

Error Aligned

Tyler Shoemaker

ABSTRACT

This essay tracks the digital afterlives of etaoin shrdlu, typographic error turned textual agent. A media effect of Linotypes, this phrase was meant to notify editors that their compositors' fingers had slipped during transcription and a hot-metal line needed to be pulled. It was an internal memo, passed around the printshop — and it is now a recurring text string in digital archives of newspaper pages, where the phrase's accidental inclusion in printed matter has been newly reset by automatic transcription processes. After examining the place of Linotypes in a long history of machine reading, I argue that the presence of this machine's error signal in digital corpora presents an opportunity to consider the extent to which automatic transcription works from an interpretive disposition.

“The Head Compositor nodded. ‘Correct, although in a manner of speaking the operator and the machine are one, in that the operator is a function of the machine and the machine is a manifestation of the operator and both are extensions of the ego of the . . . but I guess that is a little too complicated for you to understand.’”

— Fredric Brown (1943, 69)

CHARLIE WILLIS MEETS GOD, A LINOTYPE COMPOSITOR, WHILE tracking down the typographic errors wreaking havoc on his life. In the fantastic mythology of Fredric Brown’s “The Angelic Anglemorm”, hot-metal composing machines cast the course of our preordained biographies, and the one spelling out Charlie’s fate hitches whenever a bad *e* matrix cycles through it. Thus, these and other supernatural events no one but him believes: pulling an angelworm from a clod of dirt before a fishing trip, halo and all; feeling a sunburn-inducing heat, not hate, during an argument on the street; finding a teal duck in a museum display case, where once sat a Chinese coin called a tael. And thus comes Charlie knocking on the Head Compositor’s printing office door, demanding a revised edition.

Finding the ultimate source of these slips took all but a total commitment to providential inerrancy: for Charlie, every ounce of his weird, ill luck “had to MAKE SENSE!” (BROWN 1943, 62). And it did — to him, if not to those disbelievers around him. But his scripted acts need not have meaning for a class of equally programmed readers who extract, ingest, and re-present printed matter as born-digital text, readers with which the Linotype’s workings are deeply consonant. To transcribe from scanned page images, these automated readers simply compare inked glyphs to reference vocabularies and resort to guesses based on statistical distributions in language, should that comparison end in ambiguity. For them, sense seems to mean little, if anything. Of them, and the general spirit of Charlie’s sense-making insistence, this present essay has much to say. My subject is mechanized word processing, taken two ways, with the first being those readers — or better put, software processes — that otherwise go by the name of computer assisted text transcription. Working under what Mara Mills has called an “assistive pretext” (2010, 39), these processes use optical character recognition, or OCR, to identify characters in images of printed pages and then compose new plaintext files therefrom, transforming digital facsimiles into machine-readable data, ready and waiting for further computational analysis. Constant companion to both the digital humanist and the casual browser of digitized books alike, theirs is a form of reading that can quickly roam into scales we see only at long range, in glimpses and in summaries.

But it is a form of reading with many forebears, all quite legible. In what follows, I take “word processing” to also extend to an assemblage comprised of Linotypes and the compositional practices that accompanied them at the turn of the twentieth century. By that designation, I mean to demonstrate how this assemblage rehearsed much in our contemporary ways of reading alongside machinic readers. Linotypes are word processors, for they were some of the first among text technologies to mechanize character composition. From an initial 1883 patent application on, these machines kept letter molds confined to magazines, assembled them, cast them into lines, and returned them to storage with minimal intervention, save that of a few keystrokes and the pull of a lever. If, with word processing, as Friedrich Kittler once remarked, “we simply no longer know what our writing is doing” ([1995] 2014, 221), the Linotype is a wedge in the beginnings of this rift, one that stretches into present-day systems where machine readers read and write text on our behalf. When those automated readers turn, then, toward the 14.5 million newspaper pages mounted on the Library of Congress’s *Chronicling America* website, a strange, recursive loop develops. Linotypes helped make many of those pages, with the result that OCR

excavates its own history as it scans through them. This amounts to an “epistemological reverse engineering”, in which “media themselves [. . .] become active ‘archaeologists’ of knowledge” (ERNST 2013, 55). I want to use this convergence of Linotypes and OCR to explore the semiotic architecture of machine reading, and my investigations take their cue from Charlie’s exchange with the Head Compositor. If, in that story, mechanization synthesizes machines and their operators, the elliptical break trailing out from that interplay gestures toward what N. Katherine Hayles has called a “cognitive assemblage” (2017, 116), in which sense-making is distributed between both human and nonhuman agents. Sense rises across them, not from one or the other. To make sense of machine reading now, this essay traces the human-machine interactions of Linotypes as those interactions surface across touch-typing manuals, newspaper editorials, trade stories, and literary caricature.

An especially charged site for these interactions is that of the error. As recent bibliographic work on OCR has demonstrated, the historical lineaments of automatic text transcription are most legible during instances where these processes stray from source text.¹ Rather than fixate on the way errorful OCR impedes efforts to construct “clean” corpora, I follow David A. Smith and Ryan Cordell’s recent call to imagine what researchers can do with OCR’s errors, not in spite of them (2018, 10). Errors turn up decisions — design decisions, engineering ones, decisions, too, in both labor management and aesthetics — that designate what reading, and indeed text, should be. This holds equally for discourses that negotiated automation in fin de siècle text technologies as much as it does for OCR, and examining moments where digitized trade stories reflect on the potential impact of errors can potently outline how mechanization more generally augments legibility. My sense is that such discourses, both then and now, are proxies for hermeneutic certitude, what has to “MAKE SENSE!” (BROWN 1943, 62). When an error is under discussion, so too are conditions of, and assurances about, legibility — what, in the case of machine readers, quite literally counts as sense. When an error appears, then, in the output of word-processed letters, texts present an opportunity for us to identify and trace these discussions. Below, I turn to Linotypes and their traces to suggest a continuity in the way these machines’ operators handled their mistakes and how OCR presents errors in datafied textual records. For those earlier word processors, there is a nonsense phrase that thumbnails

1. See TRETTEIN 2013, ALPERT-ABRAMS 2016, and CORDELL 2017.

these negotiations, one that has carried forward into error-prone newspaper records now displayed online: etaoín shrdlu.



Its creators conceived of the Linotype as a parenthetical device, propped between two preexistent print technologies and meant to cement their logics together by blending one with the other. From its earliest stages the composing machine was under a direct order: span the centuries-old practice of printing on a pull press with the nascent, still amorphous typewriter, patented some 400 years after Gutenberg in the late 1860s. American shorthand writer and entrepreneur James O. Clephane issued the command in 1872, announcing his desire to “bridge the gap” between these two writing technologies (ROMANO 2014, 2). After achieving some small renown for his skills in stenography (he served both as secretary to the US Secretary of State William H. Seward and as a court reporter for the Supreme Court), Clephane was approached by Christopher Sholes, one of the first patent holders for a device the latter called the “Type Writer”. Sholes and his associates felt their invention would have immediate advantages for shorthand writing, and, in a moment of nineteenth-century quality assurance measures, they asked Clephane to test the machine and provide consultation on any future improvements they might make to its design.² The stenographer was reportedly harsh in his criticisms of the device but found it promising, and under his direction the first Sholes and Glidden typewriters were built for his employees.

