

# All recordable human discourse is trapped in aleph-zero

Oli Sharpe  
Go Meta  
Brighton, UK  
oli@gometa.co.uk

## Abstract

Borges' short story "The Library of Babel" is a classic literary exploration of the idea of a combinatorial library that contains all possible books of a certain format. However, this idea can be expanded into a theoretical enumeration over all possible recordable attempts by humans to communicate. From this we can deduce that the fidelity with which humans can refer is 'only' countably infinite, which is the smallest infinity known as aleph-zero. This paper constructs this enumeration and explores two consequences of it. Firstly, it is at least possible that the size of the set of 'true' things in the universe is a 'larger' infinite, such as aleph-one (as suggested by the diagonal arguments by Cantor, Gödel, Turing). If this is the case, then it would be impossible for even the full extent of our theoretically possible recordable discourse to explicitly refer to each thing that is 'true' about the universe. Maybe there are unquantifiable and undefinable aspects of the world that we cannot capture in any recordable discourse, let alone in the specialised discourse of programming. However, such an expressibility gap would be between humans and the universe, not humans and computers. Secondly, this paper looks at how this enumeration gives theoretical support for certain uses of unique identifiers in programming languages, such as Semprola.

## 1. Introduction

In this paper we are going to look at a particular limit on the fidelity with which humans can communicate. The reason we are interested in communication, rather than thought, is because the long-term value to humanity of a thought depends on our ability to communicate the thought to others.

And the reason for looking at this communication fidelity limit is to compare it to both the possible 'fidelity' of the universe and to the similar limit on the fidelity of what computers could communicate.

To examine this limit we will be building enumerations (complete, ordered listings) over sets by constructing indexes that give each member of the sets we're looking at a unique identity, a 'UID' within that set. The last part of the paper will then link the discussion about the limits of communication with the use of UIDs within programming environments.

## 2. Cataloguing Communication

One way to think about a book is as an attempt by the author to communicate an idea to all potential readers of that book. A library is therefore an attempt to collate together a fairly comprehensive catalogue of all such attempts to communicate using books. But how comprehensive could this library be?

### 2.1 The Library of Babel

In "The Library of Babel" Borges (1941) imagines a universe constructed as a single library with an indefinite number of identical looking hexagonal galleries holding books. These hexagonal galleries are stacked upwards and downwards as far as the eye can see with a shaft of air between all the floors and only a balcony to walk around to reach the books on each level. Four of the walls hold the shelves of books while the other two<sup>1</sup> walls have passageways through to yet more of these hexagonal galleries. A door on each such passageway reveals a staircase to reach the upper or lower levels.

<sup>1</sup> In the text Borges actually says only one wall has a passageway, but many readers assume that this is a slight editing error from what Borges intended, in particular see: <https://libraryofbabel.info/theory.html>

Each hexagonal gallery contains 700 books arranged and constructed as follows:

*“There are five shelves for each of the hexagon's walls; each shelf contains thirty-five books of uniform format; each book is of four hundred and ten pages; each page, of forty lines, each line, of some eighty letters which are black in color.”*

The character set used for every book includes twenty-two<sup>2</sup> letters of the alphabet and the space, comma and full stop characters. The library contains every possible 410 page combination of these twenty-five characters. Given that blank pages can be constructed entirely out of the space character, this means that every piece of text of length up to 410 pages is held somewhere in this library, including this paper (albeit with a simplified character set). Indeed, in the Appendix below there are several examples where the abstract for this paper has been found in random books in an online version of the Library of Babel. Somewhere on its shelves the library also contains all drafts and reviews and rebuttals of this paper as well.

By imagining the combinatorial library as a physical space with people attempting to make sense of the books they find as they wander around this seemingly endless universe of books, Borges manages to convey the complete lack of utility of such an exhaustively comprehensive library. Picking up and reading a random book from the library is essentially pointless.

Most of these random books are unintelligible, and even those that can be read have no real meaning as they have no author attempting to directly or indirectly communicate an idea. Without an author there is no intention behind any of the text that happens to appear on the page, even if the text has the appearance of conveying an idea. We know that for any such apparent idea in any of the random books, there will exist another book in the library with the exact opposite idea, or with a convincing counter argument to the first idea. There is no communication going on, all these books just happen to exist.

