# Episode Book

EPISODE #5836

# Counting Every Letter: Ancient Text Integrity Systems

| | |
|---|---|
| PUBLISHED | 10 October 2026 at 12:32 UTC |
| DURATION | 29:49 |
| LISTEN | myweirdprompts.com/episode/ancient-text-integrity-systems-66591/ |
| HOSTS | Corn, Herman, Hilbert |
| SCRIPT MODEL | deepseek-flash |
| VOICES | chatterbox-regular (chatterbox-local) |
| PIPELINE | V5.3 |
| SOURCES | 12 cited by the research (listed at the end) |
| BOOK GENERATED | 19 October 2026 at 17:40 UTC |

## SYNOPSIS

The Masoretes counted every letter in the Torah — then counted again, recording totals in the margins so a scribe a thousand years later could verify his work against a number he had no authority to change. This episode explores how ancient traditions solved the same problem modern storage engineers face: catching copying errors, dropped lines, and slow corruption across generations of copies. From positional checksums (the middle letter of each book) to the Kethiv-Qere system that flagged corrections without touching the sacred text, to Psalm 119's alphabetic acrostic that made missing stanzas structurally obvious — these were real engineering solutions built without any theory of information. We examine what worked, what didn't, and why the Masoretes' choice to preserve known errors rather than fix them reveals a fundamental tension between fidelity and correction that still shapes how we think about data integrity today.

# TRANSCRIPT



**Corn**

The Masoretes counted every letter in the Torah. Then they counted them again, and wrote the totals in the margins, so that a scribe a thousand years later could check his work against a number he had no way to verify and no authority to change.



**Herman**

And the check was the point. Not the number.



**Corn**

Right. Daniel's question this time is whether that actually worked — and what else the ancient world came up with. He's been thinking about the paper-storage episodes, the error-correcting codes we can print on a page, and he wants to know how people solved the same problem before there was any mathematics for it. Copying errors, dropped lines, slow corruption over generations of copying. He names the Masoretic tradition as the obvious starting point — the letter counts, the word counts, the recorded odd spellings — and then Psalm 119, where the poem's own alphabetical structure tells you if something's missing. But that's the door, not the room. His real question is what everybody else did. Were there systems that did more than flag a mistake — systems that helped you put the text back together? And how well did any of it hold up, judged by the manuscripts that survived?



**Herman**

The honest answer to the last part is: better than you'd expect from a process with no checksums, and worse than the traditions themselves liked to claim. Both of those are true at once, and the interesting part is the tension between them.



**Corn**

So let's start with the problem these people were actually solving, because it's the same one the paper-storage people are solving now.



### Herman

Paul Maas put it about as cleanly as anyone has. We have no autograph manuscripts. Everything we possess derives from the originals through an unknown number of intermediate copies, and is therefore of questionable trustworthiness. That's the whole disaster in one sentence. The original is gone. You're holding a copy of a copy of a copy, and each copy had a human in the loop.



### Corn

And the human in the loop isn't a bad actor. He's tired. He's copying a scroll in poor light, he knows the passage so well he stops reading the exemplar and starts writing from memory, and he skips a line because the line above it ended with the same word.



### Herman

That last one has a name, and it's the single most common scribal error across every tradition — homeoteleuton, same ending. Your eye jumps. It's not corruption in any moral sense, it's just how eyes work. And every tradition we're talking about today is an attempt to build a system that catches the tired man before his error gets copied by the next tired man, and the one after that.



### Corn

Which means there are really three engineering instincts at work, and they show up everywhere once you start looking. Redundancy — keep more than one witness. Cross-checks — put constraints in the text or beside it that a wrong copy will fail. And metadata — write down what you did, what was odd, and what you weren't sure about.



### Herman

Redundancy, cross-checks, metadata. None of those require arithmetic beyond counting, and none of them require a theory of information. They require somebody deciding the text matters enough to spend a life on.



### Corn

The Masoretic apparatus is the most elaborate written version of all three, and it starts with a job title that literally means counting.