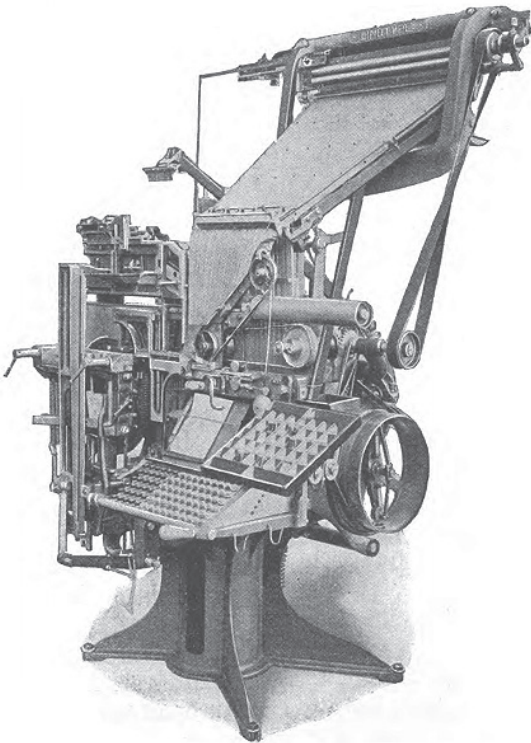
Clephane saw in typewriters possibilities for expediting the publishing process. His involvement with these machines made it clear to him that setting type on presses could neither keep pace with his stenographic notation, nor with the new compositional technique of typing on a keyboard, and Clephane began to explore ways to augment typesetting with that latter technique in mind. Initially, he planned to create a machine that would assemble entire pages of type for inking and printing, much like stereotyping, but difficulties in this design surfaced at every turn, forcing Clephane to settle on using the line as his base unit of print production (ROMANO 2014, 10). The stenographer enlisted fellow inventor Charles T. Moore for the project, and the two of them first experimented with a caster that used papier-mâché matrices (type molds) indented by mechanically assembled characters. But by their own admission, the machine was a failure, and in

2. See FOULKE 1961, 73–5 and Clephane’s obituary in the *New York Times*: “James O. Clephane Dead” 1910.

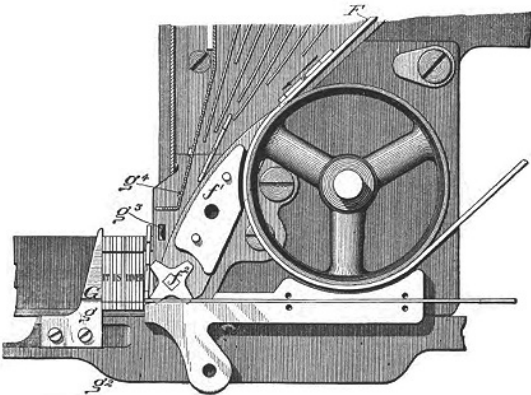
1876 Clephane and Moore reached out to a Washington, D.C. machinist by the name of August Hahl for help. Hahl referred them to his cousin, Ottmar Mergenthaler, a German immigrant who apprenticed as a watchmaker before coming to the States to work in Hahl's shop. Mergenthaler set to work, and after designing several prototypes over the next six years, he had a device ready for a patent submission in 1883.

When a former courtroom stenographer secures the necessary capital to recruit a watchmaker with the task of accelerating the pace of print shops, automation results. Mergenthaler made letter assembly a closed and mechanized loop, locking away the workings of type production into magazine channels, elevators, extruders, and distributor bars — the mere touch of a finger would make them all run. A keyboard sits front right. Each of its ninety keys are individually affixed to cams via corresponding triggers and yokes that, upon those keys' impress and subsequent upstroke, cause the cams to rotate. This engages a long, slender rod that engages an escapement lever. Crescent shaped and so engaged, the lever's bottom half lowers to release a brass matrix down a duct in the Linotype's magazine, while the top of the lever raises to keep in place the next matrix above. On the edges of these matrices are indented characters, ranging from the Latin alphabet to punctuation marks, figures, and ligatures, and from these the Linotype casts its slugs. A compositor presses a key, which drops its corresponding matrix down the magazine and into an assembler. There it waits until, with the pull of a casting lever, that matrix is sent off to the Linotype's extruder. But it can make that trip only once the assembler is full, for the machine Mergenthaler designed remains true to Clephane's original intent and uses complete lines as its basic structural units. To make casts from matrices, compositors need to fill their assembler, and only then can they send their lines to the extruder, where hot lead alloy runs over the molds' impressed surfaces. After making its cast, the extruder then turns and releases a line, or "slug", of raised print characters, type high and ready for inking. For the "simplicity of handling", slugs on this machine are full alphanumeric strings, justified automatically with expanding spacebands (MERGENTHALER Co. 1940a, 11). Hence its name, Linotype, producer of a "line o' type" (*Inland Printer* 1889, 272).

The convolutions of these workings — workings Hugh Kenner once termed "intricate intelligibility" (1986, 10) — should make clear that writing with Linotypes is dispersed, multiplex, and circulatory. They are more like miniature factories than tools, and so from the outset their proper place and use seem better suited for anonymous industrial workrooms, rather than that of the home office. Despite their ubiquity in newspaper



Figures 1a. & 1b.
Linotype illustration (T)
and assembler detail (B)
from Theodore Low de
Vinne's *The Practice of
Typography* (1904).



publishing around the globe, this alone has significantly contributed to these machines' relative invisibility in historical accounts of writing technologies; if Linotypes appear in these at all, it is often so only as footnotes or as quirky modifications to the workflows of print shops. And the yawn of this absence is widened further by a lack of narratives chronicling the Linotype's grand entrance into literary composition. Of it, there are no commonplace stories similar to the one recounting Mark Twain's use of a typewriter to produce his manuscript for *Life on the Mississippi* (1883). Though minor narratives did, in fact, exist during the first decades of its introduction to publishing, the infrastructural position of Mergenthaler's machine has made it a fringe figure in histories of writing.

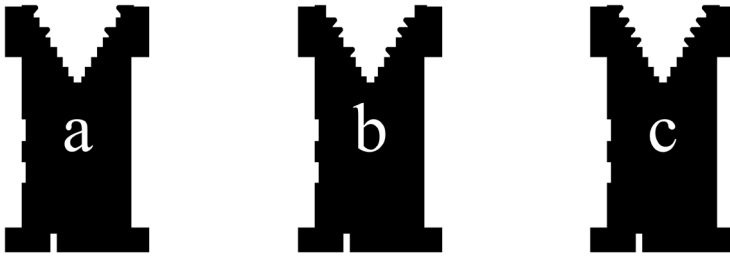
But media archaeologists in particular would do well to consider the enduring salience of the Linotype's assembler (Figure 1b), not only in the context of late nineteenth-century writing technologies, during which the workings of textual input had yet to fully solidify, but also in the long history of computing. I highlight this component because with it, the Linotype puts text into a storage state. It drops letters behind the veil of its magazine cover, briefly keeps those letters strung together, and then sends them off for casting en masse. To be sure, this is not computation — far from it. But in this early instance of automated writing, textual production on Linotypes adheres to the logic of suspended inscription, in which “the stored record of a text is separate from whatever medium or surface on which it is ultimately printed or inscribed in more palpable form” (KIRSCHENBAUM 2016, 46).³ Unlike the pen, the typewriter, or even stereotype plates, letter assembly on this machine does not coincide with those letters' printed production. Indeed, Mergenthaler's matrices never touch paper: after assembling these molds, the machine converts their character information into a different, leaden form and sends matrices back up into its magazine. Like a computer assembler, it translates language across formats.⁴ It should come as no surprise, then, that the use of the term “assembler” to designate a mechanical component predates its arrival into computing by nearly 60 years. The year 1959 marks that later occasion. However, the *Oxford English Dictionary* credits the word's first printed appearance in this sense to a 1902 entry in

3. “Suspended inscription” is Daniel Chandler's term; see “The Phenomenology of Writing by Hand” (1992).
4. An assembler translates symbolically coded instructions (written in assembly language) into those that a computer processor can directly execute (machine code).

the *Encyclopedia Britannica*. Its subject: the Linotype.⁵ With the advent of this machine, assemblers and writing join together in semantic congress.