Borges' short story therefore makes clear just how fundamentally important the history of a given text is. For any text to be a genuine attempt to communicate a meaning there must be the intentions of an author behind the text.

Note that the 'author' of the text may have generated the text via a process of some kind (e.g. recording the temperature every hour), but the process must be of the kind where the semantic intention of the author is sufficiently preserved to ensure that the text itself has meaningful content. So, while the automated generation of random books may be achieved by a process created by an author with an intention to communicate some general idea about libraries of random books, the text of each particular random book thereby generated does not individually convey this general idea.

And of course there is a more complex potential example where the human 'author' has created a sophisticated artificial intelligence (AI) which in turn is generating text in order to convey ideas that the AI 'intends' to communicate. In this case we'd take the AI to be the author of the text and thereby see this as a very different case from the situation where a random text generator happens to have created text that appears to have meaning.

As all of the books in Borges' library are randomly generated therefore they are devoid of specific semantic content and so it completely fails to perform the function of a real library.

## 2.2 Creating an index for an enumeration

None the less, the conception of Borges' library gives a nice example of how a comprehensive set of things can be constructed combinatorially thereby making it easy to conceive of an enumeration that exhaustively covers the set.

In the case of the Borges' library we can first give each of the 25 characters in the used character set a number between 0 and 24, and then give each character position in a book a unique number  $p$  where  $p = \text{page} \times \text{row} \times \text{character in row}$ , (with these positioning dimensions ranging from 0 to 409, and 0 to

<sup>2</sup> Borges uses a reduced character set for the construction of his library as explained in his essay, "The Total Library". For more information about this see: <https://libraryofbabel.info/theory3.html>

39 and 0 to 79 respectively), so that each book is considered to be simply a string of 1,312,000 characters. Now we can calculate the unique index ID number for each book in the library as being:

$$\text{Library of Babel unique ID number (LBUID)} = \sum_{p=0}^{1,312,000} c_p \cdot 25^p$$

Where  $c_p$  is the number for the character in the  $p$ th position within the given book. Note that this number can not only be used as a catalogue number for the given book, but also contained ‘within’ the single number is an encoding of the entire content of the book itself.

One thing that this formulated enumeration makes abundantly clear is that Borges’ library is in fact not infinite in size<sup>3</sup>. There is a largest book index number.

However, in the next section we will look at how we can extend this notion of a combinatorially constructed ‘library’ to include an infinite number of books (strings) of arbitrary length and then extend it yet further to include all forms of recordable communication.

But, before that, let’s briefly look at the different kinds of infinity that are referred to as aleph-zero and aleph-one.

### 2.3 What is aleph-zero?

The set of natural numbers,  $\mathbb{N} = \{ 0, 1, 2, 3, \dots \}$ , has no highest number and so if you kept counting these numbers you would continue ‘forever’. The size of this set is therefore not finite, but infinite.

Mathematicians have noticed that there is a whole class of sets that can be mapped one-to-one onto this set of natural numbers, so we can know that these other sets are no bigger than and no smaller than the size of  $\mathbb{N}$ . In other words, these sets are the same size of infinite. An important example of this is the set of rational numbers,  $\mathbb{Q}$ . The one-to-one mappings are usually achieved by creating an enumeration over the set providing a unique index number (in  $\mathbb{N}$ ) for each member of the set whose size we wish to compare to  $\mathbb{N}$ . This is why we created an enumeration over the books in Borges’ library and will continue to look at enumerations in later sections of this paper.

In 1891 Georg Cantor published his diagonalization proof (Cantor, 1891) that shows there can be no such enumeration from the set of real numbers,  $\mathbb{R}$ , to the set  $\mathbb{N}$ . The set  $\mathbb{R}$  contains more things than the set  $\mathbb{N}$  and therefore the infinitely large size of  $\mathbb{R}$  must be strictly larger than the infinite size of  $\mathbb{N}$ . The suggestion from this is that there are different ‘sizes’ of infinity!

Aleph-zero,  $\aleph_0$ , is the name of the ‘smallest’ infinity used for countably large sets such as  $\mathbb{N}$ .

Aleph-one,  $\aleph_1$ , is the name for the next largest infinity and there are larger infinities like  $\aleph_2$  and so on.

