have much smaller grammars restricting the possible range of hermeneutic interpretations. More so, the *user* of a critique will not interact with these hermeneutics directly, rather she uses compressed forms and abstractions by means of the main semiotic space.

Semiotic Space

At this point I admit this theory runs a bit thin. Kind of a let down. I know.

Realistically I am still negotiating the design of an abstract "algebra" of filter operators with real life applications in software design. I have some basic ideas, but nothing I'd like to put down on digital paper just yet. In anycase, I will move forward with a vague outline of where I intend to head.

Semiotic Space - Context: We take as our initial universe of semiotic discourse a non-empty set which we will by convention term as a **symbolic** set $\mathcal{S}$. Its elements we will term as **symbols**.

Though in an abstract sense this symbolic set is by definition complete, in practice it will generally come with some additional constraints and/or operators to help us structure and reference it internally. This will be the semiotics of our main semiotic space, but as stated above, this is still an undergoing area of research. In anycase, I can name two symbols which will likely be used:

We will first and foremost set aside and make use of two symbols which by convention do not belong to the symbolic set: '∪' and '∩'.[5]

With this practical internal structuring in mind, this set of symbols will have the following property:

Symbolic Union: If a, b are symbols in $\mathcal{S}$, then the symbol '$a \cup b$' is also in $\mathcal{S}$.

Symbolic Intersection: If a, b are symbols in $\mathcal{S}$, then the symbol '$a \cap b$' is also in $\mathcal{S}$.

Really, this is a convenience more than anything else, and even though semantic union and intersection are *implied* by the use of these symbols, please keep in mind that as of yet, the semiotics and the context are independent. It is the media that gives these two their relationship, but until then, you cannot assume any actual connotations—strictly speaking of course.

Alas, although still very vague, it is worth at least encoding the idea of the semiotics of the semiotics:

---

[5]In reality, these two symbols are arbitrary, they could be anything, but for our purposes here we will stick with the standard symbols for the set-theoretic union and intersection operators.

Semiotic Space - Semiotics: The semiotics that allow for the structuring of the symbolic set will be referred to as the **operator** set, $\mathbb{O}(\mathcal{S})$ with its elements being operators.

Even though we cannot make any assumptions about exactly how the context and the semiotics will be connected and related, we can at least prepare for the time when that comes:

Semiotic Space - Media: The **expression** set is the collection of "strings" of symbols:

$$\mathbb{E}(\mathcal{S}) \quad := \quad \{ \quad <s_k>_{k=0}^n, \quad n \in \mathbb{N}, \quad s_k \in \mathcal{S} \quad \}$$

Keep in mind this expression set is combinatorial in nature: it is all possible strings. I'm getting ahead of myself here, but when the *media space* is introduced, it will map *symbols to filters*, and by natural extension it will map expressions to hermeneutics. Yet, if we are to use these expressions to reference hermeneutics, given their full combinatorial nature, a discrepancy shows up: There are expressions which have no natural map to any hermeneutic. To put it another way, if a string of symbols represents a chain composition of filters, we would—from a natural correspondence—be interested only in those combinations of symbols where chained subsetting actually makes sense.

If this is the case, why include a broader range of expressions than hermeneutics in the first place? Because this sort of thing shows up in the real world: There's no reason to exclude it here unless there is known good reason to (entropy: "don't be wasteful"). I've brought this up as a reminder as to a common pitfall of language users who do not realize *this* subtlety of textual analysis and end up with false claims for their efforts.

In any case, we can now bring our symbols and expressions together:

Semiotic Space: A **semiotic space** $\mathbb{SS}(\mathcal{S})$ is the triple of a symbolic set $\mathcal{S}$ together with its operators $\mathbb{O}(\mathcal{S})$ and its expressions $\mathbb{E}(\mathcal{S})$.

The cumulative elements of these three sets from within this space are our **signifiers**, and sometimes even our **signifieds**. In anycase, with this we are finally ready to begin looking at bringing together our context and semiotic spaces.