In his history of word processing, Matthew Kirschenbaum sees the Monotype's coded ribbons as a formative moment for suspended inscription and pits them against Linotype slugs, which the latter created without translating keyboard input (2016, 177fn40). So too does Hayles position Mergenthaler's machine as a foil to James W. Paige's Compositor, a complicated typesetter that, in her account, was able to cognitively differentiate between type pieces and discard defective ones, before redistributing good pieces at will (2018, 1234–5). In comparison to the 18,000 separate parts required to make the Paige Compositor run, the relative simplicity of the Linotype would seem to thoroughly lock it into mere mechanization. But the information conversion that occurs in Linotype assemblers suggests a greater continuity between this machine, the Monotype, and Paige's Compositor, a continuity made all the more suggestive by an encoding system accompanying that conversion. The Linotype also works from code. Mergenthaler developed an encoding scheme that allows molds to return to their corresponding places in a magazine after the Linotype extrudes a line. This made the machine automatic, closing the loop Gutenberg left open between composition and type case redistribution. After casting, a transfer finger pushes used matrices onto an elevator, which shuttles them to the top of the machine, where a shifter lifts them onto a ridged distributor bar. Helical screws propel them along this bar as they hang from grooves, or "teeth", cut into the upper portion of each matrix. While on the course of their travels, certain ridges on the distributor bar end directly above channels in the magazine. When this happens, a matrix falls. Up to seven paired teeth line their inside edges, and cutting them away in different patterns produces 126 "matrix tooth combinations" individually linked to the specific character housed on any given mold — more than enough for the 91 channels a Linotype requires. "When the combination of a given matrix arrives at and meshes with its complimentary distributor bar segment, the matrix is released from the bar and falls by gravity into its respective magazine channel" (MERGENTHALER Co. 1934, 7). On this machine, *e* is distinguishable from *a* not only because of their engraved shapes but because encoded into their very teeth are differences that contour the course of their separate travels through its mechanisms. By means of a simple, forensic dentistry the Linotype pieces its parts back together and the loop Gutenberg leaves open finds its close.

5. See "assembler, n." 5a and 4, respectively.



Figures 2a–c. Silhouettes of Linotype matrices. Tooth combination data taken from *Useful Matrix Information* (1934).

When taken alongside the machine’s assembler, I suggest that this system positions the Linotype as a key node in the genealogy of computational media and opens up the machine’s operations to considerations regarding machinic cognition. Given Hayles’s recent work that demonstrates how cognitive activity can happen among mechanical agents (2017, 20–27; 2018), the Linotype’s active processing of words amounts to a reading operation, one which interprets alongside human operators within a wider cognitive assemblage.⁶ Its tooth combinations enable the circulation of characters from one area of its workings to the next, and they do so by working from a rudimentary form of machine-readable code. Though Mergenthaler could not program his machine to register language as language — which is to say, to register new semantic data and modify its operations accordingly — he equipped it with the ability to separate language elements, query them, and change their address locations during the composition process. To be sure, this is not consciousness, and the code this operation depends upon pales in comparison to the complexities of modern-day computing. But latent in its logics is a programmatological function later forms of code will also share. In both a literal and an idiomatic sense, machine reading cuts its teeth on the Linotype.



6. Hayles’s definition of cognition is “a process that interprets information within contexts that connect it with meaning” (2017, 22). For her, “defining cognition as a process emerging from flows of information and from interpretations of those flows [. . .] invites questions about the nature of meaning and how it differs for human and technical cognizers” (2018, 1240). Challenging an anthropocentric perspective, this expanded sense of cognitive activity “opens meaning making to nonhuman life-forms as well as to technical systems” (2018, 1240).

Of the several dozen patents specifying modifications to the original 1883 Linotype, none take time to explain the logic behind one of the machine's strangest features, its keyboard. "Finger-key" layouts (MERGENTHALER 1885, 1) on this device differ from those of both its contemporaries, such as the Remington II or the Monotype, and those of later machines, including modern day desktops. On these, QWERTY reigns, though only as a de facto standard that comes into full circulation by the early 1900s. Thomas Mullaney explains that as portions of the print industry gradually moved from type cases to typewriters, the very idea of how to organize input characters — to say nothing too of which particular interface was best suited for input in the first place — was often in question (2017, 41–2). The QWERTY layout faced competition ranging from stylus-based interfaces to chorded stenotypes and alternative keyboard layouts; the Linotype is one such example. On it, the constancy of letter frequency grids language production, with redundancy serving as a theory of compositional efficiency. While patents for this machine do not spell out their reasoning, they tacitly assume the benefits of grouping together characters with statistically-high rates of appearance in a sentence, rather than keeping them spread out, interspersed with infrequent candidates like *q* or *x*. In this way a type compositor need not stray far from one area of the keyboard while inputting a line of matrices, reducing extraneous movement and subsequently increasing typing rates. As a result, on the Linotype, *e* sits at the top, left-most corner of its keyboard, followed underneath by *t*, *a*, *o*, *i*, *n*, and then, at the start of a new column, *s*, *h*, *r*, *d*, *l*, *u*.⁷

There are more than just statistical logics undergirding this layout, however. In Mullaney's account, QWERTY took hold only insofar as it was able to suppress non-Latinate writing systems such as Chinese, which manufacturers came to see as the big Other of communication technologies well into the twentieth century (2017, 35–43). Similarly, in the decade following the typewriter's debut, women were often the subject of typing manual instructions, and they consequently learned to type on QWERTY keyboards. Mergenthaler's decision to use an alternative character organization may well have been a response to this. He believed women "ruined the reputation" of Linotypes and made for bad printers, and thus his design

7. While this layout was standard for Linotypes, machines that made their way into non-English printing offices could also be outfitted with different keyboards. In German, this resulted in keyboards whose first 12-character inputs read *e*, *n*, *i*, *a*, *t*, *x*, *r*, *d*, *g*, *o*, *v*, *c*; in Cyrillic, the same range reads *о*, *е*, *н*, *а*, *и*, *с*, *м*, *в*, *ы*, *г*, and *у*.

upheld standing and sexist disparities in gender equity on shop floors and in workforces (GOBLE 1984, 152). More, when women did take part in the printing process, employers would often “keep a woman on straight composition [regularized lines of body text], to make as much as possible an automaton of her” (ABBOTT 1910, 254–5) — precisely the type of composition Linotypes are best suited to mechanize. These machines’ setting of “straight matter” outpaces that of a hand compositor, relegating the feminized labor of producing body text to mechanics.⁸ If these mechanics are to serve, then, as forerunners to computational media technologies, as I have argued, it is also because Mergenthaler’s machine is an anchor point for the gendered logics undergirding clerical work during the era of mainframe computing and now, the labor of digitizing printed matter housed on platforms such as Google Books.⁹

This genealogy is all the more apparent when early twentieth-century touch-typing manuals carry Linotypes and their laborers into digital technologies, often quite literally. Google Books’s Library Partners program lists among its participants the Big Ten Academic Alliance and the University of California, whose holdings contain much of the extant trade literature produced for, and alongside, these machines.¹⁰ Many of these texts are now readily available online, and among their pages one finds narratives cataloging the necessities of adjusting to automation. A certain fervency dominates throughout, with manuals imploring both employers and workers to pay special attention to the way lines are composed. Once letters are no longer tied to their type case boxes, they explained, the horizon of their arrangement rests upon the swiftness of their compositor’s fingers. Proponents of the Linotype claimed it not only removed the need for such boxes, but also opened new opportunities to readjust and train workers’ bodies to the demands of high-output printing. To “set type at high rates of speed requires incessant reading of the copy”, reads a manual collected in Theodore Low De Vinne’s *The Practice of Typography* (1904, 448), continuing,

8. See also THOMSON 1997, 133–58.

9. See HICKS 2017 for a history of the feminized labor in mainframe computing; for an explanation of digitization workflows at Google and its partner libraries, see LOSH 2009, 265–72.

10. For example, HathiTrust, whose contributors often map directly onto the list of participants in Google’s Library Partners program, holds 72 volumes of *The Inland Printer*, which devoted many discussions to trade technology like Linotypes (and fin de siècle arts more generally). The database attributes these holds to libraries at the University of Minnesota and those in the University of California system.

