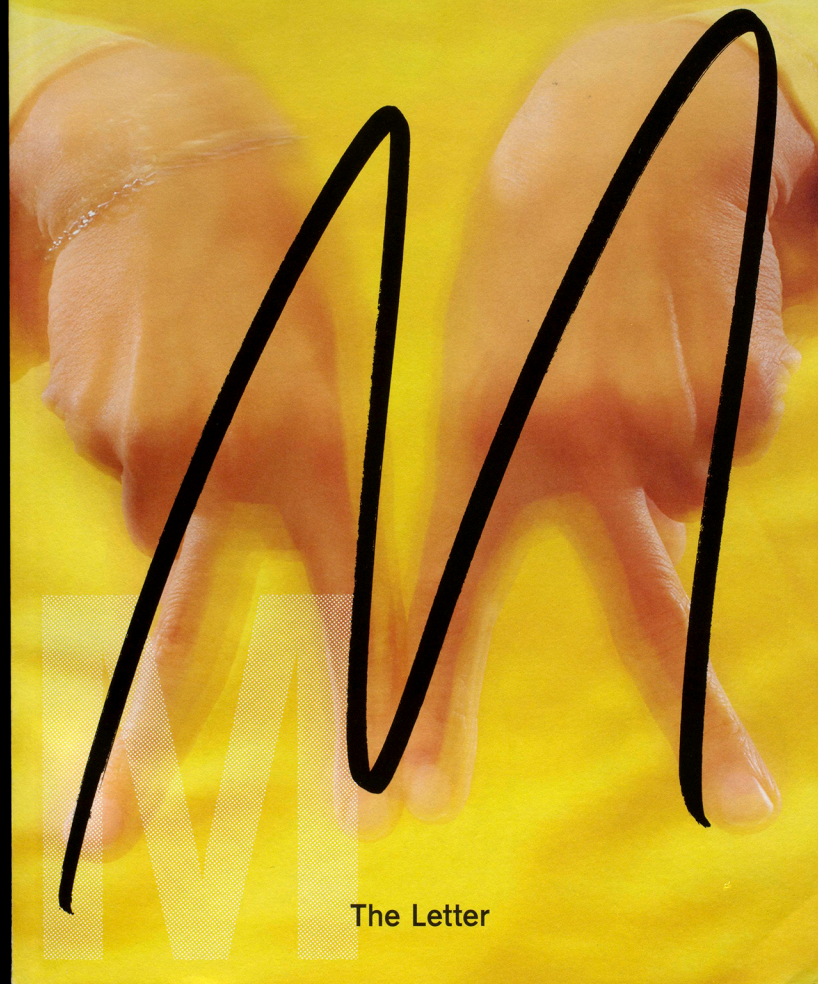


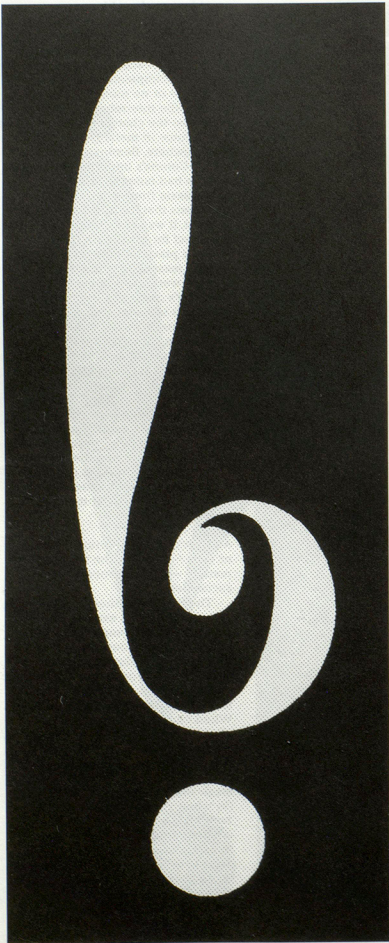
MacGuffin

The Life of Things

N° 13



The Letter



Unicode Consortium, they had created one character set to rule them all: Unicode 1.0 contained north of 28,000 characters drawn from the most popular character sets then in use and included a host of additional characters to boot.