Cantor showed that the size of the set  $\mathbb{R}$  (denoted  $|\mathbb{R}|$ ) was  $2^{\aleph_0}$  and proposed that there are no sizes of infinity between  $2^{\aleph_0}$  and  $\aleph_0$ , which is equivalent to suggesting that  $\aleph_1 = 2^{\aleph_0}$ . This is known as the continuum hypothesis and has yet to be proved.

What is known is that  $|\mathbb{N}| = \aleph_0 < \aleph_1 \leq 2^{\aleph_0} = |\mathbb{R}|$  and in this paper we simply depend on the fact that there are sets with size strictly greater than aleph-zero, and we will refer to these as simply being of size aleph-one or bigger.

### 2.4 Extending to all Unicode texts

The first way to alter our enumeration is to extend the character set from Borges’ 25 characters to the full set of Unicode ‘code points’, which are almost like characters but not quite (Unicode, 1991). Indeed, we’re going to have the index range for these code points to not only cover the existing 137,000 or so code points, but the complete range of theoretically available code points, which is 1,114,112. Similarly, rather than limit the number of code points in our ‘books’ we will instead think of strings of unlimited, but finite length.

Therefore, pieces of text of any arbitrary length using any Unicode compatible character set will be covered by this enumeration. Any such string can be given a unique index number as follows, where  $n$

<sup>3</sup> As Borges himself notes in the short story itself.

is the total number of code points in the given string and  $c_p$  is the code point value for the  $p$ th code point in the string:

$$\text{Unicode string unique ID (USUID) number} = \sum_{p=0}^n c_p \cdot 1,114,112^p$$

Unlike the enumeration of books in Borges' library, this enumeration of Unicode strings does continue for ever as  $n$  can be arbitrarily large, making this set of strings aleph-zero infinite in size.

We can construct an alternative enumeration for this same set by encoding these Unicode strings in UTF-8 format as a series of 8-bit binary numbers. This enumeration would be formulated as follows:

$$\text{UTF8 (file) unique ID number (UTF8UID)} = \sum_{p=0}^n c_p \cdot 256^p$$

Where now  $c_p$  is the number (0-255) represented by the  $p$ th byte in the file.

## 2.5 Extending to all binary encoded recordings

Clearly the last enumeration could be used for any binary file, but we happen to have decided that these are UTF-8 encoded files. What if the binary file encoded a video, or any other digitally recordable act of communication (such as a song, or video, or vector graphic, or motion captured gesture or whatever). To create an enumeration to cover all different kinds of encodings of binary files, we can combine together two enumerations using the fundamental theorem of arithmetic that any number is the product of a unique combination of prime numbers.

So, if we imagine an enumeration over all known encodings, then we could assign any particular encoding a unique '*encodingUID*' number (for example 0 = UTF-8, 1 = UTF-16, 2 = MPG4, and so on). Then, using a similar enumeration to that used for the UTF8UID numbers above, we can imagine giving every possible finite length binary file a unique '*binaryFileUID*'. We can then combine these two enumerations to give a new enumeration over all encoded binary files:

$$\text{Encoded binary (file) unique ID (EBUID)} = 2^{\text{encodingUID}} \cdot 3^{\text{binaryFileUID}}$$

The purpose of using this enumeration, rather than just the binary file enumeration is so that each index number 'knows' which encoding is being used by the binary file, therefore this enumeration keeps sufficient semantics about the recorded information that it could be decoded appropriately.

Note that we can imagine this enumeration being extended to cover any arbitrary (but finite) fidelity of recording. So, for example, every conceivable 4K high definition, 100 frames per second 3D, surround sound video is included within this enumeration as long as the video is of finite time duration. And the enumeration also contains all finite higher and lower fidelity copies of every conceivable finite time duration video.

So, we can now imagine a new audio visual (and more) 'library' that doesn't just contain books but contains multiple higher and lower fidelity copies of every conceivable way to record information. For our purposes here, we'll call this the "exhaustive digital library".

## 2.6 Why is finite fidelity enough?

A key detail in the enumeration of this exhaustive digital library is that it contains all *finite* fidelity copies of any recordable information. Without this limitation it wouldn't be possible to construct the enumeration. So, how do we know that these finite fidelity recordings are sufficient to capture any human attempt to communicate?

Well, any such attempt to communicate must be perceivable by another human and we know that the human ability to perceive and discriminate stimuli has finite limits. Therefore, for any piece of recordable information there is a finite level of fidelity of digital recording at which no human could notice a loss in information conveyed by the recording.