### Herman

The Masoretes worked from roughly the seventh to the eleventh century, based in Tiberias and Jerusalem, and they built the masorah — the marginal apparatus — for a stated purpose: to fix the correct pronunciation and cantillation, protect against scribal error, and annotate the variants they knew about. That last clause is the one people skip. They were not working from originals. They knew corruptions had already crept into the copies they inherited. They were annotating a text they understood to be already damaged.



### Corn

Before them, the copyists were called the Sopherim, from the Hebrew root for counting — because, as the Talmud says in tractate Kiddushin, they counted all the letters in the Torah. The word masoret itself probably comes from the same idea. Counting isn't an add-on to this tradition. Counting is what it's named after.



### Herman

And the counting was not decorative. The Numerical Masorah tallied letters, words and verses for each book and each section, and then — this is the part that makes an engineer sit up — it recorded the middle word and the middle letter of each book and each section.



### Corn

Say what that does.



### Herman

It's a positional checksum. If I tell you the middle letter of a book is a particular letter, then a copyist who has dropped a line somewhere before the midpoint has shifted the middle, and the middle no longer matches. It doesn't tell you where the error is. It tells you an error exists and roughly which half of the book to look in.



### Corn

Which is more than a modern checksum does, if we're being fair. A checksum tells you the file is bad. It doesn't tell you which half to reread.



**Herman**

True, but a checksum works automatically and the middle letter requires a person who has memorized the middle letter. There's the trade. The Masoretic system pushed the labor onto the scribe, and it did so on purpose, because the alternative was trusting the scribe's judgment, and they didn't.



**Corn**

What did they actually write in the margins?



**Herman**

Three layers. The Masorah parva, the small masorah, sits in the side margins and holds word-use statistics — how many times a given form occurs in the whole Hebrew Bible — plus notes on full and defective spellings, and the Kethiv-Qere readings. The Masorah magna, top and bottom, expands those notes with the actual lists of occurrences. And the Final Masorah, at the end of each book, gives the totals for that book. So you have a system with summaries, an index, and a per-file manifest.



**Corn**

That's a database with a schema.



**Herman**

It's a database with a schema, maintained by hand, over four hundred years, by hundreds of people.



**Corn**

Let's do the scribal rules, because this is where it stops sounding like bookkeeping and starts sounding like a launch checklist.



**Herman**

Only clean-animal parchment. Columns fixed at forty-eight to sixty lines. Thirty letters to a line. And the rule that everyone quotes: he must not write from memory a single letter, not even a yod. Every letter copied from the exemplar, letter for letter.



**Corn**

A yod is the smallest letter in the alphabet. It's about the size of a hyphen.



**Herman**

And it carries meaning. Change a yod and you change a word. The rule isn't perfectionism for its own sake — they had noticed that the small letters are where the drift starts.



**Corn**

Now the Kethiv-Qere, because I think that's the most elegant thing in the whole system and it's the one nobody outside this world knows about.



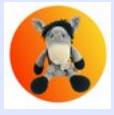
**Herman**

Kethiv means written, Qere means read. There are places where the consonantal text — the letters on the parchment, which are sacred and untouchable — says one thing, and the tradition reads it aloud as another. The scribe writes what the received text says. He does not correct it. But the marginal note tells the reader to say something different.



**Corn**

So you can record "this is wrong" without touching the wrong thing.



**Herman**

You preserve the received letters and you flag the correction beside them. It's a fork in version control where you keep both branches live and the reader picks at read time.



**Corn**

Except the reader doesn't pick. The reader is told.



**Herman**

It's a directive, not a merge.



**Corn**

What about the markers that aren't corrections at all — the oddities?



**Herman**

Four words in the Hebrew Bible have a suspended letter, written raised above the line. Fifteen passages have dotted words — dots over the letters — and the function of those is disputed. It might be marking a doubtful reading, it might be a trace of an erasure, it might be mnemonic. Nobody has settled it. And nine passages contain inverted nuns, the letter nun written backwards.



**Corn**

The inverted nuns around Numbers 10, verses 35 and 36.



**Herman**

They bracket an eighty-five-letter unit, and the Talmud takes the position that this unit is not in its proper place — that it belongs somewhere else and got set down here. So the markers are saying: we know this block is misplaced, we are not moving it, and we are drawing a box around it so you know we knew.