But Unicode was not a free-for-all. The consortium warned that it did not entertain "rare, obsolete, idiosyncratic, personal, novel, rarely exchanged or private-use characters" — an admirable stance, although one that has sometimes led to inexplicable exclusions from the standard. Over the years, a number of characters, symbols, and even entire scripts have been left out for one reason or another. (A 2015 article on the subject in *Model View Culture* magazine, by Aditya Mukerjee, was entitled "I Can Text You A Pile of Poo, But I Can't Write My Name.")

For reasons that remain unclear, the interrobang, a character that was simultaneously rare, idiosyncratic, personal, novel, and rarely exchanged, managed to sneak into the very first version of Unicode. And, since the consortium almost never removes characters, there it has stayed ever since. This is the reason that I can copy-and-paste* ? into a tweet or an email and readers across the world will see the same symbol that I do.

Martin Speckter died in 1988, just a few years shy of the interrobang's Unicode investiture. Penny lived until 2020, sustained by a daily scotch and soda, and was formidable to the last in her promotion of her husband's creation. She ran a website dedicated to the mark, gave generously of her time to those interviewers who periodically rediscovered it, and accessorized her immaculate outfits with a gold interrobang lapel pin as she did so.

As for the interrobang, it has achieved a kind of accidental immortality etched in the binary of the Unicode standard. No-one may ever use it again, but it will live on all the same, to be forgotten and rediscovered every decade or two by a new generation of users.

* Good luck finding one on a keyboard; the interrobang may live on in Unicode but it has not yet broken into the hallowed QWERTY layout.

ERROR-FREE WRITING, AT LEAST MOST OF THE TIME

Ostensibly an innocent tool to eliminate typos, autocorrect turns out to be a normative force that is forever overwhelmed by the idiosyncratic and unpredictable way humans use language.

Text by Isabel Seiffert and Christoph Miler

Most of us have been there. Eager to kill a few minutes at the bus stop, in the supermarket queue or the doctor's waiting room, we grab our smartphone and start texting. Letter by letter we touch and tap, the movements of our thumbs trying to keep pace with our fleeting thoughts. Still in the flurry of messages, we hit the send button, only to realize seconds later that our hand-crafted words were jumbled into nonsensical gibberish by autocorrect.

*A barking dog turned into a barking dad.
Learn to drive turned into like to drink.*

As the rather embarrassing mess dawns, we start to correct swiftly. Sometimes, this makes the situation worse.

*Dear husband turns into dead husband.
Fiancé turns into Beyoncé.*

In light of such devastating moments, one might be tempted to call autocorrect immature, buggy, or simply useless. When *pools* turn into *poops* once too often we want to throw our phone out of the window and never use it again. Yet autocorrect's infamous glitches overshadow its intended convenience within our digital lives, designed around the ubiquitous usage of smartphones and their virtual keyboards. Like it or not, our podgy fingertips simply exceed the size of the teeny-tiny and immaterial keys we are trying to hit. It's a fundamental physiological mismatch. Software engineers call the resulting mistypes "fat-finger errors" and hail autocorrect as their cure, capable of adjusting *of to if*, *whar* to *what* and *teh* to *the*.

Autocorrect never sleeps, as we dash off emails, to-do lists and love messages. It checks our writing against a dictionary and potential spelling mistakes, always considering the keyboard proximity of our strokes and the linguistic context of our message. Based on those variables, wrong words are spotted and their most likely replacements suggested. And if autocorrect is really sure about what we mean, it goes ahead and replaces our words on the fly, making our writing more error-free without anybody noticing, at least most of the time.

You're book

Ken Kocienda, who coded Apple's first version of autocorrect, described it as a powerful linguistic calculator that enables "countless trivial but human moments", while techno-romanticists even praise it as the overlooked ghost-writer of our era of mobile proximity. But don't be dazzled. Today's manifold versions of autocorrect are much more than language machines in service of humanity. Operating in the background of nearly every writing interface — from word processing apps to e-mail programs to Google's input field —, they do much more than replace words for the better. Cast into the mould of our smartphones, these algorithmic agents tell us what and how to write, thereby enforcing specific norms of spelling and notions of standard language. Everything outside this frame — such as particular versions of slang, vernaculars or dialects — is ruled out. Ever tried spelling the TikTok term *blud* (for 'bro') to grudgingly battle a stubborn autocorrect that insists on *blue*?

