

The Dictionary, the Definer, and How the Internet Has Changed Them Both

*Emily Brewster, Senior Editor, Merriam-Webster
Freeman Lecture, UMass Linguistics Nov. 17, 2021*

I've been a lexicographer—that is, a writer and editor of dictionaries—for more than 20 years now, and in that time I have been continually struck by the disconnect between the lexicographer's conception of what a dictionary's role is and the typical dictionary user's conception of that role. Lexicographers—and I am talking specifically about English-language lexicographers, and especially about ones who write general-use dictionaries—lexicographers tend to see the dictionary's role as being perhaps as simple as this: to tell the truth about words. That's how Philip Gove, editor-in-chief of Merriam-Webster's 1961 unabridged dictionary put it. For the English language lexicographer, there is no gatekeeping, no effort to keep the language contained, fixed. The aim is to tell the truth about words, based on the evidence available.

Truth is a multifaceted thing, though, and certainly there's agreement between lexicographers and people who use dictionaries: a dictionary provides facts about a word's spelling, its inflections, its function, its meaning, its etymology. But we also know that sometimes people are looking for something most dictionaries aren't designed to do: perennially, one of our most looked-up words is *_love_*. Lookups of *_love_* at M-W.com always spike around Valentine's Day, but they're high generally. We're pretty sure people aren't checking spelling or inflections; they're looking up *_love_* because they want to know what the band Foreigner wanted to know: They wanna know what love is. That is, they want a better understanding of what the phenomenon called “love” is, a kind of philosophical explication. They have, in fact, told us as much:

>Email received May 21, 2018

>You need to change the definition of love because what I feel for my girlfriend [name withheld] is far more than what the definition states.

>Email received October 15, 2015

>As a dictionary, can you please tell me the true meaning of love :3 Because i'm going crazy now. Help me please Lovelots <3 [emoticon comprising a less than symbol and 3]

Sometimes they want to offer helpful corrections:

Email received April 21, 2006

>A new definition of love

>Love—when two life forms use one another for leisure, trust, information, and/or work.

>

>(webster can add more if they wish)

My replies to queries such as these point the writer to the poets and philosophers: they're the ones who dig into the nature of love itself. It's just not the dictionary's job. In lexicography we talk about this as being a matter of real defining vs word defining. The investigation into the nature of phenomena?—that's called real defining. I admit the terminology is imperfect—I think it means that the band Foreigner does real defining and I don't?

Whatever. The term that is contrasted with “real defining” or “real definition” is itself apt: in lexicography our interest is the word itself. And so we call what we do “word defining” or “word definition.”

Within the category of word definition, there are two varieties. The first is called “stipulative definition”: this is the kind of defining that aims to reflect cold facts about words that have precise and largely static functions in the language. It doesn't matter what the layperson thinks angina pectoris is; the definer who handles medical terms will define it according to what the medical literature says it is.

At Merriam-Webster definers are categorized as either science definers or general definers: science definers handle vocabulary relating to math, chemistry, astronomy, biology, medicine, and the like, and they do primarily stipulative defining. A general definer, which is what I am, handles any non-science terms, and must also do plenty of stipulative defining. At love, for example, the tennis use of the word gets a stipulative definition: “a score of zero (as in tennis)” (the parenthetical makes definition broad enough to cover the use of love in other games). But the bulk of the entry requires what we call “lexical” defining: the writing of definitions that reflect actual usage.

While the entry for `_love_` is a big fail according to some of our correspondents, it does in fact reflect the varied ways that speakers of English use the word `_love_`; `_love_` can refer to many things—it can be strong affection with or without sexual desire; it can be a sexual episode, or sex itself; it can be benevolent concern for the good of others; it can be adoration of God; it can be great enthusiasm; or it can be appreciative attention or praise; and it can be a score of zero—among other things.

Lexical definitions require that the definer read a great many examples of the word in use. For a term like `_love_` one has to be sure that the resulting definition addresses current and historical uses of the word; religious and literary uses; formal and informal uses—which means consulting hundreds of sources. I'll get into how that's done a bit later on. But for now, we're going to move on from Real vs Word defining because the disconnect between how the lexicographer sees the dictionary's role and what dictionary users often want a dictionary to do doesn't stop there.

In addition to wanting a dictionary to do the poet's job, people who use dictionaries also frequently want dictionaries to pass judgment on the language's constituents.

There is an anecdote involving Samuel Johnson, creator of the first full dictionary of the English language, published in 1755. The oldest version of the anecdote has an unnamed woman approaching Dr. Johnson and complimenting him on his prudent omission of improper words, to which Johnson is said to have replied: "No, Madam, I have not daubed my fingers. I find, however, that you have been looking for them."

