archive.today
webpage capture

Saved from https://via.hypothes.is/https://www.wired.com/2014/07/history-of-autocorrect/ search
history ←prior next→
Redirected from https://via.hypothes.is/http://www.wired.com/2014/07/history-of-autocorrect/
history ←prior next→
Via https://www.wired.com/2014/07/history-of-autocorrect/
history ←prior next→
Original http://www.wired.com/2014/07/history-of-autocorrect/
history ←prior next→
All snapshots from host via.hypothes.is
from host www.wired.com

9 Mar 2019 01:59:07 UTC

Webpage Screenshot
share download .zip report bug or abuse Buy me a coffee

SHARE

GIDEON LEWIS-KRAUS GEAR 07.22.14 12:01 AM

THE FASINATNG... FASCINATING HISTORY OF AUTOCORRECT



David Sparshott

Invoke the word autocorrect and most people will think immediately of its hiccups—the sort of hysterical, impossible errors one finds collected on sites like Damn You Autocorrect. But despite the inadvertent hilarity, the real marvel of our mobile text-correction systems is how astoundingly good they are. It's not too much of an exaggeration to call autocorrect the overlooked underwriter of our era of mobile prolixity. Without it, we wouldn't be able to compose windy love letters from stadium bleachers, write novels on subway commutes, or dash off breakup texts while in line at the post office. Without it, we probably couldn't even have phones that look anything like the ingots we tickle—the whole notion of touchscreen typing, where our podgy physical fingers are expected to land with precision on tiny virtual keys, is viable only when we have some serious software to tidy up after us. Because we know autocorrect is there as brace and cushion, we're free to write with increased abandon, at times and in places where writing would otherwise be impossible. Thanks to autocorrect, the gap between whim and word is narrower than it's ever been, and our world is awash in easily rendered thought.

As someone who typed the entire first draft of his book on a phone, I want to shake the hand of the person responsible for this heedlessness, to meet the man who taught machines to draw sense from our maniacal tapping. I find him in a drably pastel conference room at Microsoft headquarters in Redmond, Washington. Dean Hachamovitch—inventor on the patent for autocorrect and the closest thing it has to an individual creator—reaches across the table to introduce himself.

“Dean,” I say.

“Gabriel,” he says.

“Gideon.”

“Sorry.” He frowns. Then he smiles and adds: “Autocorrect.” Apparently his brain, long trained for efficiency, replaces implausible terms with likelier alternatives.

Hachamovitch, now a vice president at Microsoft and head of data science for the entire corporation, is a likable and modest man. He freely concedes that he types teh as much as anyone. (Almost certainly he does not often type hte. As researchers have discovered, initial-letter transposition is a much rarer error.) The day I meet him he has dark circles under his eyes and short bangs that look self-cut, and his skin looks as though it has struggled valiantly to draw whatever vitamin D it can from his cinema display. He's clearly swamped, but autocorrect is close enough to his heart that he is happy to take the time to talk about it.



David Sparshott

When Hachamovitch first joined Microsoft, he was given a job on the Word team. This was back in the early '90s. Word processing was at a crossroads, split into factions. On one side were the people who wanted adornments and frills—improved desktop publishing, color separation, and the like. On the other side was the functionality gang, with whom Hachamovitch threw in his lot. This camp simply wanted to help people get out of their own way. As Hachamovitch saw it, the main thing that people do on a word processor is type—and typing, in his estimation, is a matter of “a little bit of creativity and a whole lot of scutwork.” He could improve the typing experience by delivering us from scut. His aim was to make our typing sleek and invisible, smooth as speaking from a teleprompter.

The notion of autocorrect was born when Hachamovitch began thinking about a functionality that already existed in Word. Thanks to Charles Simonyi, the longtime Microsoft executive widely recognized as the father of graphical word processing, Word had a “glossary” that could be used as a sort of auto-expander. You could set up a string of words—like insert logo—which, when typed and followed by a press of the F3 button, would get replaced by a JPEG of your company's logo. Hachamovitch realized that this glossary could be used far more aggressively to correct common mistakes. He drew up a little code that would allow you to press the left arrow and F3 at any time and immediately replace teh with the. His aha moment came when he realized that, because English words are space-delimited, the space bar itself could trigger the replacement, to make correction … automatic! Hachamovitch drew up a list of common errors, and over the next years he and his team went on to solve many of the thorniest. Seperate would automatically change to separate. Accidental cap locks would adjust immediately (making dEAR grEG into Dear Greg). One Microsoft manager dubbed them the Department of Stupid PC Tricks.