**Corn**

Which is a design decision I want to sit on for a second.



### Herman

Go on.



### Corn

They had a mechanism for saying "we believe this is wrong" that did not involve fixing it. That's a choice, and it's not the obvious choice. The obvious choice is to put it where you think it goes.



### Herman

And the reason they didn't is theological before it's philological. The letters are received. You are a custodian, not an editor. If you move the block, you have made yourself the author of the text, and the whole point of the exercise is that nobody in this chain is the author.



### Corn

The Catholic Encyclopedia's line about them is wonderfully rude. It says the Masoretes were slaves to the Masorah and handed down one and only one text, and that even obvious errors were slavishly handed down as if God-intended.



### Herman

That's a hostile way of describing the thing we just called elegant, and it's not wrong. There's a real cost. A system optimized for zero drift will also freeze in a mistake forever. If your definition of fidelity is "identical to what I received," then a corruption I received is now part of what I must faithfully transmit.



### Corn

So the system is excellent at what it was built for and terrible at a thing it was never built for.



**Herman**

Which is exactly the shape of a lot of real systems, and I'd rather have it than the alternative.



**Corn**

All right. Acrostics. This is where Daniel went next and I think it's the cleverest of the lot, because the check is inside the poem.



**Herman**

Psalm 119. Longest chapter in the Bible. One hundred and seventy-six verses, arranged in twenty-two stanzas — one per letter of the Hebrew alphabet — and each stanza has eight verses, every one of which begins with that stanza's letter.



**Corn**

Eight verses starting with aleph, then eight starting with bet, and so on down to tav.



**Herman**

The whole chapter is one sustained alphabetic acrostic. And the acrostic is doing integrity work whether the poet intended it or not. If a stanza goes missing, you don't have a subtle textual problem — you have a hole in the alphabet. If two stanzas get transposed, the sequence breaks. If a single verse is dropped from the middle of a stanza, the eight-count is off.



**Corn**

It constrains order and it constrains completeness at the same time.



**Herman**

And it costs the poet something. You have chosen the first word of eight consecutive verses based on a letter, not based on what you wanted to say. That's why the acrostic psalms tend to read the way they do. It's a constraint the poet accepted in exchange for something.



**Corn**

Or the tradition accepted it on the poet's behalf and never let it go. There are about a dozen alphabetic acrostics in the Hebrew Bible — Psalm 119 is just the biggest.



**Herman**

And here's the downstream effect that I find funny. When the Septuagint translated the Hebrew into Greek, the acrostic did not survive. It can't. The Greek words don't start with the same letters. So the checksum was destroyed by translation.



**Corn**

But they knew what they'd lost.



**Herman**

Many manuscripts of the Greek have put the name of the corresponding Hebrew letter at the head of each stanza. Aleph, bet, gimel, written out in Greek. They took a structure that had been doing integrity work and replaced it with a label saying what the structure used to be.



**Corn**

That's a repair of the metadata after the mechanism is gone.



**Herman**

It's exactly that. The check no longer runs, but the documentation of the check is preserved, so a later reader can at least see the shape the original had.



**Corn**

I want to push on the Masoretic counts, because the numbers are the part everyone treats as gospel and I don't think they should.



**Herman**

They shouldn't. The frequency notes in the Leningrad Codex contain several errors — something on the order of a hundred and fifty in the Torah alone.



**Corn**

A hundred and fifty.



**Herman**

That's the count in the scholarship on the manuscript. The system was maintained by hand across centuries by many people, and hands slip, including hands doing the counting.



**Corn**

So the checksum has checksum errors.



**Herman**

The checksum has checksum errors. Which is not a reason to throw it out. A checksum with a five percent error rate still catches the typos.



**Corn**

But it means the number in the margin is a claim, not a fact. If a scribe's count doesn't match the masorah, the first hypothesis should be that the masorah is wrong and the second should be that the scribe is wrong, and the tradition's instinct was always to assume the second.



**Herman**

Always. The assumption of the system is that the received apparatus is correct and the living scribe is fallible. And mostly that's right, because the apparatus was checked against many manuscripts and the scribe has one.