Johnson did in fact daub his fingers with the ink required to define improper words. He defined `_fart_` (prohibited in my childhood home) as "wind from behind," and defined `_bitch_` as "A name of reproach for a woman" in addition to its female canid[cannid] meaning. He also included the potent sobriquet `_pricklouse_`, a contemptuous term for a tailor.

But what Johnson found still holds true: sometimes people look up words they don't want to find. And when they find the word (or the meaning) they don't want to find, they sometimes let us know just how dissatisfied they are with us.

Email received March 2, 2016:

>I know you've heard it before, but as a life-long educator ..., I simply feel that you, and other dictionary entities, are absolutely wrong by allowing slang, non-words, misused words, such as "ain't" and "irregardless" to become words, or even entries in the dictionary. Why have rules if they aren't applied? I already know the answer...just voicing a view and casting my vote. Thanks.

Ah *_irregardless_*. It's not a word, people say. But from a lexicographer's perspective it is indeed a word. It's a combination of sounds that communicates a meaning; there's evidence of it in use as far back as 1795, though it was not common in American dialectal speech til early 20th c; it's likely a blending of *irrespective* and *regardless*; and it is primarily used in informal speech.

Irregardless was one of our 2020 Words of the Year (the words are chosen based on what people look up at merriam-webster.com): lookups for it increased dramatically in July, up 464%, when actor Jamie Lee Curtis tweeted that we had just entered the word. Ms. Curtis was wrong: we first entered *_irregardless_* in one of our dictionaries in 1934, and an entry for it has been at Merriam-Webster.com since the website debuted in 1996. What people are really saying when they say that *_irregardless_* isn't a word is that it isn't a respectable word that people should use.

And here is where we see another pair of terms: prescriptive and descriptive. Many dictionary users want and expect a dictionary to be prescriptive, to tell them what words they should and shouldn't use, to tell them what's good English and what's not. All of the major English language dictionaries are, however, not prescriptive but descriptive: they are designed to describe the language as it is used. They provide guidance about controversial words—there's a note at our entry for *_irregardless_* that recommends you use *_regardless_* instead so that your listener's disdain for your word choice does not distract them from your meaning; but dictionaries are dedicated to collecting all the established words of the language, in all their messy variation.

As Sidney Landau writes in *“Dictionaries: The Art and Craft of Lexicography”*, “All dictionaries based on usage—and all competently done dictionaries must be based on usage—are descriptive. Prescription is impossible to distinguish from bias. Any preferred usage or condemnation of existing usage necessarily reflects the educational or cultural background of the editor making such a judgment.”

And a dictionary is no place for reflections of editors’ backgrounds. A dictionary must reflect broad usage to have any legitimacy.

The commitment to descriptive lexicography extends to pronunciations: and so it is that Steve Wilhite, the person credited with creating what is referred to in writing with the letters *“gif”*, can report that it is correctly pronounced “jif,” while the dictionary provides both “jif” and “gif” as established pronunciations at its entry. Coining a word does not grant a person any special privileges with regard to a dictionary’s treatment of a word. We’re really only interested in what the speakers of the language have done with it.

These facts about dictionaries—they’re about words, not concepts; they reflect actual usage, not ideal usage—do not mean that a dictionary cannot rightly function to settle arguments and justify claims. In truth, it is exactly because of these facts that a dictionary can refute mistakes and misunderstandings. But those using a dictionary for such would be wise to acquaint themselves with the text’s stated aims before they use it for such.

It's a tough ask though: most people avoid articles and discussions that deal with the precepts of dictionary making, and no one wants to read the front matter of a print dictionary.

The chore is unavoidable for some of us: Merriam-Webster editors begin their lexicographical careers reading the front matter. It’s a task you’re assigned on your first day.

And it’s there that you learn such information as the structure of a dictionary entry, the order of entries (“alphabetical” is only the beginning), the function of the divisions in headwords (they designate not syllables but where printers typically allow a hyphen), and the basics of defining itself.

The noun side-eye dates at least as far back as 1797 (the verb appears to be a 20th century innovation) but it wasn't until 2016 that it met our criteria for entry: sustained meaningful use in a variety of sources over an extended period of time.

As the definer handling the term I considered 29 examples in our files including the following:

-a 2016 example from Time magazine: Though there's humor in the unexpected pairing, the actors play it with the innocence of children who do not yet count the judging side-eye as part of their vocabularies.

-a 2009 quote in the Sunday Mail from Carmen Van Kerckhove: "People have given him the side-eye for a while."

-a 2011 movie review in the Detroit Free Press referring to "some mysterious third guy who keeps giving him the side eye and chasing him around Manhattan."

-a 1967 uncredited quote from the Seattle Daily Times: "He side-eyed me to see if he had chosen the right approach."

