

Chapter 1

Informal Writing

Imagine learning to talk from recordings rather than people. If you learned how to have a conversation from movies, you might think that people regularly hang up the phone without saying goodbye and no one ever interrupts anyone else. If you learned to think out loud from news programs, you might believe that no one ever “ums” or waves their hands while searching for an idea, and that people swear rarely and never before ten p.m. If you learned to tell stories from audiobooks, you might think that nothing much new had happened with the English language in the past couple hundred years. If you only ever talked when you were public speaking, you’d expect that talking always involves anxious butterflies in your stomach and hours of preparation before facing an audience.

Of course, you did none of these things. You learned to speak English domestically, conversationally, and informally long before you could sit through an entire news report or deliver a speech. You might never be wholly comfortable with public speaking, but of

course you can complain about the weather to a friend. Sure, they both involve moving the same body parts, but they're hardly the same task at all.

And yet this is exactly how we all learned to read and write.

When we think about writing, we think about books and newspapers, magazines and academic articles—and the school essays in which we tried (and mostly failed) to emulate them. We learned to read a formal kind of language which pretends that the past century or two of English hasn't really happened, which presents words and books to us cut off from the living people who created them, which downplays the alchemy of two people tossing thoughts back and forth in perfect balance. We learned to write with a paralyzing fear of red ink and were taught to worry about form before we even got to consider what we wanted to say, as if good writing were a thing of mechanistic rule-picking rather than of grace and verve. Naturally, we're as intimidated by the blank page as we are by public speaking.

That is, we were until very recently. The internet and mobile devices have brought us an explosion of writing by normal people. Writing has become a vital, conversational part of our ordinary lives. In the year 800, Charlemagne managed to get himself crowned as Holy Roman Emperor without being able to sign his own name. Sure, he had scribes to write up his charters, but illiterately running an empire? Today it's hard to imagine even organizing a birthday party without writing. One type of writing hasn't replaced the other: the "Happy Birthday" text message hasn't killed the diplomatic treaty. What's changed is that writing now comes in both formal and informal versions, just as speaking has for so long.

We write all the time now, and most of what we're writing is informal: our texts and chats and posts are quick, they're conversational, they're untouched by the hands of an editor. If you define a

"published" writer as someone who's had something they've written reach over a hundred people, practically everyone who uses social media qualifies—just make an announcement post about a new job or a new baby. It's not that edited, formal writing has disappeared online (there are plenty of business and news sites that still write much like we did in print), it's that it's now surrounded by a vast sea of unedited, unfiltered words that once might have only been spoken.

I'm a linguist, and I live on the internet. When I see the boundless creativity of internet language flowing past me online, I can't help but want to understand how it works. Why did emoji become so popular so quickly? What's the deal with how people of different ages punctuate their emails and text messages so differently? Why does the language in memes often look so wonderfully strange?

I'm not alone in wondering about these things. When I started writing about internet linguistics online, I quickly ran into more follow-up questions from readers than just another article could answer. I went to conferences and dived into research papers and ran a few of my own queries. I realized that in many cases there were answers, just not from an internet native speaker, not all together in one place, not in a form that's fun to read regardless of how much you already know about linguistics. So I wrote this book.

Linguists are interested in the subconscious patterns behind the language we produce every day. But traditionally, linguistics doesn't analyze writing very much, unless it's a question about the history of a language and written records are all we have. The problem is that writing is too premeditated, too likely to have gotten filtered through multiple hands, too hard to attribute to a single person's linguistic intuitions at a specific moment. But internet writing is different. It's unedited, it's unfiltered, and it's so beautifully mundane.

And, as I've continued rediscovering with every chapter of this book, when we analyze the hidden patterns of written internet language, we can understand more about our language in general.

Internet writing is also useful because speech is an absolute nightmare to analyze. First of all, speech vanishes as soon as it's said, and if you're just taking notes, you might be misremembering things or not noticing everything. So you want to record the audio, but that's your second problem: now you need to physically transport people into a recording lab or travel around with a recorder. Once you've got recordings, you've got a third problem: processing. It takes about an hour of skilled human work per minute of audio recording to get speech into a transcript usable for linguistic analysis: to transcribe the overall gist, to go back and add detailed phonetic information, to extract parts and analyze their acoustic frequencies or sentence structure. Many a beleaguered linguistics grad student has spent years of their life doing precisely this, in search of the answers to just a handful of specific questions. It's hard to do at a massive scale. All the while, there's a fourth challenge: your participants probably won't talk to an academic interviewer the same way they'd talk to a friend. Want to analyze a signed language instead? Instead of analyzing audio in just one dimension, now you're facing video in two. Want to skip a step and use preexisting recordings? Good luck: most of that is news, acting, and other formal varieties.