MICROSOFT WORD COULDN'T VERY WELL GO AROUND RECOMMENDING THE CORRECT SPELLING OF MOTHREFUKCER.

It wasn't long before the team realized that autocorrect could also be used toward less productive—but more delightful—ends. One day Hachamovitch went into his boss's machine and changed the autocorrect dictionary so that any time he typed Dean it was automatically changed to the name of his coworker Mike, and vice versa. (His boss kept both his computer and office locked after that.) Children were even quicker to grasp the comedic ramifications of the new tool. After Hachamovitch went to speak to his daughter's third-grade class, he got emails from parents that read along the lines of "Thank you for coming to talk to my daughter's class, but whenever I try to type her name I find it automatically transforms itself into 'The pretty princess.'"

Hachamovitch and his team couldn't get carried away with the pranks, though, because they were still hammering out the feature's basic functionality. The two-letter-caps substitution (THis to This) was scheduled to go live in Word 6, but it immediately posed the problem of how to handle exceptions (like CDs). What they needed, they realized—not just for this type of exception but for all of them—was a master list, a kind of artisanal concordance. The task fell to Hachamovitch's intern, Christopher Thorpe, a 19-year-old on leave from Harvard. Thorpe wrote a script that compiled all the manual entries that Microsoft employees had made to their custom dictionaries—words the built-in dictionary hadn't recognized but that were judged by users to be legitimate. Thorpe gave the list a quick edit and then gathered the remaining entries into a corpus. It began with such words as abuzz and acidhead. That prospectus became a database of the additional words that Word should find legit. In later Word releases, another list itemized changes that ought to be made to problematic homophone phrases. Word should always, for example, change elude to to allude to, and could of been to could have been.

For our meeting, Hachamovitch dug up a stack of historical documents like that list, files that had languished for 20 years in office moves, making their way eventually to his basement. He clearly takes a marmish pride in the artifacts, patting a shrink-wrapped Word 95 box and turning it over to point out where it touted the great triumph of autocorrect. He flipped open a perfect-bound manuscript to a page he'd marked, with a table that showed when they added judgement as a correctible error. That one, he explained, was a very common misspelling, now eradicated like smallpox.

On the subject of judgment, though, it became clear even in those early days that a sort of editorial consciousness was at work in Word's spell-check and autocorrect systems. Judgement, for example, isn't a misspelling—just about every dictionary lists it as an acceptable alternative. But autocorrect tends to enforce primary spellings in all circumstances. On idiom, some of its calls seemed fairly clear-cut: gorilla warfare became guerrilla warfare, for example, even though a wildlife biologist might find that an inconvenient assumption. But some of the calls were quite tricky, and one of the trickiest involved the issue of obscenity. On one hand, Word didn't want to seem priggish; on the other, it couldn't very well go around recommending the correct spelling of mothrefukcer. Microsoft was sensitive to these issues. The solution lay in expanding one of spell-check's most special lists, bearing the understated title: "Words which should neither be flagged nor suggested."





David Sparshott

I called up Thorpe, who now runs a Boston-based startup called Philo, to ask him how the idea for the list came about. An inspiration, as he recalls it, was a certain Microsoft user named Bill Vignola.

One day Vignola sent Bill Gates an email. (Thorpe couldn't recall who Bill Vignola was or what he did.) Whenever Bill Vignola typed his own name in MS Word, the email to Gates explained, it was automatically changed to Bill Vaginal. Presumably Vignola caught this sometimes, but not always, and no doubt this serious man was sad to come across like a character in a Thomas Pynchon novel. His email made it down the chain of command to Thorpe. And Bill Vaginal wasn't the only complainant: As Thorpe recalls, Goldman Sachs was mad that Word was always turning it into Goddamn Sachs.

Thorpe went through the dictionary and took out all the words marked as “vulgar.” Then he threw in a few anatomical terms for good measure. The resulting list ran to hundreds of entries:

anally, asshole, battle-axe, battleaxe, bimbo, booger, boogers, butthead, Butthead ...

With these sorts of master lists in place—the corrections, the exceptions, and the to-be-primly-ignored—the joists of autocorrect, then still a subdomain of spell-check, were in place for the early releases of Word. Microsoft's dominance at the time ensured that autocorrect became globally ubiquitous, along with some of its idiosyncrasies. By the early 2000s, European bureaucrats would begin to notice what came to be called the Cupertino effect, whereby the word cooperation (bizarrely included only in hyphenated form in the standard Word dictionary) would be marked wrong, with a suggested change to Cupertino. There are thus many instances where one parliamentary back-bencher or another longs for increased Cupertino between nations. Since then, linguists have adopted the word cupertino as a term of art for such trapdoors that have been assimilated into the language.