This is a fundamentally important observation for this paper as it is this that allows us to know that there will be no recordable human communication that cannot be sufficiently represented by some entry in our exhaustive digital library. So our digital library that is exhaustive by construction has also been shown to comprehensively cover everything we wish to hold in such a library.

## 2.7 An enumeration over all recordable human communications

Now that we have an enumeration that gives anything recordable a unique EBUID, we can combine this with an imagined enumeration over all humans that will ever live<sup>4</sup> (giving each one a unique ‘*humanUID*’) and another imagined enumeration over all milliseconds since the big bang (‘*millisecondTimeCounter*’). Putting these together we can formulate an imagined index of Recordable Human Communication Unique IDs (RHCUIDs) for all conceivable acts of recordable communication by any human<sup>5</sup> ever:

$$\text{Recordable Human Communication UID} = 2^{\text{humanUID}} \cdot 3^{\text{millisecondsTimeCounter}} \cdot 5^{\text{EBUID}}$$

The above description of the RHCUIDs gives a basic proof by construction that such an enumeration would in theory be possible by a god-like observer of the universe.

Furthermore, the existence of such an enumeration over all conceivable recordable human communication demonstrates that the total cannon of all actual human communication that will ever happen is also no larger than aleph-zero.

Indeed, assuming that there will be a largest *humanUID* (even if this enumeration includes the evolutionary descendants of humans) and a highest *millisecondsTimeCounter* value in which humans exist then there will only be a finite number of actual attempts by humans to communicate. Also, during all of that time that humans exist there will be an EBUID with the largest, finite size that would be needed in order to have recorded each communication act by humans in sufficiently high fidelity so as not to have any noticeable loss in any of the information being communicated.

In other words, assuming there is a “last human”, then to faithfully record all human attempts to communicate ever will ‘only’ require a finite description length! (the ‘only’ is in quotes because this finite number will obviously be extremely large).

### 3. Our recordable discourse is trapped in aleph-zero

So, what is the purpose of creating all of these rather absurd enumerations? As with Borges’ library these enumerations are of no practical use. However, even if we stay neutral on the question of whether or not there will be a “last human”, we can now confidently make the following three statements:

1. The total number of recordable communication acts that are actually made by all humans ever will be in aleph-zero (where “in aleph-zero” means less than or equal to aleph-zero).
2. The total description length of a set of high fidelity recordings of all these actual communication acts will also be in aleph-zero.
3. Even the description length of all *conceivable* high fidelity recordable content that humans could ever produce is in aleph-zero.

Hence, all conceivable human discourse, everything expressible by humans is ‘trapped’ in aleph-zero. Or in other words, the fidelity with which humans can refer is only countably large.

But why is this fidelity limit worth noting?

#### 3.1 Can’t we refer to some real numbers like pi?

The possibly surprising thing about our discourse being trapped in aleph-zero is that we have been able to discover and write about larger infinities, like aleph-one, and we are able to refer to and use real numbers, such as  $\pi$  and  $e$  and  $\sqrt{2}$  that seem to belong to sets of size aleph-one or bigger.

Similarly, the diagonal arguments such as Cantor’s, Gödel’s (incompleteness theorem) and Turing’s (undecidability of the halting problem) seem to depend on our ability to see truths that go beyond the enumerations being used in the proofs. And, a lot of mathematics (such as calculus) depends on the fact that real numbers form a genuine continuum (in aleph-one or bigger), unlike the rational numbers (which are in aleph-zero).

<sup>4</sup> And we’ll include in this theoretical enumeration all the evolved descendants of humans too.

<sup>5</sup> We can easily imagine extending this enumeration to all agents in the universe, but humans will do for our purposes here.

So, in the realm of logic and mathematics there seems to be a genuine importance to the larger infinities such as aleph-one. This in turn suggests that the set of things that are ‘true’ in this universe is greater in size than aleph-zero. Also, given that we have discovered this mathematics of larger infinities and are able to work effectively with real numbers it seems at least plausible that we are somehow able to think in ways that go beyond aleph-zero.

However, even if we are able to think with higher fidelity, the enumeration presented above provides a definite limit to the fidelity with which we can express ideas. And this includes every possible way that we could recordably communicate our ideas about aleph-one and beyond. So, what is going on?