There were difficulties in studying informal writing before the internet, too. It existed, in forms like letters, diaries, and postcards, but by the time a collection of papers is donated to an archive, they've generally been moldering in boxes for decades, and of course they also need to be processed in order to be analyzed. Deciphering old-timey handwriting on fragile paper is only marginally easier than

transcribing audio. Studies of Victorian letters and medieval manuscripts can tell us that a particular word is older than we thought, or provide evidence of changing pronunciations through idiosyncratic spelling, but we don't want to limit our studies of present-day English to a fifty-year time delay, based solely on the highly biased sample of the kinds of famous people whose papers get donated to archives. But if we wanted more recent stuff, we'd again face the logistical challenges of getting people to write, for instance, sample postcards for our study and hoping that they're not too self-conscious about researchers reading their words.

Lucky for us, internet language is both easier to work with, since the text is already digital, and less likely to get distorted because someone's observing it, since much of it is already public as tweets and blogs and videos. (Although the would-be internet researcher must also consider the ethics of working with linguistic data that is functionally public but would embarrass or harm the people that made it if distributed out of context.) Even the logistics of distributing fun language surveys or asking people to donate archives of their private messages has gotten easier online. Internet linguistics isn't just a study of the latest cool memes (though we'll get to memes in a later chapter): it's a deeper look into day-to-day language than we've ever been able to see. It brings new insight to classic linguistic questions like, How do new words catch on? When did people start saying this? Where do people say that?

Now, I like me a good book. I've watched a few TED Talks in my time. I'm very aware of the hours of craftwork that go into making ideas flow gracefully through formal language, and there's much to be admired there. But there's already plenty of admiration for literature and oratory. As a linguist, what compels me are the parts

of language that we don't even know we're so good at, the patterns that emerge spontaneously, when we aren't really thinking about them.

Even keysmash, that haphazard mashing of fingers against keyboard to signal a feeling so intense that you can't possibly type real words, has patterns. A typical keysmash might look like "asdjlklgafdljk" or "asdlkfjas;dlf"—quite distinct from, say, a cat walking across the keyboard, which might look like "tfgggggggggggggggggsgsdxzzzzzzzzz."*

Here's a few patterns we can observe in keysmash:

- Almost always begins with "a"
- Often begins with "asdf"
- Other common subsequent characters are g, h, j, k, l, and ;, but less often in that order, and often alternating or repeating within this second group
- Frequently occurring characters are the "home row" of keys that the fingers are on in rest position, suggesting that key-smashers are also touch typists
- If any characters appear beyond the middle row, top-row characters (qwe . . .) are more common than bottom-row characters (zxc . . .)
- Generally either all lowercase or all caps, and rarely contains numbers

Sure, a lot of these patterns relate to the fact that we're mashing on the home row of the QWERTY keyboard rather than using random-letter generators, but they're reinforced by our social expectations. I conducted an informal survey, asking if people retype their keysmash if it doesn't look, er, smashing enough. While there were a





*With thanks to Eliza, the Very Good Cat of A.E. Prevost.

few keysmash purists, who posted whatever came out, I found that the majority of people will delete and remash if they don't like what it looks like, plus a significant minority who will adjust a few letters. I also heard from several people who use the Dvorak keyboard, where the home row begins with vowels rather than ASDF, who reported that they just don't bother keysmashing anymore at all because their layout makes it socially illegible. Keysmashing may be shifting, though: I've noticed a second kind, which looks more like "gbghvjfbtghchc" than "asafjlskflskf," from thumbs mashing against the middle of a smartphone keyboard.

It's not just that we make patterns. It's that even when we're not trying to make patterns, when we think we're just a billion monkeys mashing incoherently on a billion keyboards, we're social monkeys—we can't help but notice each other and respond to each other. Even when something looks incoherent to an outsider, even when it's intended as incoherent for an insider, we as humans are still practically incapable of doing things *without* patterns. My mission with this book is to map out what some of those patterns are, to examine why they fall into the patterns that they do, and to give you the tools to look at internet language and other cutting-edge linguistic innovation through the lens of a pattern-seeker.