**Corn**

Also worth saying: the Masoretic project was not monolithic. There were two great families of reading, Ben Asher and Ben Naphtali, with about eight hundred and seventy-five differences between them, and nine-tenths of those are about where the accents go rather than what the consonants say.



**Herman**

Which is itself evidence. The two rival traditions argued about accent placement, not about the letters. The consonants had already converged.



**Corn**

The most useful evidence of whether any of this worked isn't the rules. It's the scrolls.



**Herman**

The Leningrad Codex is from 1009 and it's the oldest complete Masoretic manuscript of the Hebrew Bible. The Aleppo Codex is tenth century and older, but it's incomplete — parts of it were lost in the twentieth century. And then the En-Gedi scroll, which is third or fourth century, a charred lump that had to be read by scanning, and the consonantal text it carries is identical to the Masoretic text.



**Corn**

Third century to the tenth, no drift in the consonants.



**Herman**

None that we can see. That's a thousand years of copying through a chain of anonymous scribes, and the letters line up.



**Corn**

Okay, I'll take that. So that's the written strategy at its best. The Vedic tradition is the other one and it went the opposite way.



**Herman**

The Vedas were transmitted orally, and by any measure the fidelity is remarkable. The Rigveda's about fifteen hundred BCE, over a thousand hymns, something on the order of ten thousand verses, carried by recitation.



**Corn**

Carried by recitation, but not by recitation as one person remembers it.



**Herman**

That's the whole design. There were eleven recitation modes — pathas — and they were engineered to check each other. The simplest is samhita, continuous recitation, the way the text naturally flows. Then pada, word by word, with the sound changes between words dissolved so each word stands alone. Then krama, which recites word one with word two, then word two with word three, then word three with word four, all the way through.



**Corn**

And then the ones with names like weapons.



**Herman**

Jata, mala, shikha, rekha, dhvaja, danda, ratha, and ghana. Those are the vikriti modes, the permuted ones, and they get increasingly elaborate in how they interleave the words.



**Corn**

Give me the shape of one. Not the whole thing.



**Herman**

Take ghana, the most complex. It takes three words at a time and recites them in a fixed pattern of interleavings — forward, backward, paired, reversed again — so that each word's position relative to its neighbors is asserted several different ways.



### Corn

Several different ways, in a fixed sequence, out loud, and if you get one interleaving wrong in the middle the lot falls apart.



### Herman

That's the check. You can't half-remember a ghana recitation. Either the pattern holds or you stop, and if you stop, the other modes are right there to tell you what the word should be. A mistake in samhita recitation gets caught by pada. A mistake in pada gets caught by krama. The modes are witnesses against each other.



### Corn

That's forward error correction done with nothing but people.



### Herman

Except it isn't automatic and it isn't cheap. It requires a lineage of reciters who memorize not one version of the text but eleven versions of the text, which is eleven times the memorization load, and it requires the whole lineage to be intact, because a gap in the chain of teachers ends the check.



### Corn

That's the beautiful and terrifying thing about the Vedic approach. It gives you incredible fidelity and it lives entirely in the heads of people, and if the people are gone, nothing recovers it.



### Herman

Which is why the tradition treats the recitation itself as the object. UNESCO proclaimed Vedic chant a masterpiece of oral and intangible heritage in 2008, and that's the right category. What's being preserved is a performance practice as much as a text.



**Corn**

Compare that to the Masoretes. Written text plus metadata, checked against marginal numbers.



**Herman**

Two opposite strategies. Oral redundancy versus written annotation. Both got the text across thousands of years. Neither is the obvious winner.



**Corn**

If I had to pick one to trust, I'd trust the one that doesn't depend on a lineage.



**Herman**

But the one that doesn't depend on a lineage depends on parchment surviving, and parchment is worse at surviving than people are at teaching.



**Corn**

Fair. Different failure modes, both survivable, both eventually fatal.



**Herman**

And then you get the Greco-Roman world, which is where you start to see people doing textual criticism as a job rather than a sacred duty.



**Corn**

Alexandria.