“The keyboard must of necessity be operated without looking at it”, lest a compositor lose their place or produce an error in the line. “Therefore,” the manual concludes, “it is apparent that [. . .] the location of the keys must be so fixed in the operator’s mind that the fingers seek them automatically, and the eyes be devoted to the continuous reading of the copy” (DE VINNE 1904, 448). No room to think: devotion, utter observance of the line and the keyboard articulating it, leaves no place for anything but the mechanical scanning of copy. Eyes are to operate with full and perfect independence from fingers once thought steps out the door. With touch-typing, the medium does not extend cognition, as in McLuhan’s dictum, but rather brackets it so as to better bifurcate work into muscle memory and visual scanning, demanding bodies structured like machines.¹¹

This sentiment appears among many such manuals in late nineteenth-century America, a time during which a new profession Lisa Gitelman (1999) calls “the scribal technician” emerged. If Hayles finds evidence of distributed cognition among the machines of this period, anxieties about the thought patterns of those new mechanical agents are also traceable in an “underlying conflict over how much intelligence the scribal technician had to supply” to the cognitive assemblage of which they were now a part (GITELMAN 1999, 203). There were many open questions about “when and how much the head and the fingers worked,” Gitelman explains, or “just how automatic stenographers, telegraphers, and other scribal technicians had to be” (1999, 203). When, as typing manuals would have it, copy is a matter of the eye and production, that of fingers, said technician’s experience of embodiment would seem to be a purely mechanical affair, properly gauged for both efficient word processing and good typographic aesthetics alike. A certain *Linotype Keyboard Practice* maintains that a compositor’s “subconscious attention to the machine must be such that he constantly produces slugs with clean sharp-printed face and good body, properly trimmed to uniform size” (MERGENTHALER CO. 1940b, 4). Here, a technosomatic continuity — an “intextuation” (CERTEAU 1984, 149) — implicates cognition with copy-text reading and type production. The workings of the machine spread to those of the human, and the facticity of letter frequency comes to govern not only the space of a keyboard layout but the space of discursive manufacture. For Linotype manuals, character assembly is human-machine feedback. It requires bodies to be no more than mechanistic reflexes, a series of inputs and outputs working independently of conscious activity. Type composition turns technosomatic when making

11. See McLuhan [1964] 1994, 3–21.

words “has nothing to do with the ‘I think’” (KITTLER 1999, 189): type with “good body” simply needs to be produced by good (read: disciplined) bodies. The sentiment of such manuals is writ large quite literally across their fingering instructions, inscriptive overlays of letter combinations and governed digits diagramming a body in pieces. . . .

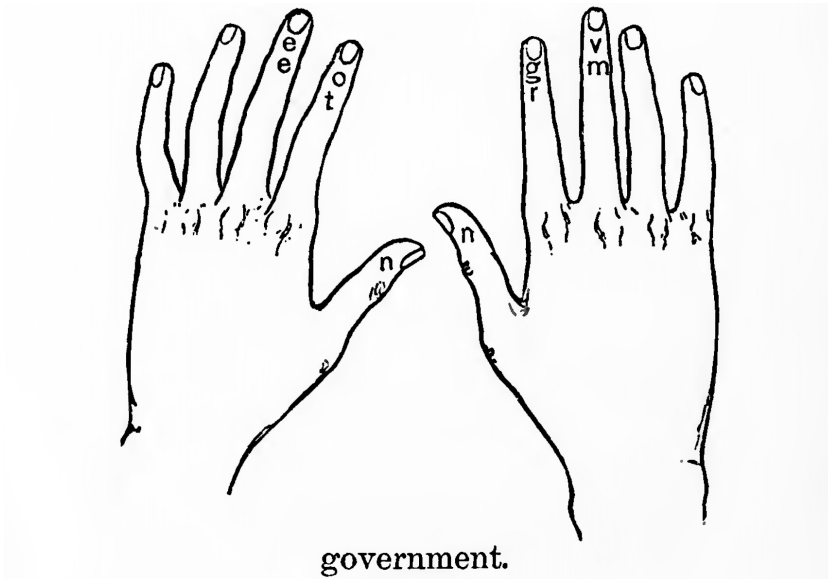


Figure 3. Typing instructions from Theodore Low De Vinne’s *The Practice of Typography* (1904).

So thoroughly set as they are on efficient and effective word processing, these manuals rarely entertain the possibility of failure. But typographic errors were (and are) inevitable, and they especially trouble Linotypes. The very same components that separate keyboard input and character assembly on these machines demand the special handling of typos — or no handling, rather. Pulling a mistake is not particularly feasible when an assembler and magazine channels keep matrices out of human reach; doing so disrupts the loop Mergenthaler introduces into the print shop. All a compositor can do is clear the assembler and start over. This, however, will still activate the machine’s extruder mechanism, producing the incorrect line anyway, so that a typo does not disappear once registered, as on twenty-first-century word processors, but rather becomes all the more weighty —

weighty, and imbued with the potential for someone to overlook the error, place it into a galley, and send it off to be inked and printed. In response to this threat it became common practice for compositors to run their fingers down the first two rows of the machine's keyboard when they noticed their errors, creating a nonsense phrase after the mistake meant to stand out during proofreading. The eye would then snag where the finger had slipped.

The outcome of this practice: etaoín shrldu, pure, leaden letter frequency. Thumping one's fingers down key by key of the 12 most recurrent characters in the English language produces letter salad; bookends to a prior mistake; noise pointing to noise; a clash of awkward syllables strung together only to indicate that they should not be there, that something is wrong, that something needs a fix, a second look, another set of eyes. And once they have caught the eyes of editors these letters are to simply disappear as easily as they came, sliding back into the molten lead alloy above the extruder that cast them, their matrices pulled upwards, back into the guts of the machine, to be released by fingers trained — this time — to touch the right keys.

But traces of these errors persist, and this poses an opportunity to consider how born-digital records can point text mining methods on large corpora toward medium specificity. Despite the canny eyes of compositors (or the intuitions of their fingers), editors would occasionally fail to see those garbled lines of type earmarked for remelting and would instead send them through the rest of the printing process. In newspapers especially, etaoín shrldu appears nestled in paragraphs, hanging under show times and sale prices, or nearly buried underneath photos in their captions. The Library of Congress's *Chronicling America* houses some 485 instances of this letter string, which serve as photo negatives to the positive proclamations otherwise found among digitized typing manuals on Google Books, HathiTrust, and the Internet Archive.¹² Examples occur on digital exemplars from Duluth to Los Angeles: "Now on the ear sounds srish8!tsecaofycea ETAOIN SHRDLU the 'Dead March' [. . .]" (*Labor World*, 13 Nov., 1920); "WHERE THE SHOE PINCHES etaoni Mv etaoín shrldu srdlu cmfwyp [. . .]" (*The Los Angeles Herald*, 17 Nov., 1907). The mistakes that have prompted a compositor to type etaoín shrldu are there too, but the locations of these are murkier for keyword searches, since they have no defined syntax. Etaoín shrldu is easier to find, predictable; it stays gridded to the same patterned

12. The number of instances of etaoín shrldu is likely to increase; *Chronicling America* is an ongoing effort and periodically gains new content. This count is gathered from a search I conducted in September 2018.

structure keeping compositors' fingers in line, and it stays on when those fingers' sentinels fail to keep watch. The phrase persists despite itself, existing by way of a strange irony. Only because an editor fails to see a typo their compositor has recognized will yet another error go to print.

Once it has, the phrase stands as an indexical trace of the human-machine feedback loop that created it. Etaoin shrdlu is an example of what Hans Ulrich Gumbrecht terms the "materialities of communication" (2004, 17–18), a "presence effect" in symbolic work that directly evinces interactions between media technologies and their users. The phrase says that a Linotype was there, and, just as important, there was a practiced and acculturated compositor there using it, re-appropriating the machine's logics to fit both the communicative and material needs of their print shop. German media archaeology has a name for this: etaoin shrdlu signifies one of a multitude of *Kulturtechniken* ("cultural techniques") that accompany media technologies and do so as "ontic operations", operations that quite literally make sense as they pass between human and nonhuman agents.¹³ Where etaoin shrdlu lies on paper, or rendered as a digitized exemplar, a Linotype and its strange rituals are somewhere nearby. Consider the multilingualism the phrase strangely accrues when it appears in *L'Italia* (San Francisco, 1887–1943), which spells out the accommodations Italian printers made when working with English keyboards. "Prima di lasciarsi però gli assessor si scambiarono la parola d'ordine che nessuno doveva st-sa etaoin shrdlu cm nessuno avrebbe parlato con estranei del dissidio, e difatti nessuno parlò" (31 May, 1900; emphasis added). Though it has changed in both material form and textual format, etaoin shrdlu is still present and still signifies. When digitized exemplars of the phrase sit on databases like *Chronicling America*, building a query with an ostensibly dematerialized character string can serve as a pathfinder back to the ink and paper surfaces Linotypes and their operators produced. Querying for an error on digital collections that remove text from paper produces paper trails pointing toward analog media.