### 3.2 What if the physical universe is ‘larger’ than aleph-zero?

Our models of physics, including quantum mechanics, use mathematics that depends on the continuous nature of real and imaginary numbers. This suggests there is also a certain physical importance to the infinities beyond aleph-zero.

However, a key unknown about the nature of the physical universe is whether or not the Heisenberg uncertainty principle describes a limit on the fidelity with which we could ever observe the universe, or a limit on the ‘fidelity’ of the universe itself. Or to put it another way, is space time a genuine continuum (in aleph-one or beyond) or is it actually countable (in aleph-zero)?

There is some disagreement among physicists and philosophers about this issue and indeed given the Heisenberg limit on what can be measured it is unlikely that this question could ever be solved empirically! But there is no need to take a firm view in this paper, as here it suffices that we note the two possibilities. Either the scale of the universe is in aleph-zero, or it is larger than aleph-zero.

### 3.3 An aleph-zero scale universe

If the scale of the universe is within aleph-zero, then there would be no discrepancy between the scale of what humans and computers can express and the scale of the universe itself. There would still be problems of tractability (both in terms of the time and resources required to calculate or express certain things), but there would be no fundamental fidelity gap between reality<sup>6</sup> and expressibility.

### 3.4 An aleph-one or larger scale universe

On the other hand, it is possible that the scale of the universe is in aleph-one or larger. Even if we could never observe the full fidelity of the universe, it might be that space-time is indeed a continuum just like the real and imaginary numbers that we theoretically use in our models of it.

In this case there would be a fidelity gap between the reality of the universe and the expressibility of humans and computers. It would rule out the possibility of using digital computation to accurately simulate the universe at the fidelity of the universe itself. It might also give some embodiment in the universe of the kind of expressibility ‘gap’ exposed by logical diagonal arguments, such as Gödel’s incompleteness theorem. Any mode of human or computer expression could only reach countably many ‘truths’ and yet there would be circumstances where what we are trying to express refers to a feature of the universe with aleph-one or more ‘truths’.

Indeed, if humans couldn’t possibly express all of the truths of the universe, then this would give an additional reason for the postmodern concerns about the gap between what we can express, what we mean to convey and what is understood by others. Or to put it another way, the idea that all human recordable discourse can be enumerated should not be seen as giving support for a modernist, totalising form of rationalism. Quite the reverse. It is a limit within which we are ‘trapped’ and the full truths of the universe might be beyond our expressible grasp.

And yet there is still the apparent conundrum of how we are able to use the concepts of real numbers and the aleph-one infinity (and higher) if all our expressions are trapped in aleph-zero.

### 3.5 How we refer to pi ( $\pi$ )

Pi is probably the irrational real number that has been studied in the most detail. In 2016 the record for enumerating the digits of pi stood at over 22 trillion digits (Trüb, 2016). Ironically, even listing out

<sup>6</sup> In this paper we’re admittedly working with a very simple, naïve realism.

all these trillions of digits would be a slightly less accurate way to refer to pi than writing its well known symbolic signifier: ‘ $\pi$ ’.

With this one character we refer precisely to an infinite series of digits because we know the algorithm with which to calculate any arbitrary precision of  $\pi$ . It is a non-halting algorithm, so the expansion from the symbol,  $\pi$ , via the algorithm, to the infinite series of digits is intractable, but that does not render our ability to refer inaccurate.

Furthermore, in mathematical formulae we can use the symbol  $\pi$  to stand in for the exact number and in some situations different uses of  $\pi$  will cancel each other out.

So, we can refer to  $\pi$  accurately because although it is irrational we can describe it accurately with a finite length algorithmic description. We can then meaningfully refer to this description (and thereby the accurate number) via its even simpler, naming signifier.

Note, however, that the description, “the next real number bigger than pi” does not refer correctly to anything as there is no ‘next’ operator on the real numbers. So, not all attempts to refer to a real number using a description will be successful. However, the conjecture here is that for all real numbers (rational or irrational) that we can refer to accurately we will be doing so via a finite length description or algorithm of some kind.