As with any period of tremendous disruption, the explosion of informal writing is changing the way we communicate. The norms that we worked out for books and newspapers don't work so well for texts and chats and posts. Imagine how weird you'd think ordinary conversation was if you'd only ever seen scripted TV monologues! We have a sense, more or less, of how informal speech works. We have a long history of doing it, and it's the primary thing that linguistics studies, much as literature and rhetoric study formal writing and formal

speaking. But the combination of writing and informality has been neglected—and this quadrant is precisely where internet writing excels. How does it fit in among these known quantities?

	SPOKEN	WRITTEN
INFORMAL	conversations, talking to yourself 	texts, chat, social media, diaries, notes 
FORMAL	speeches, radio, television, acting 	books, articles, static websites 

One way to think about informal writing is through the lens of efficiency. Across languages, short words tend to be more common words, which contribute a small amount of information to a sentence, while longer words occur less frequently and contribute more information. Think about the English words “of” and “rhinoceros.” “Of” is clearly more common, and it’s also much shorter—a simple vowel + consonant sequence that can even be reduced into a single neutral vowel, as in “sorta” or “outta.” “Rhinoceros” is longer and way more informative: if you hear “rhinoceros!” out of the blue, you can form a pretty solid hypothesis about what’s going on, and if it’s accidentally omitted (“I am fond of this _____”), many other words could take its place. Hearing “of!” out of the blue is pretty much meaningless, and if it’s accidentally omitted (“I am fond ___ this rhinoceros”), you can be almost certain that it was meant to be there. It would be a waste to use the short, versatile monosyllable “of” for the relatively uncommon concept of an odd-toed ungulate. Similarly, if we assigned the meaning of “of” to a sequence of sounds as long as “rhinoceros,” it would be a clear drop in efficiency. In this chapter alone, the word “of” occurs over one hundred times, and

making them all five times longer would be a lot rhinoceros sound for a small amount rhinoceros meaning!

Frequency isn’t completely static: the word “rhinoceros” entered English around the fourteenth century, but as the animal became more common in the lives of English speakers, we shortened it to “rhino” by 1884. “Rhino” splits the difference. It’s not quite as short as “of,” but then again, even a zookeeper still says “of” more often. Truly obscure animals, like the axolotl (a type of salamander) or the *Wunderpus photogenicus* (a type of octopus which, true to its name is very photogenic), don’t have nicknames in common use, although I expect to hear from the Association for Researchers of the Axolotl and the *Wunderpus Photogenicus* (ARAWP?) any day now informing me that they say them often enough that they’ve devised more efficient names for them.

Sometimes, as with “of” and “rhinoceros,” efficiency in writing, and speaking amounts to basically the same thing: more letters on the page equals more sounds in the mouth. Other times, they take different paths. In speech, we often make language more efficient by dropping unnecessary syllables or squishing sounds together, even if it’s not writable. We truncate words without regard for spelling: you can say the first syllable of “usual” or “casual” and everyone knows what you mean, but do you write it “yooj”? “uzh”? “cazh”? “casj”? It’s simply not clear, but speech proceeds merrily along anyway. An even more extreme example comes in how English speakers smooth out “I do not know.” We’ve been saying it out loud for generations, long enough for it to have worn down to “I don’t know,” “I dunno,” and even a simple triplet “uh-huh-uh” or “mm-hm-mm” to the low-high-low melody of “I dunno.” “I dunno” is easier to articulate than “do not know,” but it’s not really much shorter to write (even though we sometimes write it to evoke speech). The melodic triple hum

is exceedingly easy to produce (you can even do it with a mouthful of sandwich) but not efficient at all in writing, requiring a full-on explanation. We also try to maintain a constant rate of information flow: to say predictable words more quickly and unpredictable words more slowly. One study showed that people say the word "mind" quite quickly in a sentence like "Mama, you've been on my mind," where it's very predictable thanks to a certain off-covered Bob Dylan song, but they say it much slower in an unpredictable context, like "paid jobs degrade the mind," one of Aristotle's more obscure sayings. (Of course, if you're a big Aristotle fan who's never heard of Bob Dylan, you may find that the inverse is true for you.)