Such a query adds a forensic dimension to distant reading practices, putting digital humanists in conversation with both media archaeologists and

13. See SIEGERT 2015, 9–12. Siegert's translator, Geoffrey Winthrop-Young, writes that *Kulturtechniken* is difficult to render into English not because of *Kultur* but because of *Technik*: "Its semantic amplitude ranges from gadgets, artifacts, and infrastructures all the way to skills, routines, and procedures — it is thus wide enough to be translated as *technology*, *technique*, or *technics*" (2015, xv).

bibliographers.¹⁴ That conversation would begin to ameliorate the lack of “data-rich literary history” in computational text analysis by supplying its methods with what Katherine Bode has called a “scholarly edition of a literary system” (2017). Shaping digital corpora, she argues, should be a matter of identifying and representing how literary works “circulated and generated meaning together at particular times and places” (BODE 2017, 94). More, this shaping should self-reflexively make plain the “scholarly infrastructure” that informs a corpus’s bibliographic selections. Showing this infrastructure, Bode explains, changes the question of doing literary history with big data from one that speculates on a future synthesis of readings distant and readings close, to one that instead asks about the appropriate amount of documentation required to articulate historical context in concert with the interpretive decisions that inform corpus construction. Inasmuch as the Linotype’s legibility in digital corpora indicates how particular texts got to where they are, I take this machine to be one such entry in this documentation. Indeed, media technologies are crucial nodes in the scholarly infrastructure Bode discusses, and her projected records of literary systems will need to account for the way such devices enabled circulation in their contemporary milieux and still continue to do so — albeit it after any number of remediations.¹⁵ Identifying the traces of Linotypes with this in mind “translates”, as Bode puts it (2017, 97), methodological achievement into a historical insight that looks both ways: at the past, and at those history-making selections that go on now.



While attention to Linotypes may buttress Bode’s method, her cross-scale and distributive approach to literary history is also essential for animating these machines’ digital afterlives, especially inasmuch as they continue to be marked by *etaoin shrdlu*. As I have already indicated, that phrase indexes more than just the mere presence of Mergenthaler’s invention; so too does *etaoin shrdlu* locate the compositional practices that accompanied these machines. Complicating this, though, is the fact that within 20 years of the Linotype’s introduction into print shops, a diffuse mesh

14. See KIRSCHENBAUM 2008 and, more recently, HUCULAK 2015.

15. An analysis that susses out the full extent of those remediations would need to contend with the performativity of a text, with its place conditionally situated inside a broader “knowledge ecology”, existing “in a co-dependent relation to the cultural systems of production/reception in which it functions” (DRUCKER 2014, 22).

of tropes and characterizations begins taking those very practices as its subject. By the early 1900s, etaoín shrdlu no longer means what it originally meant. Or rather, it widens in semantic sense when the phrase begins circulating as a print shop in-joke, published in editorial quips and short stories, poems and apologies. Where these moments of symbolic work comment upon the role and effects of automated print technologies, they also put pressure on the indexicality for which I have just advocated, diverting it, forking it, sometimes leading it altogether astray. But they do so productively, for these departures offer up so many chances to trace out a recursive play between technology, technique, and trope that distributes the presence of Linotypes across digital collections.¹⁶ As I discuss below, the *Kulturtechniken* that accompanied those machines serve as key frames for understanding bibliographic criticism's relationship to automatic text transcription in our present moment.

Exhibits A, B, and C are editorials, instructions advising readers on the proper interpretation of etaoín shrdlu. "Using 'Etaoín Shrdlu Cmfwyp' for a headline", writes the "Jayhawker Jots" section of the *Topeka State Journal*, "the Sedan Times-Star hastens to explain that it is only 'linotype' for a brand new set of matrices" (23 April, 1914). So says A. B. for the Salem, Oregon *Daily Capital Journal*, "If the war correspondents would just substitute 'Shrdlu' and 'Etaoín' for some of those badly spelled and never pronounced names of men and places, it would be as intelligible to the readers, more simple for the editors and a joy to the linotype operators" (7 Oct., 1916). And C: the Crystal Falls *Diamond Drill* makes a plea: the "linotype operator has a 'Volapuk' all his own. Every time he makes a 'pi' line it is 'etaoín shrdlu shrdlu.' So dear reader, whenever you are reading along and come to a couple of 'etaoín shrdlus' don't let it bother you — it's only the operator's way of telling his troubles to his fellow craftsman. Just hurdle over the 'shrdlus' and proceed with the story" (20 July, 1918).

In one of the earliest instances of this self-conscious usage, a poem (Figure 4) begins a round of reprintings after its initial publication in the May 1903 edition of *The Inland Printer*, a key trade journal that often reported on developments in the American print industry. First attributed to a Chicago letter artist by the name of Edgar Yates, *The Deadly Pi Line* quickly loses its byline in later printings across *Chronicling America*, which catalogs its

16. One such instance of this recursion is a particularly astute essay of Whitney Trettien's (2013), in which she ties the "zombie-like" materials of algorithmically generated print-on-demand books back to nineteenth-century reprints of Milton's *Areopagitica*.

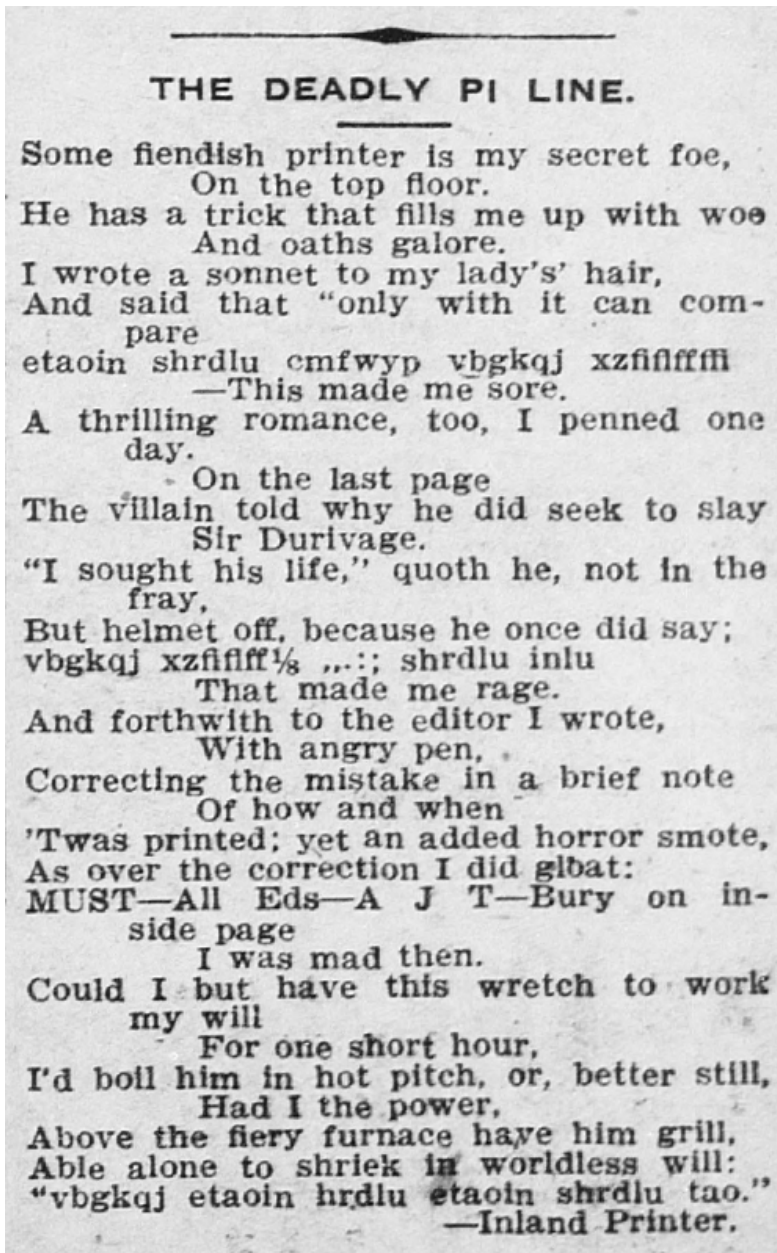


Figure 4. *The Deadly Pi Line*, in *The St. Paul Globe* (3 October, 1903).
 Available on Chronicling America at <https://chroniclingamerica.loc.gov/lccn/sn90059523/1903-10-03/ed-1/seq-7/>.