-an Inverse headline from 2016:

With Zika Virus Headed North, American Scientists Side-Eye Asian Tiger Mosquitoes

-a 2016 use in Essence magazine: I naturally and perpetually side-eye every woman who can rock a pointed-toe shoe with ease.

-and from exactly a century before that, a 1916 example from the Arizona Republican: In his mind's-eye he saw himself associating with actor-folk, who invariably side-eye him and whispered among themselves: "That's Alonzo Gubbins—frightfully wealthy—..."

The 29 examples in our files do indeed show that side-eye had met our criteria for entry—sustained meaningful use in a variety of sources over an extended period of time: the term's function is clear in the examples; it's found in both quoted speech and prose writing; in newspapers and magazines with wide readership; there are examples from the US and Canada, from the UK, and from

Australia and New Zealand; and evidence going back a century, with the bulk of evidence from recent years.

Prior to the late 20th century, Merriam-Webster definers would have had to make do only with this kind of evidence, with what was in our files. These files, which we call the citation files, are examples collected by Merriam-Webster editors beginning in the late 19th century and now containing about 100 million indexed words in about 16 million individual citations. Each citation is an excerpt from text that shows a highlighted word in context, and that has been chosen by an editor because of its possible use in defining. Anything can be a source: citations come from comic strips, menus, gas station signs, novels, newspapers, academic journals, blogs, anything. Editors are supposed to spend time each workday collecting this evidence and most of us find that we can't help collecting it on non-workdays too.

The Merriam-Webster definers of this century have significantly more resources though: in addition to the citations in our files, I was also able to search for examples of *_side-eye_* in various electronic databases and corpora that Merriam-Webster licenses and subscribes to. Nexis showed the term to be gaining in use from the early years of the current century with significant use beginning gradually and increasing more dramatically from 2010 on. Various corpora at Mark Davies' English-Corpora.org offered still more evidence. Google Books had examples too. If I were to consider the term today I would also have the surfeit of corpora that we now have access to through Sketch Engine, our newest lexicographical tool.

The quantity of evidence can be overwhelming, and knowing how to sort and weigh it is very important. As I mentioned a moment ago, examples from menus and signage have been part of our evidence for a long time. We continue to pay attention to formal and informal use, to written and spoken use; we collect examples from Twitter and Reddit in addition to examples from the *_Times_* of London, New York, LA, Chicago-Sun, Tampa Bay, etc. But as has been the case for our entire history, we aim to define the established words of the language as reflected especially in published, edited text. We know full well that this means there are terms covered extensively in Urban Dictionary that we will never define; it is not our assertion that these terms are not words—of course they are words (at least most of them are). But when a word regularly appears in a variety of

published, edited texts over a period of time it has demonstrated a likely longevity; that's the kind of word we define.

Our defining criteria— sustained meaningful use in a variety of sources over an extended period of time—are intentionally vague. *_Fardel_* passes the test of sustained use mainly by appearing in Hamlet's soliloquy; *_yeet_* has had a hard time with the "meaningful" piece until quite recently (it's still in the not-yet pile); the name of a new element gets an enormous boost with entry in a single source, the scientific journal *_Pure and Applied Chemistry_*. And sometimes an "extended period of time" is almost no time at all: *_Covid-19_* was entered in our dictionary a mere 34 days after its coinage. The criteria for entry is the same for *intraocular pressure* as it is for *beer goggles*, but the lexicographer has to hold each term up to those criteria one at a time.

Side-eye met the entry criteria in a more typical way than any of those terms I just mentioned, and it was entered with a definition reading:

"a sidelong glance or gaze especially when expressing scorn, suspicion, disapproval, or veiled curiosity"

The entry also provides two examples, the second prefaced by the note "often used with *_the_*" and demonstrating that use.

The headword is hyphenated— *_side_ hyphen _eye_*—and the entry includes the hyphenless open variant *_side_ space _eye_*, identifying it as less common than the hyphenated form. (The most tedious part of creating an entry like this is tabulating the evidence to provide accurate information about these variant forms. I was relieved that the closed form *_sideeye_* was not in the running.)

The verb is not defined at all but is covered as what we refer to as an "undefined run-on": it's styled with a hyphen only and identified as both transitive and intransitive; its forms are given; and an example is included. The meaning of the verb is considered to be easily enough understood from the definition of the noun. The entry has a pronunciation and a date of first known use of 1797. The example justifying that date is not included but I'm going to read it to you just for fun:

It's from a text called *Remembrancer For Lord's Day Evenings*, dated 19 Mar. 1797. The text reads:

“Here we come to what calls for the strongest eye-sight, the most steadfast gazing. Our being in Adam [the writer by this means something like a deep kinship with Adam]... our being in Adam has been looked on with a side eye. The subject has provoked dislike; I may almost say, contempt. It is now painful to speak of it.”