Looking back to the early days of autocorrect makes it seem almost inevitable that a one-sided version of correctness would be baked into the software from the start. Introduced by Microsoft as a brand-new feature of Word 6.0 in 1993, it was advertised as fixing your typos faster than you can make them. However, autocorrect wouldn't do so by checking your writing against an official Oxford Dictionary or the like. Instead, it compared your words with a list of 'supposedly incorrect' terms hand-picked by Microsoft's own engineers and their interns.

The initial list was compiled by Dean Hachamovitch, whose team invented autocorrect and anticipated a number of prosaic slips of the pen, like *dEar*, *friday*, or *SHip*. Later, autocorrect's linguistic worldview was refined by feeding it private dictionary entries from Microsoft's employees. In parallel, hundreds of supposedly obscene terms — like *asshole*, *butthead*, and even *booger* — were removed from the list by Hachamovitch's trainee. Microsoft decided that it didn't want to recommend the correct spelling of vulgar terms. Most curiously though, this happened shortly after the investment bank *Goldman Sachs* complained about being autocorrected to *Goddam Sachs*.

Needless to say, Microsoft Word's custom-made dictionary was heavily twisted. The way it was acquired, compiled and edited clearly reflected the lingua franca of the company's engineers and the politics of its managers. By design, autocorrect would normalize the company's version of a 'correct language' and press you as a user of Word 6.0 to write like the average employee of a prudish US software giant. And if you dared to fall out of line with its vision, the all-too-familiar red squiggly underline would notify you accordingly. The extent of such intrusions cannot be overstated, since any version of autocorrect over time "becomes a normative force, because they're built as systems for telling you what language should be," says Allison Parrish, an expert for computation and language.

Fat Imagination

While Microsoft was busy collecting precious words from their personnel, computational linguists over at IBM were working on a radically different idea that would change autocorrect forever. There, computers were taught not only to spell, but also to recognize letters in relation to one another. Statistical patterns of occurrence would allow correct spelling to be predicted, at least in theory. *I love you* is more likely to be correct than *I live you*, since the former combination of letters appears more often. The challenge of this method was, however, that it required enormous amounts of real-life text — training data — to be analysed. Yet the age of big data was nowhere near and early digitized texts had to suffice. From technical manuals and stock market reports to laser technology patents and half a million internal Enron emails (released after its fraud investigation), researchers grabbed as many words as they could find, heralding an age of ruthless data extraction in the name of 'correct' spelling.

Today, so-called crawlers are deployed by big companies to keep up with

ever-evolving language, adding to the historical sources. These data-hungry bots roam the internet in a pedantic fashion to collect enormous quantities of public text. The goal of this profitable endeavour is to determine which words most people use and their likely constellations. Our erroneous expressions may be thus corrected by probabilistically savvy replacements based on the simple formula of probable, plausible and popular words.

As a result, a wealth of words falls through the cracks of autocorrect. For instance, the forenames *Rashmi* and *Fatima* are flagged red and offered the suggestions *Sashimi* and *Fat Imagination* in some applications. At the same time, the Afro-American slang word *finna* (for 'getting ready to') is turned into *Finns* and the queer term *gender* prompts the suggestion *agenda*. Those expressions, like myriads of other vernacular terms and non-Western names, are not seen as probable choices. They are ruled out, because they don't appear often enough in autocorrect's training data. In other words, autocorrect may even contribute to the erasure of linguistic worlds and communities that are already underrepresented in public and virtual space. It's a vicious cycle in which a subtle yet structural violence comes with the seemingly innocent replacement of a single letter.

Ultimately, autocorrect's attempts to anticipate what we mean to write will always be limited, because the way humans use language is idiosyncratic and unpredictable. Even if autocorrect learns a bit about our personal way of writing over time, our situational quirks and linguistic singularities still overwhelm its statistical models. This is why, for years, autocorrect would change a joyful *yeahhhh* to *uranium*, and why the cheerful spellings *hiiii*, *woooow* and *heeeey* are still error-flagged today. With its defaults to an aggregate norm and its fetish for probability, autocorrect is not prepared to anticipate our improvised and improbable modes of expression. It's not designed to overlook those deliberate mishaps, and alerts us duly when we actually try to say something special.