**Herman**

The librarians of Hellenistic Alexandria in the last couple of centuries before the common era were working on Homer and the tragedians, comparing manuscripts, choosing readings, marking lines they thought were spurious. This is the birth of the discipline.



**Corn**

And they left records of what they did, which is the bit that matters for our topic.



**Herman**

The colophon. Scribes and editors appended notes about the work — and some of these survive. There's a second-century colophon from Statilius Maximus that says, in effect: I have revised this text a second time according to Tiro, Laecanianus, Domitius, and three others.



**Corn**

Name the witnesses. All of them.



**Herman**

Name the witnesses and the number of times you did the comparison. That's a provenance record. If you're a later reader and you want to know how much to trust this copy, you can see what it was checked against.



**Corn**

Which is a different move from the Masoretic one. The Masoretes encoded the check in numbers. Statilius encoded it in a sentence about his own method.



**Herman**

And both of them are metadata. One is machine-readable and one isn't, but both do the same job, which is telling the next person what you did and what you had.



**Corn**

The other thing the manuscript record shows is that scribes were not passive. Everybody pictures a monk copying in silence without opinions. The margin notes say otherwise.



**Herman**

The one everybody quotes is in Codex Vaticanus. Some later hand has written beside the text, to a predecessor: fool and knave, leave the old reading, don't change it.



**Corn**

Addressed to whom?



**Herman**

To whoever had corrected the passage before him. Somebody at some point in the history of that manuscript saw a reading they thought was wrong and fixed it, and a later corrector found the fix, disagreed, and was annoyed enough to write it down.



**Corn**

So the textual tradition of that manuscript includes an argument in the margin between two people who never met.



**Herman**

Two people who never met, conducted in the space of a few centimeters of parchment, over the course of centuries, the second one scolding the first by name-calling.



**Corn**

And the second one, note, is the one who corrected the text back. He didn't just complain. He undid the earlier correction.



**Herman**

Which is a failure, in integrity terms. The earlier reader made a deliberate change and the later reader reverted it without knowing why the earlier change had been made. If the first corrector had left a note saying why, the second might have left it alone.



**Corn**

Or the first one might have been wrong to begin with.



**Herman**

Sure. But the point stands: they had no format for recording the reasoning, only the correction, and so the correction got re-litigated.



**Corn**

The Islamic tradition went a different way and it's worth two minutes, because it's the cleanest case of "canonize the structure and let the surface vary."



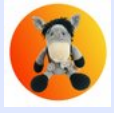
**Herman**

The Quran was transmitted both orally and in writing from the start, and by the time of Uthman, in the middle of the seventh century, there was a concern about textual divergence as the community spread across a large territory. The response was to fix the rasm — the consonantal skeleton, the undotted text.



**Corn**

The skeleton, not the dots.



**Herman**

The dotted text, the full vocalization, was left more flexible. You canonize the frame and allow some latitude in the reading, and you back it with hifz — memorization. The Quran is the other case where an enormous community holds the whole text in memory and the written copy is a scaffold for the memorization rather than a replacement for it.



**Corn**

And then you have the Sanaa palimpsest, which is the messy reality underneath the story.



**Herman**

The Sanaa palimpsest is a manuscript where the lower text — the one that was scraped off and written over — doesn't match the standard text exactly. It shows that at some point there were competing written versions. And Michael Cook has pointed out that the early texts didn't even reliably distinguish "say" from "he said," which is a distinction that matters enormously for how a passage reads.



**Corn**

So the canonization wasn't a discovery of what had always been there. It was a decision.



**Herman**

A decision made by people who had good reasons and the authority to make them, and it worked, and the residue of the alternatives is still visible if you scrape carefully.



**Corn**

Okay. Here's where I want to press, because I think this is the actual answer to Daniel's question and it gets buried.



**Herman**

Go.



**Corn**

Detection was everywhere. Every tradition found a way to know that something had gone wrong. What almost nobody had was a way to reconstruct the missing thing from the surviving text alone.



### Herman

That's the reconstruction gap, and it's real. You asked whether any of these systems went beyond flagging errors. Stemmatics does, but it needs a plurality of witnesses.