travels to Indianapolis (29 May), Marshalltown, Iowa (8 June), Minneapolis (15 June), and St. Paul (21 June). Further printings in the California Digital Newspaper Collection and HathiTrust see it go as far as Los Angeles (10 June) and Augusta, Georgia (August edition) before returning to the *St. Paul Globe*, where it makes a final appearance in the October 3rd edition of 1903. Generically, the poem follows the conventions of “fugitive verse”, which featured “narratives of authorship that provided lurid interest, fostered sentimental identification, or otherwise helped readers connect” with poems as they were reprinted across newspaper publications (CORDELL and MULLEN 2017, 43). Here, that identification broaches the conditions of newspaper manufacture. The poem’s speaker pens and publishes a sonnet “to my lady’s hair”, merely to have the second half of a simile frustrated by gibberish: “only with it can compare / etaoín shrdlu cmfwyp vbgkqj xzfllffffi”. A simple mistake, perhaps, but a subsequent attempt at writing “a thrilling romance”, and then a letter to the editor written “with angry pen” both suffer similar bouts of logorrhea. Fuming, since typographic errors seem to mar any attempts at writerly expression, the poem’s speaker fantasizes about all they would do to their “secret foe”, if given the chance. *The Deadly Pi Line* ends with their plotting:

Had I the power
 Above the fiery furnace have him grill,
 Able alone to shriek in wordless will,
 vbgkqj etaoín hrdlu etaoín shrdlu tao.

Whether readers are to understand this final line as a tormented howl or further ironic censorship, one cannot tell: they are one and the same after so many wordless phrases have muddled the print copy from start to finish.

“Pi Line”, not “By Line”. That this poem of frustrated writing registers a typographic error at the site of authorial attribution suggests other forces shaping intent and the production of meaning beyond the Romantic ideal of a transcendent Author. “Pi Line”, not “By Line”, because the poem harkens back to Gutenberg’s logic of printing, where type sorts are in high demand and a printer may have need to substitute the phonetically identical but visually distinct *i* for *y* — in type cases, there are only so many of each letter. “Pi Line”, not “By Line”, because this displacement spells out the numerical value π , that ratio of circumference to diameter that, in the idiom of the print shop, means a line has been jumbled up. To “pi the type” is to remix characters so their order resembles that random sequence of numbers after the decimal point of π . And ostensibly, the errors this poem’s

meditation on typos highlights are themselves infinite, errors constituting a series of random permutations and chance occurrences that foreclose the effective transfer of idea to page to public sphere. “Pi Line”, not “By Line”, because *The Deadly Pi Line* presciently figures poststructuralist theories of the author’s death — Foucault or Barthes: your choice, it is all the same to the free play of mechanically processed language.

If, in this poem, troublesome printers threaten authorship, the etaoin shrdlu trope will soon lay siege to print shops as well. The phrase comes alive in the form of a proper name, and as it begins to move beyond the editorial page and into the print industry more generally, it brings along anxieties about automating trade work. In a parody of the disciplined bodies of touch-typing, Elmer Rice’s 1923 play *The Adding Machine* sees its protagonist, Mr. Zero, meet a certain Mr. Shrdlu in the Elysian Fields. The former had been hanged for killing his boss, upon discovering his employer would soon replace him with an adding machine, while the latter, a morose copy-editor, snapped and murdered his mother during Sunday dinner. Details of that grim meal were published far and wide, for newspapers record the sins of typos and those of murderers alike.

Zero. I remember readin’ about you in the papers.

Shrdlu. Yes, my guilt has been proclaimed to all the world.

(RICE [1923] 1965, 38)

Because of his actions, Shrdlu lands in purgatory, where he is to remain “until I understood” (RICE [1923] 1965, 43) — until, that is, he can work through the psychic break that drove him to run a knife across his mother’s neck, a break as illegible to him as the letter salad mimicking those shrieks of pain in *The Deadly Pi Line*. Until he, like Charlie Willis, can “MAKE SENSE” of these events (BROWN 1943, 62), here he will wait. Much as the typographic errors Shrdlu was to watch for in shoe catalogues stay inked on printed pages, the copyeditor stays in an afterlife limbo for his indelible crimes.

In the years after Shrdlu’s confinement, a Linotype named “Etaoin Shrdlu” gains sentience in another of Fredric Brown’s short stories and puts George Ronson, typesetter, to work: “the Linotype no longer worked for him; he was working for the Linotype” (1954, 61). “Or”, as Ronson’s friend Walter suggests, the machine was merely “interested in learning. And it read by assimilating the process of typesetting” (BROWN 1954, 61). Machine reading ends with machines reading — a situation to which I will soon turn. “Etaoin and Shrdlu” by Anthony Armstrong follows in tow,

along with newspaper typesetter Mr. Etaoin in Charles G. Finney's *The Circus of Dr. Lao* and a mischievous Étienne Cherdlu in Thomas Pynchon's "The Secret Integration". Across the scriptural economy whose expansion Linotypes underwrite, the etaoin shrdlu myth takes hold, and by means of the very same speculative registers these stories so often use, nonsense is named.¹⁷ Eventually, that named nonsense will come to rely solely on those registers, for by the midcentury, other print technologies like phototypesetting begin making substantial gains on the Linotype's lead in newspaper publishing. The errors only that earlier machine produces must migrate into the mythic space of literary reference if they are to stay alive. The typographic error tropologically figured: in this form, the myth lives on.

Right up until the Linotype's death knell, compositors continued using etaoin shrdlu to mark off mistakes. The phrase leads a double life, and its strange polysemy necessitates the documentary records Bode proposes, for when typo crosshatches with trope, querying for this letter string in digital corpora does not always lead to a mistake. Etaoin shrdlu stays uncertain. The mythic structure of its surplus significations blurs the indexical traces of Linotypes, while its joint use as error and errant signifier frustrates its unambiguous reading among collocations and topic models. When it shows up in these, it remains noisy, undetermined, like a probability space. Letter frequency indicating a statistical distribution of topicality encoded into machine reading from 1883 on: this is the trace Linotypes leave on pages and in files. To "MAKE SENSE" (BROWN 1943, 62) of these machines in digital collections, bibliographic forensics on computational platforms must supplement evidentiary claims with speculation.



Such supplementarity marks a wider condition of working with automatically transcribed records, one that arises from the probability spaces in which optical character recognition itself works. Numerous OCR methods have been in use since the early twentieth century, but on computational media these processes generally implement template matching, feature extraction, or a mix of both to generate data from print sources. In template matching, software engines compare the overall shape of a glyph

17. Here I have in mind de Certeau's concept of myth: "fragmented discourse which is articulated on heterogenous practices of a society and which also articulates them symbolically" (1984, 133–4). See also Lisa Gitelman's use of the scriptural economy in her work on the "embarrassment of material forms" that surged into use with the advent of job printing in nineteenth-century America (2014, 6).

with exemplars in a reference vocabulary to make their identifications, while those using feature extraction isolate smaller, more idiosyncratic features such as a letter's line intersections or curvatures during comparison. After an initial scan both methods compile a short list of potential character matches for every printed mark and advance hypotheses therefrom. As with Linotype keyboard composition, statistical distributions of letter and word frequencies often aid in making these decisions: if they cannot discern a glyph's outlines, OCR engines consult frequency lists to probabilistically guess which word or letter they should compile next. To these matches software engines add a confidence rating and then move on, inputting characters into a plaintext readout, all with their own ratings.¹⁸