Sashimi

In previous texting eras, various groups resisted the oppression of such crude, machine-made limitations. During the heyday of T9 (a predictive text technology for good-old button-press phones), US teenagers would replace the term *cool* with *book* in their real-life conversations. It was the suggestion that appeared as they attempted to write c-o-o-l. "You're book!" became a way of passing a compliment. Similarly, a dialect spoken by Asian teens in the early 2010s was based entirely on autocorrect suggestions. Some say it was a secret code that helped them to communicate safely, a world away from the linguistic rules their parents and the police agreed upon. Strategies like these might even be seen as forms of resistance, turning autocorrect against itself. In an almost Dadaist fashion, those groups rejected its suppressive nature and challenged its built-in power structures by playfully appropriating its nonsensical flaws.

Henceforth, however, resistance will get more difficult. With ever-growing amounts

of training data and better algorithms, autocorrect is becoming smoother. As a result, hilarious hiccups and obvious mistakes, ready to be criticized and turned against the machines, are less and less frequent. Instead of *fuck* being turned into *duck*, powerful features like 'auto-complete' and 'auto-predict' are on the rise. Those functions not only replace individual words eloquently, but also adjust phrases and rewrite whole sentences seemingly perfectly, although they are still based on biased training data and corporate censorship. For example, Google's autocorrect function doesn't want you to see offensive content, so it would ignore the misspelled *crine*, making it harder to search for phrases like *Donald Trump's crimes*. At the same time, Grammarly, a popular writing assistant, was criticized for gender-bias: Corrections of misspelled words such as *paked* or *earnings* would depend on the gender pronoun placed in front of them. Whereas "he is *parked* and collects his *earnings*", she is *naked* and collects her *earrings*". If gender prompts autocorrect to reproduce skewed worldviews like this, one has to wonder what kind of subtle but significant manipulations age, ethnicity or sexual orientation will trigger.

While companies report that their autocorrect systems are getting smarter by the day, their fundamental omissions, gaps and biases are clearly still in place. Their operational mechanics are traversed by asymmetrical power relations and dotted with blind spots for underrepresented communities and linguistic subjectivity. Tech companies seek to assure their critics that, with time and the superpowers of AI, they will fix the problem. Though actually they won't. No level of autocorrect, fed with the largest troves of data and run by the smartest algorithms, can "escape the fundamental slippages that occur when an infinitely complex world is simplified and sliced into categories," as AI researcher Kate Crawford points out.

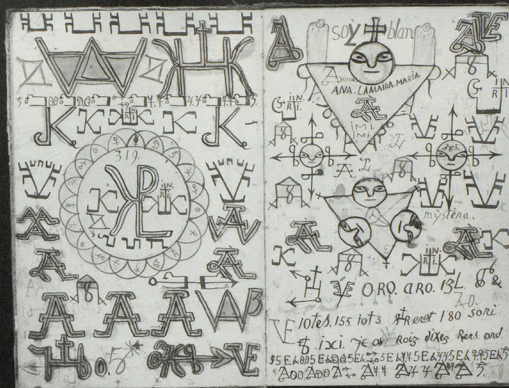
Ultimately, there is no neutral ground for correct language. Words are always deeply entrenched in time, place, culture and politics and their selection, editing and replacement is a serious socio-political act of world-making. Maybe it is time for auto-

She is naked and collects her earrings

He is parked and collects his earnings

correct's engineers to acknowledge this complexity and include all of us in its internal struggles. But not — as matters stand now — as anonymous data points that have to be squeezed into predefined modes of expression, but as unique individuals with the right to choose their own words.

73



REPAIRING THE WORLD

A posthumous letter from Ginevra Petrozzi to a kindred spirit, Pierre Richard, a 19th-century peasant from Lorraine, about the latter's magic books filled with drawings, monograms, alphabets and cabalistic writings.

Kédange, 1802. Pierre Richard Rome, 1997. Ginevra Petrozzi