As each of these descriptions will have at least one EBUID that encodes the description, therefore there are only countably many irrational real numbers that we can accurately refer to in a similar way to which we refer to  $\pi$ . Obviously the other well known irrational real numbers, like  $e$  and  $\sqrt{2}$ , are members of this set, but the conjecture here is that all irrational numbers that we can refer to accurately must have a finite description length algorithm that can generate any arbitrary precision expansion of that irrational number (given enough computational time and resources). Furthermore, the size of that set will be aleph-zero, it will itself be a countable set.

Conversely, the overwhelming majority of irrational numbers cannot be referred to accurately with a finite description length. These numbers require all of their infinite series of digits to be listed in order to be accurately referred to and to write down such an infinite signifier would be an intractable undertaking.

In other words, we can only actually work with a countable set of irrational numbers. Our descriptions of these numbers and of aleph-one and so on are all finite length descriptions, and this is how we are able to usefully refer to aleph-one and some irrational numbers within our countable set of all possible acts of communication.

### 3.6 How we recordably refer to unrecordable thoughts, feelings and experiences

We often use language to refer to experiences that many people will have shared, but for which the full content of the experience could never be recorded or communicated in its entirety. Just as with irrational numbers, like  $\pi$ , we use descriptions to indirectly refer to the otherwise inexpressible. For example, we may use a description that tries to convey to the reader which of their own unrecordable experiences we are trying to refer to. Or, if the reader has never had such an experience we may try to describe how the reader could get into the right kind of situation in which they would experience something similar. We then use short signifiers, like ‘pain’ or ‘love’, to refer to the experience ‘via’ a lifetime’s collection of these longer descriptions of what, say, ‘love’ is. However, we could never communicate the experience itself. And our descriptions of experiences are never (or rarely) as accurate and repeatable as our algorithmic descriptions of specific irrational numbers like  $\pi$ .

The point being made in this paper is that the total conceivable cannon of all such recordable descriptions of unrecordable thoughts, feelings or experiences would also have to be ‘in’ aleph-zero.

### 3.7 Comparing humans and computers

The purpose of this paper is not to make the claim that humans and computers can refer to the same things in the same way. When a human who has bitten into an apple refers to this experience via the description, “it was just like biting into an apple” they are referring in a way that no non-human could ever fully understand or achieve. A robot that can bite apples might have an experience when doing so, but this would not be a qualitatively similar experience to those had by humans. Therefore, if the

robot were to say, “it was just like biting into an apple” it is not quite referring to the same thing as a human would be. If an unembodied, text-analysing AI (that doesn’t have the capacity to even potentially experience biting into an apple) were to ‘say’, “it was just like biting into an apple” then it’s questionable whether or not the generated text describing an experience is meaningfully referring to anything.

But this paper is not examining the general question of how recordable communication (like text) can meaningfully refer. Rather the purpose of this paper is to highlight that the difference between how computers could potentially refer compared with humans is *not* a difference in the scale or fidelity of communication as both are trapped in aleph-zero.

And so, if we’re looking to understand the difference between how humans and computers can meaningfully refer we should not think in terms of humans being able to refer to ‘more’ than computers, or that humans could refer in a ‘finer grained’ way than computers. Any difference must come from elsewhere, such as the difference between the ways that humans and computers are engaged with the world around them.

Note that to motivate these conclusions we do not need to claim that human thought is necessarily bounded in a similar way to human expression. It may turn out that human cognition taps into quantum computation in a meaningful way and so is vastly superior to traditional digital computation. However, even if this were true in an interesting way, our ability to communicate ideas to each other would still be bounded by countability as laid out in this paper.

It is similarly worth noting that even if analogue computing or quantum computing are able to work in aleph-one fidelity by exploiting aleph-one features of the universe (if they exist) any attempt to communicate the ‘results’ to humans would again be bound by the aleph-zero countability constraints discussed above.

#### **4 Use of Unique IDs (UIDs) in programming**

Finally, we’ll take a quick look at what the preceding discussions imply for the appropriate use of unique IDs (UIDs) in programming environments such as in Semprola (Sharpe, 2018).

For any attempt to improve the semantic depth of programming to get closer towards the level of meaning imbued in text and other forms of communication by humans it would seem, at first glance, that the programming environment should be ever more like something that humans would normally work with or be ever more biologically inspired. And, humans do not normally use UIDs in their daily lives nor is there any hint of any suggestion that anything like UIDs are in operation within the mechanisms of the brain.

So, the use of UIDs feels at odds with any project to improve the semantics within programming.