In writing, we often make language more efficient by selecting just a few important letters or squishing symbols together into a new shape, even if the resulting formations aren't pronounceable. The types of ideas that get acronyms or abbreviations have evolved along with what sorts of things people wanted to write efficiently. The Romans found it much easier to inscribe coins and statues with SPQR than the full *Senatus Populusque Romanus*. Medieval scribes smooshed common words together into new symbols such as & and %. When the Renaissance brought with it an increased interest in the classics and sciences, scholars abbreviated Latin phrases like *eg.* ("for example") and *ibid.* ("in the same reference already cited"). But the true golden era of acronyms began surprisingly recently. The word "acronym" itself entered English only in 1940, and acronyms, especially the kind that are pronounced as a single word rather than a series of letters, began flourishing during World War II, when soldiers used acronyms like AWOL, snafu, WAAF, and radar.* After the war, acronyms just kept proliferating, especially for organiza-

* Absent Without Official Leave, situation normal all fucked up, Women's Auxiliary Air Force, and Radio Detection And Ranging.

tions, new discoveries, and other names: laser, NASA, NATO, AIDS, NAACP, codec, Eniac, UNESCO, UNICEF, OPEC, FIFA, NASDAQ, FDR, CD-ROM, MoMA, DNA, and so on. These forms are shorter in writing, but not necessarily more efficient in speech, even though we sometimes speak them aloud when we're talking about specialized topics: it takes longer to say "ampersand" or "WWF" than "and" or "World Wildlife Fund." Technical acronyms reflect writing as a formal domain, which aims to maximize efficiency on bureaucratic procedures and long-winded names.

The internet has acronym'd some technical terms as well, like url, jpeg, or html, but a lot of what we're writing is informal and conversational. A new kind of social acronym has come into use, based on common conversational phrases rather than technical jargon—less BAC (blood-alcohol content) and more btw (by the way), less OBE (Order of the British Empire) and more omg (oh my god), less LAX (the airport code for Los Angeles) and more lol (originally "laughing out loud," though now a more subtle meaning, which we'll get to in Chapter 3). I think it's disingenuous to follow formal tradition at the expense of regular usage in a book that's entirely about regular usage, so I've made the stylistic decision to write social, internet acronyms in all-lowercase, while often keeping technical acronyms in uppercase, because people on the internet primarily reserve LOL and OMG for when they're SHOUTING.

Internet acronyms are a perfect example of the intersection between writing and informality. Their form comes from the writing side: acronyms reduce the number of letters you type, although not necessarily the number of syllables you articulate. In other words, "I dunno" is efficient in speech, whereas "idk" is efficient in writing. Their function comes from the informal side: the phrases are personal expressions of our feelings and beliefs, like "I don't know," "what the fuck," "just so you know," "as far as I know," "in my

opinion," "today I learned," and "that feeling when." With technical acronyms, the long version and the short version are invented at the same time, sometimes with an eye to how the initial letters of a phrase will fit together as an acronym. Social acronyms are instead made out of phrases that are already common: a sure sign of bad internet linguistics is EIAFUP (Elaborate Invented Acronyms for Uncommon Phrases). We're not pure efficiency maximizers, however; we also sometimes respell words when we want to make speech evoke speech, or speak acronyms when we want to make speech evoke writing. Efficiency simply points to where and why a particular abbreviation originated.

Another example of how the internet melds writing and informality is in how we use visuals. In casual speech, we don't generally converse in pitch-blackness, with our backs to each other, or with our hands tied behind our backs, a paper bag over our heads, and our voices in a robotic monotone. I mean, you could get a message across that way, but you'd be missing something. Often, we gesture: the next time you're in a public place, look around at groups of people and notice how you can tell who's talking by looking at who's gesturing. We use gesture and tone of voice to reinforce our message or add another layer of meaning to it: "Good job!" plus a thumbs up is sincere, while with rolled eyes it's sarcastic.

Formal speech uses fewer and more stylized gestures. Television reporters may point at a weather map or shuffle a few pieces of paper, but for the most part, their hands remain still and their faces remain in a narrow range between serious and cheerful—no waving, eye-rolling, sobbing, or uproarious laughter. Public speakers are often advised to cut down on their natural movements: a classic tip for improving your public speaking is to watch a video of your per-

formance so you can notice and reduce repetitive gestures. Both styles are often cropped just below the shoulder, by a camera or a lecturer, so we can't see the speaker's hands even if they did move them. People have been prescribing formal gestures for a long time: the Roman orator Quintilian advised rhetoricians that acceptable gestures included pointing and gestures of admiration, wonder, rejection, certainty, pleading, and others which "naturally proceed from us simultaneously with our words," but that pantomime and similar attempts to act out one's actual words should be reserved for the theater rather than the law courts.