Media Space

At this point I would hope it's somewhat obvious to the reader the theme of this article—I will hence streamline the following definitions:

Media Space - Context: A **denotation** $\mathcal{D} : \mathcal{S} \to \mathbb{F}(\mathcal{L})$ is a mapping from a collection of symbols to a collection of filters (with respect to a given lens).

As I have already (though informally) introduced symbolic union and intersection as a practical consideration, I will explain here that a denotation in practice will preserve these operations. This is to say:

$$\mathcal{D}(a \cup b)(x) = \mathcal{D}(a)(x) \cup \mathcal{D}(b)(x)$$

$$\mathcal{D}(a \cap b)(x) = \mathcal{D}(a)(x) \cap \mathcal{D}(b)(x)$$

Note, although not formally stated, a natural constraint is the two filters in question must take the same function domain to be compatible.

As much as this is the case for these operators, I should mention one other operator:

$$\mathcal{D}(a \uplus b)(x) := \{\mathcal{D}(a)(x), \mathcal{D}(b)(x)\}$$

which takes two filters and collects their output into a single pair. This operator provides a way of moving across *pillar boundaries* as per the context. Again this is informal. There will be a need for an operator that lets us "collect" sets together, but as of yet I have not worked out a proper definition that is not problematic.

In anycase, with a denotational mapping from symbol to filter at hand, the next step would be to define a mapping from *expression to hermeneutic*. Here's the thing though: If we have a denotational mapping

in front of us, it tends to want to generate a *natural mapping* from at least some of its expressions to its hermeneutics. This is so because let's say a, b, c are symbols, and "abc" is an expression,[6] if we know that

$$
\begin{aligned}
a &\rightarrow f \\
b &\rightarrow g \\
c &\rightarrow h
\end{aligned}
$$

then wouldn't it be natural that

$$
abc \quad \rightarrow \quad h \circ g \circ f \qquad ?
$$

This is what's meant by a "composition preserving mapping" in the following:

Media Space - Semiotics: Given a denotation $\mathcal{D}$, a **correspondence** $\mathcal{N}_{\mathcal{D}} : \mathbb{E}(\mathcal{S}) \rightarrow \mathbb{H}(\mathcal{L})$ is a composition preserving mapping (with respect to the denotation) from an expression set to a collection of hermeneutics.

First note that even though a correspondence is *composition preserving*, nothing is said about *unnatural* expressions—expressions which exist in the expression set but don't correspond to the natural extension of the denotation. They can map however they want to be mapped. Also notice one particular limitation of a correspondence: It must map from the whole of the expression set. In the practical situation, this is unnecessarily strict, and in the following we will relax that restriction through the use of *restricted functions*.

Let's recap: Within the discourse of a media space, denotations (the exact "wiring" arrangement of our speaker-system as example) are our context of interest. Nothing is said about having to use the whole symbolic set to define our denotations, which greatly expands our context of media. To represent these denotations then, we look at the correspondences generated by "full" denotations. We would build up more complicated restriction functions from simpler ones to better approximate and access the true nature of our correspondences—an algebra of restrictions. finally we look to restrict such correspondences on their domain to approximate and represent the less easily accessible denotations.

In the broadest sense, a grammar to represent functions within set theory suffices for these purposes. With this said, category theory might be a better framework to take a closer look at this grammar of correspondences—or at least ideas borrowed from it might save from reinventing wheels, or speaker-systems.

My reality is—although I've written this much of the article and it is likely to stabilize—this final piece of the puzzle is still too experimental to pin down with complete certainty. The packaging of these underlying ideas may still need some negotiating.

Media Space: A **media space** $\mathbb{MS}$ is a triple, the first of which is a set of denotations. The second of this triple is a set of restrictions of correspondences generated by full denotations. Lastly, the third of the triple is as of yet an unnamed set which provides a language to this media.

The nature of this definition, aside from being still very vague, is that there are many possible media spaces with respect to a given context and semiotic space.

As far as this introduction goes, I have nothing else to say but *thank you* for taking the time to read it.

The End.

---

[6]More explicitly this would be represented as $< a, b, c >$.