However, as has just been discussed above, the entirety of human discourse is enumerable and therefore it would be possible (in theory) to assign to every recordable act of communication a UID along the lines of RHCUID above.

For most of history we have had no mechanism or indeed purpose to do anything like this, but with the emergence in the last decades of digital communication technologies and indeed a growing number of people “life logging” all recordable aspects of their own life, it is the case that a growing number of communication acts by humans are actually being given UIDs even if they do not belong to a single, universal enumeration.

All phone calls, text messages, Skype calls, emails, documents, photos and more each have some form of explicit or implicit UID. Indeed these UIDs are invaluable to ensure that the identity and thereby history of a recorded act of communication remains stable even as the binary file encoding that communication act is copied and transported around a multitude of computer infrastructure.

As mentioned in relation to Borges’ library in section 2.1 above, it is the history of a piece of text (or encoded binary file more generally) that is the vital link between the true semantics of the text with its original author. Without this link a randomly generated piece of text only has the appearance of signifying an intended meaning.



With a physical book the physical object itself can act as the maintainer of the continual identity through which the history of the book to an author can be traced (or more usually assumed). But with virtual bits of text this ‘metadata’ relationship has to be explicitly maintained with the text if it is not to be lost. This is one of the crucial roles that UIDs can perform for virtual text or, of course, any binary file recording of an actual communication act. This is one of the reasons why Semprola ensures that every piece of text (for example) has a Semiotic Programming Unique ID, SPUID.

There are of course other, important ways that UIDs can help maintain high levels of semantic information (such as by helping distinguish the identities of relata with greater semantic accuracy than would be possible by just using human readable text labels), but these are not so relevant to the discussion here.

Having explained the reasons why a programming environment like Semprola would systematically use UIDs, it is still possible for this abundance of UIDs to give an impression that what Semprola will be capable of is so obviously countable in scale that this must somehow be less than what humans are capable of. Hence, the key message of this paper was that actually all of human communication is also only of countable scale, so this apparent ‘limit’ on an environment like Semprola is no greater limit than already applies to humans!

## 5. Conclusion

In this paper we have compared the fidelity of the universe with the fidelity with which humans and computers can communicate recordable ideas. With the use of various enumerations it was shown that humans, just like computers are only able to communicate to a fidelity equal to or less than the aleph-zero infinity and so there is no fidelity gap between humans and computers. And yet, it is possible that the universe has an even greater fidelity of aleph-one or higher. If true this would imply that there is a gap between what could ever be communicated by humans or computers and the scale of things that are true about the universe.

However, a key point of the paper is to highlight that there is no such ‘scale’ gap between the fidelity with which humans and computers can refer. So, if we wish to understand the differences between how humans and computers refer we cannot simply suggest that humans can refer to ‘more’ than computers or that humans can refer “in more subtle ways” than computers. Instead, any gap between the ways that computers and humans can meaningfully refer must lie somewhere other than scale.

## 6. Appendix

A lovely online version of the Library of Babel has been constructed by Jonathan Basile (<https://libraryofbabel.info>). Not only can you virtually browse through the shelves in the hexagonal library, but due to its computational construction, you can also search for books that contain particular pieces of text. Below is a list of four such books where the abstract of this paper can be found on a page in their random text:

- Page 115 of book at location: <https://libraryofbabel.info/bookmark.cgi?ppig2018.1>
- Page 15 of book at location: <https://libraryofbabel.info/bookmark.cgi?ppig2018.2>
- Page 351 of book at location: <https://libraryofbabel.info/bookmark.cgi?ppig2018.3>
- Page 179 of book at location: <https://libraryofbabel.info/bookmark.cgi?ppig2018.4>

## 7. References

Borges, J.L. (1941) The Library of Babel.

Cantor, G (1891) via Wikipedia [https://en.wikipedia.org/wiki/Cantor%27s\\_diagonal\\_argument](https://en.wikipedia.org/wiki/Cantor%27s_diagonal_argument)

Unicode (1991) The Unicode Standard. <https://www.unicode.org/standard/standard.html>

Trüb, P (2016) 22.4 trillion digits of pi. <https://pi2e.ch/blog/>

Sharpe, O. (2018) Semprola: A Semiotic Programming Language. Salon des Refusés Workshop, <Programming> 2018