Writing is a technology. Speaking and signing require only our human bodies and the energy we infuse them with, and we've never met a society without one or both. But writing requires something external to the body: even if you write on your own arm with your own blood, you'll need to prick yourself to do so, thereby making the blood external. Writing systems, therefore, are greatly affected by the tools available to make them: it's easier to carve wood or stone in a straight line, but easier to swirl and loop with ink. The images that go along with our writing also reflect the available tools. Medieval illumination was on calfskin vellum with pigments made of ground beetles or stones; print drawings were cut into wood so the raised areas could be inked and pressed onto paper; cameras fixed a tiny pinhole of light onto a ribbon of film with silver compounds. In any circumstance, some colors and shapes were easier than others.

Formal writing is disembodied in the same way that formal speech is. Just like a news anchor is supposed to be a conduit for the news, not its maker, the images of formal writing represent the content, not the author. Sometimes formal images present information, like graphs and diagrams. Sometimes they create a record, like maps and portraits. Sometimes they're eye-catching and tell stories, like in stained glass and picture books.

What's cool about informal writing is that, once we had the technology to send any image anywhere, we used it to restore our bodies to our writing, to give a sense of who's talking and what mood we're in when we're saying things. Take emoji, those small images that enliven our digital messages. There are thousands of them, ranging from animals to foods to nature to common household and workplace objects. And yet the most commonly used sets of emoji are the faces and hands, like the smile, the face with tears of joy, the thumbs up, and the crossed fingers. We use emoji less to describe the world around us, and more to be fully ourselves in an online world.

We do the same with gifs. In theory, you can put any image in an animated eight-frame loop. (In the 1990s, people used grainy, animated "under construction" gifs, ornamented with hard hats and traffic cones, to apologize for portions of their homepages that were incomplete.) In practice, the gifs of the 2010s are reaction gifs: silent, looping animations of people, animals, or cartoon characters doing a specific, expressive gesture which you invoke as a representation of your own body under the circumstances, saying that right now you're laughing, or applauding, or looking around in bewilderment, just like the person in the gifis. Early visualizations of cyberspace thought we might want to manipulate three-dimensional figurines of ourselves in order to interact with each other in virtual space. But it turned out, what we really wanted was less about dressing up our avatars in fanciful digital clothing and more about conveying what we're thinking and feeling. We can recruit a wide range of tools, from acronyms to emoji to punctuation, in order to do so.

The first writing systems were deeply aware of their limitations. They wrote only words, a mere aid to the memory of the reader, who had to infuse them with life again on saying them. Gradually, over the centuries, we began adding punctuation and other typographical enhancements. Just as crucially, we began expecting more subtlety

from the written text: we began to see writing as a thing that could represent verbatim speech and stream of consciousness, even if most of us were reading it rather than writing it. The internet was the final key in this process that had begun with medieval scribes and modernist poets—it made us all writers as well as readers. We no longer accept that writing must be lifeless, that it can only convey our tone of voice roughly and imprecisely, or that nuanced writing is the exclusive domain of professionals. We're creating new rules for typographical tone of voice. Not the kind of rules that are imposed from on high, but the kind of rules that emerge from the collective practice of a couple billion social monkeys—rules that enliven our social interactions.

Whatever else is changing for good or for bad in the world, the continued evolution of language is neither the solution to all our problems nor the cause of them. It simply is. You never truly step into the same English twice. When future historians look back on this era, they'll find our changes just as fascinating as we now find innovative words from Shakespeare or Latin or Norman French. So let's adopt the perspective of these future historians now, and explore the revolutionary period in linguistic history that we're living through from a place of excitement and curiosity.

If you're worried that this revolution is leaving you behind, or if you're so cutting-edge that it's hard to explain yourself to non-internet people, this book will help you bridge that gap. It will help you understand how we got here, why people get so passionate about a stray period or three, and how linguistic changes on the internet fit into the broader picture of the incredible capacity of human language. You'll never look at your quickly dashed-off text messages the same way again.



You've lived somewhere—perhaps in a city where your friends or family have influenced you. You've probably even thought about how you like some linguistic features and want to avoid others. People have long been aware that these factors influence why different groups speak differently, but the systematic study of dialect began in the eighteenth and nineteenth centuries, as part of the same scientific movements that gave us the Linnaean catalogue of the living world and the periodic table of the elements. While some cataloguers set out with nets to study butterflies, or burned candles inside jars to distill gases, others pored over ancient scripts and compiled lists of verbs.