In the two decades since Hachamovitch moved from the manual coding of corrections like judgement to his loftier executive role in the ambit of data science, autocorrect has followed suit. Autocorrection is no longer an overqualified intern drawing up lists of directives; it's now a vast statistical affair in which petabytes of public words are examined to decide when a usage is popular enough to become a probabilistically savvy replacement. The work of the autocorrect team has been made algorithmic and outsourced to the cloud.

A handful of factors are taken into account to weight the variables: keyboard proximity, phonetic similarity, linguistic context. But it's essentially a big popularity contest. A Microsoft engineer showed me a slide where somebody was trying to search for the long-named Austrian action star who became governor of California. Schwarzenegger, he explained, “is about 10,000 times more popular in the world than its variants”—Shwaranegar or Scuzzynectar or what have you. Autocorrect has become an index of the most popular way to spell and order certain words.

When English spelling was first standardized, it was by the effective fiat of those who controlled the communicative means of production. Dictionaries and usage guides have always represented compromises between top-down prescriptivists—those who believe language ought to be used a certain way—and bottom-up descriptivists—those who believe, instead, that there's no ought about it.

The emerging consensus on usage will be a matter of statistical arbitration, between the way “most” people spell something and the way “some” people do. If it proceeds as it has, it's likely to be a winner-take-all affair, as alternatives drop out. (Though Apple's recent introduction of personalized, “contextual” autocorrect—which can distinguish between the language you use with your friends and the language you use with your boss—might complicate that process of standardization and allow us the favor of our characteristic errors.)

There is, of course, some legacy prudishness to autocorrect—the tendency, for example, of hell to become he'll—but for the most part the global menagerie has, in its unflagging vulgarity, produced a linguistic corpus that skews blue. Where Hachamovitch did away with the scutwork, the new autocorrect introduces the slutwork and the smutwork. When one reads such hilarious-error collections as Damn You Autocorrect, one can't help but feel skeptical. Some entries beggar the imagination; it's hard to believe that Volvos could become vulvas as often as they seem to. But even if we assume a significant rate of fraud, we are forced to conclude—given that autocorrect draws from group behavior—that the unpublished typing of our society is more unpublishable than we ever imagined.





David Sparshott

In an earlier generation, when players in the party game of charades needed to indicate the idea of "work," they might have mimed the laying of brick, the digging of a trench. Today, given what passes for work in the smartphone era, players would be much likelier to fiddle their thumbs in the air—possibly just one of them. If a peculiarly hyperspecific virus suddenly laid to waste the world's right thumbs, world GDP would tumble. And none of this digital (in both senses) productivity would be possible without autocorrect.

Given how successful autocorrect is, how indispensable it has become, why do we stay so fixated on the errors? It's not just because they represent unsolicited intrusions of nonsense into our glassy corporate memoranda. It goes beyond that. The possibility of linguistic communication is grounded in the fact of what some philosophers of language have called the principle of charity: The first step in a successful interpretation of an utterance is the belief that it somehow accords with the universe as we understand it. This means that we have a propensity to take a sort of ownership over even our errors, hoping for the possibility of meaning in even the most perverse string of letters. We feel honored to have a companion like autocorrect who trusts that, despite surface clumsiness or nonsense, inside us always smiles an articulate truth.

Just now, for example, I reached for my phone and bashed my finger pads against the glass to see what wisdom autocorrect might read from me today. I started in the general vicinity of the letter d and then just let loose, trying to tap at random across the characters. The first time I tapped out dcisnence and drew existence. The random string dzyjzynxe produced distance. The third time I went a little longer and beset my keyboard with descinnztsb. This instantly transformed itself into deacon stab. And there it was, a little potted history of humanity: first birth, then exile, and before you know it somebody's gone and shanked a priest.

We write something and immediately take responsibility for it; we see something in the world and, as charitable interpreters, want to believe that it contains meaning. Following the invention of Samuel Morse's new code, spiritualists claimed that a series of knocks might be understood as a communication from the dead. Likewise, it often feels as though our autocorrect errors make us into media—not for the dead but for ghosts in our language, summoned forth by Big Data. Ultimately we come back around, via probabilistic inquiry, to the old lit-crit idea of the death of the author—the proposition that we don't speak through language but that language speaks through us. We are merely its clumsy-thumbed vessels.