Collectively, these matches amount to varying degrees of confidence. Because statistical distributions in corpora inform OCR's transcriptional matches, this process is error-prone, susceptible to misrecognition. Its guesses do not always line up with their paired page images, and Ryan Cordell has argued that digitized facsimiles and the born-digital data derived therefrom amount to two separate editions of a text. OCR, he writes, is a compositor "setting text in a language it does not comprehend" (2017, 196) — and, I would add, it is in this sense an heir to the automatic writers touch-typing once demanded. Composition in machine reading unthinkingly inputs characters from proof texts to generate statistically informed readout containing differences from those sources. It produces new bibliographic objects with no guarantee that these will maintain utter fidelity to their imaged variants during a side-by-side inspection. Those objects are, as Hannah Alpert-Abrams argues, "interventions" in the historical record of a text, not transcriptions (2016, ¶ 34), and claims to the contrary uphold a reigning "myth of surrogacy" in digitization (MAK 2014, 1520), which treats the presence of computational processes as a certifying seal for completeness, comprehensiveness, and accuracy. Here, *etaoin shrdlu* is instructive: beyond just serving as their pre-digital forebear, the mythic status of that phrase demonstrates the need for a skeptical view of those substitutions. Linotypes, surrogate and automated word processors, probabilistically generated words that no compositor would otherwise type, and readers were left to contend with the task of assimilating nonsense into discourse. The joint force of mechanized word processing and *etaoin shrdlu* produced semantic residue, and now, a correlative excess in OCR keeps

18. For an overview of how these ratings impact analyses of newspaper archives, see HOLLEY 2009. See also SMITH 2007 for an in-depth explanation of how Tesseract, one of the most widely used OCR engines, works.

automatically generated text in an approximate, supplementary relationship to the page images from which it derives.

Inasmuch as my own thinking focalizes this errancy through its Linotype forebears, I want to suggest that the enduring traces of these machines are best suited to unravel the implications of that last and most important letter, R, in OCR. As it both replicates and misprints printed sources, automatic text transcription adheres to the complex and shifting sites of recognition etaoín shrdlu first marks. OCR readouts are above all catalogues of recognized and recognizing agents, and they “inscribe the scene of their production into plaintext forms” (ALPERT-ABRAMS 2016, ¶ 34).¹⁹ Such scenes remain open to interpretation in ways etaoín shrdlu underscores. Consider a key feature of the phrase, which continues to haunt its digital exemplars, even if curatorial efforts manage to account for its joint use, figured or mis-fingered: while it is an error, it also represents yet *another* error to which it is meant to point. In this way etaoín shrdlu reflects warnings retrospectively, where, looking back, one notices anomalies without being able to pinpoint them directly. Beyond generalities, the phrase can do little more. It simply indicates that an editor needed to revisit the trace of a mistyped key or series of keys within the last 30 to 42 picas of newly forged lead before ink can meet paper. There lies, then, a hidden error on the page, undisclosed, intimated but unrevealed, an errant string that gives rise to the letter salad accidentally appearing at present in *Chronicling America* and elsewhere. Etaoín shrdlu is an unrecognized error stemming from an error previously recognized, and this prior mistake, that one that accompanies etaoín shrdlu as its initial catalyst, now haunts the semiotic architecture of machine reading.

It haunts machine reading both literally and because the two errors etaoín shrdlu indicates produce the very same signifying structure — and signifying is the word — that errant OCR produces now. Anytime automatic transcription outputs text that diverges from its sources, it recreates the recursive chain of recognition editors working alongside Linotypes faced, in which nonsense text calls out to readers and points them elsewhere in a document. For those pages produced by Linotypes, that location

19. Alpert-Abrams’s own case study details how the automatic transcription of colonial contact narratives often reproduces the very same marginalization such narratives enforce. “Automatic transcription, itself a mechanical and practical tool,” she writes, “also and simultaneously participates in this transfer of power, with practical consequences for scholarly work and our work as actors in the public sphere” (2016, ¶ 10).

was often earlier in a line; for errorful strings in OCR, substitute characters lay in place of a page image’s direct reproduction, and that “elsewhere” is no longer present in the text file itself. In both, error results from, indexes, and demands further recognition, which shifts the function of OCR from transcription to that of a hermeneutic act — or better, indicates that from the very start OCR entails a reading operation. Following Alpert-Abrams and Cordell, if OCR transcriptions intervene in the historical record, the digital afterlives of etaoïn shrdlu show how the mode of that intervention has, at base, an interpretive disposition. We readers of those automatic readouts must learn to recognize, and then to read alongside and within, the slippery probability spaces of machines reading. The table below gives four such examples in *Chronicling America* that necessitate this recognition. In the first two, the afterimage of leaden letter frequency indicates an error forensically identifiable in page images but masked in plaintext by errors produced during the statistical analyses informing decisions in automated reading. New typos replace existing typos. And the statistical logics that make that replacement possible have, in this table’s third and fourth entries, produced “etaoïn” where no such etaoïns were printed — produced them in an anticipatory move, in fact, for these newspaper pages appear well ahead of the Linotype’s debut.

Table 1. *Chronicling America* OCR interpretations. The first two entries register misprinted misprints; the third and fourth, anticipatory replacements in text printed before the Linotype’s invention.

Newspaper	Original Text	OCR Interpretation
<i>Misprinted Misprints</i>		
<i>The Irish Standard</i> 13 February, 1897 (2)	2 cows etaoïn shrdlu cmfpetaola sbrdlu cmfp
<i>Virginian-Pilot</i> 25 July, 1900 (7)	SOLD etaoïn shrdlu cmfwypvb	SOLD elaoïn slirdlu omfwypvb
<i>Anticipatory Replacements</i>		
<i>New-York Daily Tribune</i> 3 April, 1844 (1)	The public are invited to call and examine them — also, all other kinds of Sofas always on hand.	The public are invited to call and etaoïns them?also, ill cther lundi of .-> .fi* liway? ou hand.
<i>New Orleans Daily Crescent</i> 8 July, 1852, morning (7)	Agents Louisiana Dry Dock Co. / New Orleans, February 4, 1852	N...O.in., Febrn..y, IILS. etaoïn .rDok.

Under these conditions, interpretive conjecture must accompany computational forensics. When, as with automatic text transcription, a process poses and resolves decision points before any output goes on to human eyes, our interpretive activities are channeled along through cognitive assemblages that perform semantic difference ahead of our review. OCR engines “MAKE SENSE” (BROWN 1943, 62), despite their being not conscious of that fact. Though they may not comprehend what they set, with OCR processes we are nevertheless faced with an uncanny, almost unthinkable situation in which comprehension is no longer a necessary and sufficient condition for hermeneutic activity.

In this sense, that the following appears as a header for every plain-text view of newspaper pages in *Chronicling America* is entirely apposite. There, errant letters leave their columns for an HTML render, finding themselves set not by a Linotype but by a web browser. Above them; above the page image from which they have derived (diverged); above a link that asks, “What is OCR?”; above a “persistent link” that directs readers to the present born-digital edition; above information specifying the state collection in which the aforementioned page image lies; above that page image’s title, publisher location, span of publication, and its particular date, there is a phrase that both describes the transcription below it and that names a mode of nonconscious reading whose implications we have yet to fully think through: “OCR Interpretation”.

University of California, Santa Barbara

Works Cited

Sources from Chronicling America

- Daily Capital Journal*. 7 October 1916. Salem, Oregon. <http://chroniclingamerica.loc.gov/lccn/sn99063957/1916-10-07/ed-1/seq-4/>.
- The Diamond Drill*. 20 July 1918. Crystal Falls, Michigan. <http://chroniclingamerica.loc.gov/lccn/sn96076817/1918-07-20/ed-1/seq-6/>.
- Irish Standard, The*. 13 February 1897. Minneapolis, Minnesota. <http://chronicling-america.loc.gov/lccn/sn90059959/1897-02-13/ed-1/seq-2/>.
- Labor World*. 13 November 1920. Duluth, Minnesota. <http://chroniclingamerica.loc.gov/lccn/sn78000395/1920-11-13/ed-1/seq-1/>.
- L'Italia*. 31 May 1900. San Francisco, California. <http://chroniclingamerica.loc.gov/lccn/sn85066408/1900-05-31/ed-1/seq-1/>.
- Los Angeles Herald*. 17 November 1907. Los Angeles, California. <http://chronicling-america.loc.gov/lccn/sn85042462/1907-11-17/ed-1/seq-12/>.