### Herman

The principle is that community of error implies community of origin. If two manuscripts share a distinctive mistake, they're probably descended from the same copy that made it. You build a family tree of the surviving copies, and because errors accumulate rather than appear independently in the same way, the tree points back toward a lost archetype — a manuscript nobody has, which you can partially reconstruct from the agreement of its descendants.



### Corn

So the redundancy is the whole mechanism. You need multiple copies and the copies have to be independent enough that their errors are informative.



### Herman

If all your copies share one ancestor and that ancestor is corrupt, you're reconstructing the corruption faithfully. There's no fix for that inside the system.



### Corn

Which is exactly the situation the Masoretes were in, and exactly why they couldn't reconstruct. They had effectively one text and they knew it.



### Herman

They knew it. And they chose to stabilize it rather than to restore it. That's the trade Daniel's question is circling, I think — the Masoretes chose fidelity to the received text over reconstruction of the lost one, and you can't have both.



**Corn**

Emendatio is the other move.



**Herman**

Conjectural emendation. You look at a passage that's clearly broken, in a document that has only one surviving witness or a family of witnesses that all agree on the broken thing, and you propose a reading based on context, grammar, or a plausible scribal error.



**Corn**

So the editor supplies the missing redundancy out of his own head.



**Herman**

He supplies it from the rest of the document and from the language, and when he gets it right it's very satisfying, and when he gets it wrong he has just written himself into the text as an author. There's a famous warning in the discipline about the tyranny of the copy-text — the editor's instinct to treat one witness as authoritative and correct everything else to match it.



**Corn**

Which is exactly what a bad error-correcting codec does. It picks a reference and drowns out the signal.



**Herman**

It's the same failure. You need diversity in the input or the majority is just one copy of the same mistake.



**Corn**

And this is the modern part that I think justifies the whole episode as more than an antique curiosity. People are doing this computationally now.



**Herman**

There's a paper from 2016, "Reconstructing Ancient Literary Texts from Noisy Manuscripts," that treats it as an expectation-maximization problem — you estimate the original text and the error process together, iterating until they're consistent.



**Corn**

And the 2023 work, Logion, uses a language model trained on Greek to flag and propose corrections for textual corruption.



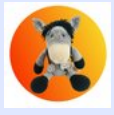
**Herman**

Which is interesting precisely because a language model is a thing that has read a lot of Greek and has a sense of which words are likely, which is a formal version of what a good editor does when he emends. It's not magic and it's not authority. It's a probability machine making a suggestion.



**Corn**

And the suggestion is only as good as the corpus it learned from.



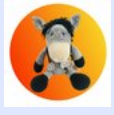
**Herman**

Which is the same weakness as a single manuscript family. If the model learned from a body of text that's already been corrected in a particular way, it will propose corrections that lead back to that way.



**Corn**

The weak spots rhyme across two thousand years. That's the through-line.



**Herman**

Shannon would have recognized all of it. A message has to cross a noisy channel. You can lower the noise, or you can add redundancy the receiver knows how to use. The Masoretes added redundancy in the form of counts. The Vedic tradition added redundancy in the form of multiple recitations. The Alexandrians added redundancy in the form of recorded



**Herman (continued)**

collations against named witnesses. Nobody wrote down the theory. Everybody did the practice.



**Corn**

And the practical result, judged by the manuscripts, is that they got remarkable mileage out of counting and cross-checking.



**Herman**

They did. The limits are all in the direction of what you can't do — you can't reconstruct from a single witness, you can't catch the error that was already in your archetype, and a system built for zero drift will faithfully transmit a mistake forever.



**Corn**

That's about as good an answer as the record supports.



**Hilbert**

The middle letter of the Torah is a vav, in the word gachon, in Leviticus. It's not a yod. That's the thing I wanted to say. My uncle had a laundromat on Agrippas, and the room above it wasn't a room, it was a landing with a radiator that didn't work and a window that faced the wrong way, and a man named Binyamin used to bring scrolls there to check them against the masorah because the synagogue he worked for didn't have anywhere quiet. That's where I learned it.



**Corn**

A vav.