After all, it's almost impossible not to read some hidden intentionality behind these errors, since they derive from a probabilistic analysis of our aggregate behaviors. It's almost too good that the foundational autocorrect mistake, the one that gave the whole genre its name, was the unsolicited

appearance of Cupertino, home of Apple's headquarters—since in the past seven years it's been Cupertino, not Redmond, that has elevated autocorrect to its role as a kind of necessary arch-demon.

Today the influence of autocorrect is everywhere: A commenter on the Language Log blog recently mentioned hearing of an entire dialect in Asia based on phone cupertinos, where teens used the first suggestion from autocomplete instead of their chosen word, thus creating a slang that others couldn't decode. (It's similar to the Anglophone teenagers who, in a previous texting era, claimed to have replaced the term of approval cool with that of book because of happenstance T9 input priority.) Surrealists once encouraged the practice of écriture automatique, or automatic writing, in order to reveal the peculiar longings of the unconscious. The crackpot suggestions of autocorrect have become our own form of automatic writing—but what they reveal are the peculiar statistics of a world id.

I'm almost relieved to report that none of these thoughts trouble the superegos of the Microsoft data scientists who continue to work at improving autocorrect. In my second meeting there, with the current head of the European autocomplete team for Bing, Nitin Agrawal, I can only pay attention for so long as he drones on about the statistical probability of a Schwarzenegger. My mind begins to drift. Finally I turn to him.

“Do you text in any languages beside English?” I ask.

He pauses. “I use only English.” He pauses again. “But occasionally I might use Hindi.”

“Are there funny autocorrect errors in Hindi?”

He doesn't seem to know of any. I do my best to explain. I mean, like, vulvas and deacon stabs.

“I don't have a ready answer for you to that,” he says. “I think we're going to have to take that question offline.”

I look around at the expanse of conference table, lit with bounced whiteboard glare. The only place we could get more offline would be under the ocean.

He begins to speak again. “We have errors, yes. But what percent are amusing? We don't have a metric around amusement. We don't track amusement errors as a metric.” His metrics, he explained, were such things as “time to task completion.”

When I later transcribe my notes on my phone, I find, to my great delight, that his metric has become time to tusk contraction. You may not go looking for amusement, but when you allow autocorrect to speak on your behalf, amusement finds you regardless.

#ARTPHONE #MAGAZINE-22.08 #WIRED CLASSIC

VIEW COMMENTS

SPONSORED CONTENT POWERED BY OUTBRAIN



JEFFREY VAN CAMP

LG's OLED TVs Are Cheaper Than We've Ever Seen Them

SPONSORED



ARIELLE PARDES

The Perfect Pair of Pants Is Just a 3D Body Scan Away

SPONSORED



JEFFREY VAN CAMP

Best Nintendo Switch Deals, Bundles, and Essentials Online

SPONSORED



LAUREN GOODE

Microsoft's HoloLens 2 Puts a Full-Fledged Computer on Your Face

SPONSORED



WIRED STAFF

You've Got to Know When to Fold 'Em

SPONSORED

GADGET LAB PODCAST

A Former Google Exec Tells You How to Quit Your Tech Job

WIRED STAFF

EULOGY

My Jibo Is Dying and It's Breaking My Heart

JEFFREY VAN CAMP

BUYING GUIDE

Last Call: How to Pick the Right Galaxy S10, and Preorder It

JEFFREY VAN CAMP

BUYING GUIDE

Kick the Keyboard: Other Ways to Get Text Onto Your Screen

RICHARD BAGULEY

GEAR

Fitbit's New Lower-Cost Wearables Undercut the Apple Watch

LAUREN GOODE

HOW TO

A BEGINNER'S GUIDE TO

LOUISE MATSAKIS

GET GADGET LAB'S NEWSLETTER

Sign up to receive the latest gadget news, reviews, and expert advice.

Enter your email

SUBMIT

CNMN Collection

© 2018 Condé Nast. All rights reserved.

Use of and/or registration on any portion of this site constitutes acceptance of our User Agreement (updated 5/25/18) and Privacy Policy and Cookie Statement (updated 5/25/18). Your California Privacy Rights. The material on this site may not be reproduced, distributed, transmitted, cached or otherwise used, except with the prior written permission of Condé Nast. Ad Choices.