- New Orleans Daily Crescent*. 8 July 1852, Morning. New Orleans, Louisiana. <http://chroniclingamerica.loc.gov/lccn/sn82015753/1852-07-08/ed-1/seq-7/>.
- New-York Daily Tribune*. 3 April 1844. New York, New York. <http://chroniclingamerica.loc.gov/lccn/sn83030213/1844-04-03/ed-1/seq-1/>.
- St. Paul Globe*. 3 October 1903. St. Paul, Minnesota. <http://chroniclingamerica.loc.gov/lccn/sn90059523/1903-10-03/ed-1/seq-7/>.
- Topeka State Journal*. 23 April 1914. Topeka, Kansas. <http://chroniclingamerica.loc.gov/lccn/sn82016014/1914-04-23/ed-1/seq-4/>.
- Virginian-Pilot*. 25 July 1900. Norfolk, Virginia. <http://chroniclingamerica.loc.gov/lccn/sn86071779/1900-07-25/ed-1/seq-7/>.
- Wichita Daily Eagle*. 22 March 1895. <http://chroniclingamerica.loc.gov/lccn/sn82014635/1895-03-22/ed-1/seq-3/>.

Printed Sources

- ABBOTT, Edith. 1910. *Women in Industry: A Study in American Economic History*. New York: D. Appleton and Company.
- ALPERT-ABRAMS, Hannah. 2016. "Machine Reading the *Primeros Libros*". *Digital Humanities Quarterly* 10.4.
- BODE, Katherine. 2017. "The Equivalence of 'Close' and 'Distant' Reading: or, Toward a New Object for Data-Rich Literary History". *Modern Language Quarterly* 78.1: 77–106.
- BROWN, Fredric. 1954. "Etaoin Shrdlu". *Angels and Spaceships*. New York: Dutton.
- . 1943. "The Angelic Angleworm". *Unknown Worlds* 6.6: 49–70.
- CERTEAU, Michel de. 1984. *The Practice of Everyday Life*, translated by Steven F. RANDALL. Berkeley: University of California Press.
- CHANDLER, Daniel. 1992. "The Phenomenology of Writing by Hand". *Intelligent Tutoring Media* 3 (May–August): 65–74.
- CORDELL, Ryan. 2017. "'O i-jtb the Raven': Taking Dirty OCR Seriously". *Book History* 20: 188–225.
- CORDELL, Ryan, and Abby MULLEN. 2017. "'Fugitive Verses': The Circulation of Poems in Nineteenth Century American Newspapers". *American Periodicals: A Journal of History & Criticism* 27.1: 29–52.
- DE VINNE, Theodore Low. 1904. *The Practice of Typography: Modern Methods of Book Composition; a Treatise on Type-setting by Hand and by Machine and on the Proper Arrangement and Imposition of Pages*. New York: The Century Company.
- DRUCKER, Johanna. 2014. "Distributed and Conditional Documents: Conceptualizing Bibliographic Alterities". *MATLIT* 2.1: 11–29.
- ERNST, Wolfgang. 2013. *Digital Memory and the Archive*, edited by Jussi PARIKKA. Minneapolis: University of Minnesota Press.
- FOULKE, Arthur. 1961. *Mr Typewriter: A Biography of Christopher Latham Sholes*. Boston: Christopher Publishing House.

- GITELMAN, Lisa. 2014. *Paper Knowledge: Toward a Media History of Documents*. Durham: Duke University Press.
- . 1999. *Scripts, Grooves, and Writing Machines: Representing Technology in the Edison Era*. Palo Alto: Stanford University Press.
- GOBLE, Corban. 1984. "The Obituary of a Machine: The Rise and Fall of Ottmar Mergenthaler's Linotype at U.S. Newspapers". Doctoral dissertation, Indiana University. Ann Arbor: University Microfilms International.
- GUMBRECHT, Hans Ulrich. 2004. *The Production of Presence: What Meaning Cannot Convey*. Palo Alto: Stanford University Press.
- HAYLES, N. Katherine. 2018. "Human and Machine Cultures of Reading: A Cognitive-Assemblage Approach". *PMLA* 133.5: 1225–1242.
- . 2017. *Unthought: The Power of the Cognitive Nonconscious*. Chicago: Chicago University Press.
- HICKS, Marie. 2017. *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. Cambridge, MA: The MIT Press.
- HOLLEY, Rose. 2009. "How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs". *D-Lib Magazine* 14.3–4.
- HUCULACK, J. Matthew. 2015. "Reading Forensically: Modernist Paper, Newfoundland, and Transatlantic Materiality". *The Journal of Modern Periodical Studies* 6.2: 161–190.
- "James O. Clephane Dead". 1 December 1910. *The New York Times*. ProQuest Historical Newspapers (accessed 18 November 2017).
- KENNER, Hugh. 1988. *The Mechanic Muse*. Oxford: Oxford University Press.
- KIRSCHENBAUM, Matthew. 2016. *Track Changes: A Literary History of Word Processing*. Cambridge, MA: Harvard University Press.
- . 2008. *Mechanisms: New Media and the Forensic Imagination*. Cambridge, MA: The MIT Press.
- KITTLER, Friedrich. 1999. *Gramophone, Film, Typewriter*, translated by Geoffrey WINTHROP-YOUNG and Michael WUTZ. Palo Alto: Stanford University Press.
- . [1995] 2014. "There Is No Software". *The Truth of the Technological World: Essays on the Genealogy of Presence*, edited by Hans Ulrich GUMBRECHT and translated by Erik BUTLER, 219–229. Palo Alto: Stanford University Press.
- "The 'Linotype' Composing Machine". 1889. *Inland Printer* 7.3: 272.
- LOSH, Elizabeth. 2009. *Virtualpolitik: An Electronic History of Government Media-making in a Time of War, Scandal, Disaster, Miscommunication, and Mistakes*. Cambridge, MA: The MIT Press.
- MAK, Bonnie. 2014. "Archaeology of a Digitization". *Journal of the Association for Information Science and Technology* 65.8: 1515–26.
- MCLUHAN, Marshall. [1964] 1994. *Understanding Media: The Extensions of Man*. Cambridge, MA: The MIT Press.
- MERGENTHALER LINOTYPE COMPANY. 1940a. *Linotype Keyboard Practice*. Brooklyn: Mergenthaler Linotype Company.

- . 1940b. *Linotype Machine Principles*. Brooklyn: Mergenthaler Linotype Company.
- . 1934. *Useful Matrix Information*. Toronto: Canadian Linotype Limited.
- MERGENTHALER, Ottmar. 1885. US Patent 317, 828.
- MILLS, Mara. 2010. "Deaf Jam: From Inscription to Reproduction to Information". *Social Text* 28.1: 35–58.
- MULLANEY, Thomas. 2017. *The Chinese Typewriter*. Cambridge, MA: The MIT Press.
- OXFORD ENGLISH DICTIONARY. "assembler, n.". Oxford University Press (accessed 26 October 2017).
- RICE, Elmer. [1923] 1965. *The Adding Machine*. 3 Plays, 1–62. New York: Hill and Wang.
- ROMANO, Frank. 2014. *History of the Linotype Company*. Rochester: Rochester Institute of Technology Press.
- SCHANTZ, Herbert F. 1982. *The History of OCR*. Vermont: Recognition Technologies User Association.
- SIEGERT, Bernhard. 2015. *Cultural Techniques: Grids, Filters, Doors, and Other Articulations of the Real*, translated by Geoffrey WINTHROP-YOUNG. New York: Fordham University Press.
- SMITH, David A., and Ryan CORDELL. 2018. "A Research Agenda for Historical and Multilingual Optical Character Recognition". <http://hdl.handle.net/2047/D20297452>.
- SMITH, Ray. 2007. "An Overview of the Tesseract Engine". *Proceedings of the Ninth International Conference on Document Analysis and Recognition*: 629–633.
- THOMSON, Ellen Mazur. 1997. *The Origins of Graphic Design in America, 1870–1920*. New Haven: Yale University Press.
- TRETTIEN, Whitney. 2013. "A Deep History of Electronic Textuality: The Case of English Reprints Jhon Milton Areopagitica". *Digital Humanities Quarterly* 7.1.