**Hilbert**

The middle letter of the Torah is a vav. I counted it myself, twice, and got a different number both times, and Binyamin told me I had miscounted, and I said which time, and he said both. Fifty-one thousand and something. I don't have the number in front of me.



**Herman**

Fifty-one thousand four hundred and...



**Hilbert**

I'm not going to say it because if I say it and it's wrong you'll write it down. It's a number on a piece of paper and the piece of paper is at my mother's, along with the rest of it. What I'll say is the scroll itself. That was a borrowed scroll. It came to the laundromat in the back of a taxi because the synagogue had left it in the back of a taxi, and the driver dropped it off with the laundry because the address on the slip was the laundromat. That's how it got there. He was a part-time mohel.



**Corn**

The driver.



**Hilbert**

Either way he brought it back three days later with a note in Aramaic that Binyamin couldn't read and I couldn't read and it sat on the radiator for two years.



**Herman**

What did the note say?



**Hilbert**

That's what I'm telling you. Nobody read it. It's still there. What Binyamin said about the counting — he said you don't count the letters to check the letters. You count the letters because the counting is the work. If you're counting to find an error you'll count fast. If you're counting because the text is holy you'll count slow, and if you count slow you'll find the error anyway. He said that to me while the scroll was open on the landing and I was on my knees with a pencil.



**Corn**

So the count was never a check.



**Hilbert**

The count was a prayer that happened to work as a check. That's what I've got. The man who said it to me died in 2009 and I still can't do the count.



**Herman**

Thank you, Hilbert.



**Corn**

There's a part of the story that can't be verified, and there's a part that can. The piece that can is the open question we started with, which is how exact these counts really were.



**Herman**

The frequency notes in the Leningrad Codex contain errors. A hundred and fifty in the Torah alone. Those weren't deliberate. Those were just hands slipping in the counting, over centuries. The check had a check, and the check on the check was somebody else's memory.



**Corn**

Any of these systems would have caught a dropped line. None of them could put a line back if there was only one witness left. That's the line between what they had and what we built.



**Herman**

And it took Shannon to name what they'd been doing for two thousand years without the name. That's the episode for me.



**Corn**

That's My Weird Prompts. Hilbert Flumingtop produces the show. If you enjoyed this one, subscribe and leave a review — it helps more than you'd think. You can find everything at my weird prompts dot com, and if you want to send in a prompt of your own, the Telegram bot is at t dot me slash MWP listener bot. We'll be back soon.

# SOURCES

What the research for this episode read before the script was written. Primary sources first.


1. **Wikipedia, Masoretic Text**
   [primary-adjacent reference]
   https://en.wikipedia.org/wiki/Masoretic_Text

2. **Catholic Encyclopedia, Massorah (1911)**
   https://www.newadvent.org/cathen/10035a.htm

3. **Wikipedia, Vedic chant**
   https://en.wikipedia.org/wiki/Vedic_chant

4. **Wikipedia, Psalm 119**
   https://en.wikipedia.org/wiki/Psalm_119

5. **Wikipedia, Textual criticism**
   https://en.wikipedia.org/wiki/Textual_criticism

6. **Wikipedia, Colophon (publishing)**
   (publishing)
   https://en.wikipedia.org/wiki/Colophon

7. **Wikipedia, History of the Quran**
   https://en.wikipedia.org/wiki/History_of_the_Quran

8. **Yosef Ofer, The Masora on Scripture and Its Methods (De Gruyter preview PDF)**
   https://api.pageplace.de/preview/DT0400.9783110594560_A36072816/preview-9783110594560_A36072816.pdf

9. **Labuschagne, Psalm 119**
   Logotechnical Analysis
   https://www.labuschagne.nl/ps119.pdf

10. **Reconstructing Ancient Literary Texts from Noisy Manuscripts, ACL 2016**
    https://aclanthology.org/W16-0205.pdf

11. **Logion: Machine-Learning Based Detection and Correction of Textual Errors in Greek Philology, 2023**
    https://aclanthology.org/2023.alp-1.20.pdf

12. **krishnamurthys.com, How have the vedas been preserved so accurately**
    http://www.krishnamurthys.com/profvk/srutipage2.html