I want to look for a minute at the definition itself. It is the most common of the four basic types of definition, what we call an “analytical definition.” An analytical definition begins by putting the word being defined—the definiendum—into the most limiting category that will include it; that category is referred to as the definition’s genus term. In the case of *_side-eye_*, the genus term is the phrase “glance or gaze.” That’s what the definition hangs on; a side-eye is a kind of glance or gaze. To the genus term are added differentiae that distinguish the definiendum from other members of that genus; in this case differentiae are “sidelong”—a sidelong glance or gaze—and then the specifying phrase “especially when expressing scorn, suspicion, disapproval, or veiled curiosity.”

If the verb had been considered to warrant its own definition it would likely have received a definition of a second type: the much-despised truncated analytical definition:

: to look at (someone or something) with a side-eye

Its definition includes the headword (or actually a homograph of the headword) in its text. Truncated analytical definitions are incredibly useful when you must keep a physical book portable and affordable: they allow the definer to take advantage of the reader’s access to an entry that is typically mere millimeters from the line on which their eyes currently rest. Truncated analytical definitions continue to be useful to the definer of an electronic dictionary, as they save effort there too. Such definitions are key to efficiency, as they permit the dictionary to say a good many things just the one time.

But especially in an electronic dictionary, such definitions are understandably maddening—look up *_sagacity_* and you find the truncated analytical definition “the quality of being sagacious,” which means you have to actually click on a hyperlink—horrors!—to learn the meaning of the word. But the truncated

analytical definition allows the dictionary to avoid repeating everything at the entry for *_sagacious_*.

A third type of definition is the periphrastic definition: sometimes no genus term is to be found, and a definer has to orient the reader by describing the semantic job the genus term would otherwise do. Definienda requiring periphrastic definitions are often simple terms—the ones that make good genus terms themselves. *_Ability_* has among its definitions “competence in doing something”; *_form_* is “the shape and structure of something as distinguished from its material”; *_hot_* is “having a relatively high temperature”; *_off_* is “from a place or position.”

A fourth type of definition is the definition by synonym: we typically use these for the less common word for a thing. *_Whataboutism_* gets a full definition: “the act or practice of responding to an accusation of wrongdoing by claiming that an offense committed by another is similar or worse; also: the response itself.” The British synonym *_whataboutery_* is defined with only a link to *_whataboutism_* (a synonymous cross-reference, we call it), but like *_whataboutism_* it gets three examples of the word in use at its entry.

There are two other common defining methods; they don’t strictly qualify as definitions but are more correctly understood as substitutes for definitions. Neither is preceded by the boldface colon that introduces a definition in Merriam-Webster’s dictionaries. (That boldface colon can be understood to mean “is here defined as,” so: “*_side-eye_* is here defined as”) The first non-definitional defining method is the functional description by usage note—this is used for such lexical items as trademarks (the only dictionary treatment *_Kleenex_* gets is an entry with the functional label *_trademark_* and the description “used for a cleansing tissue”). Functional description by usage note is also used for interjections: *_holy moly_* gets the description “—used to express surprise, amazement, or bewilderment.” And it’s common for function words: sense 1a of the preposition *_for_* is described with “used as a function word to indicate purpose.”

And the final common defining method is the expansion, which is simply the full words that are the source of an abbreviation or symbol. *_Tbh_* has “abbreviation” as its functional label; it gets an informal register label and three examples, and without even a dash just the words “to be honest” where a definition would go.

At the time that I defined *_side-eye_* my defining assignment was to look through a very long list of words that editors have over the years identified as being worth considering for entry and choosing the ones I thought most important to assess. A definer's job varies though: we commonly work through alphabetical sections of the dictionary, particularly when we are working on a new edition of a dictionary; or we have assignments relating to groups of entries: my assignment when I revised *_love_*, noun and verb, was to fully revise the definitions of words commonly looked up.

That assignment—to revise entries for words commonly looked up—wasn't possible before 1996, when the entire Merriam-Webster Collegiate Dictionary, Tenth Edition, was put online, for free.

Up to that point, there was no way for lexicographers to know which of their definitions was being consulted how often. But as the information superhighway started to get real traffic we started to get real information about where everyone was going.

In addition to *_love_*, there are some other perennial favorites at Merriam-Webster.com, and they will surprise no one. Right up there at the top? *_Affect_* and *_effect_*. Also the pair *_i.e._* and *_e.g._*. And then a bunch of common terms with abstract meanings: *_culture_*, *_integrity_*, *_pragmatic_*, *_ubiquitous_*, *_disposition_*. It's clear that people are often looking for clarification about a word they encounter pretty regularly and probably use themselves. Of course the dirty words and swear words are in there too—so also words people encounter regularly and likely use themselves. But perhaps clarification is a lesser motive.

We let this information guide our work: entries for commonly looked up words get mostly full analytical definitions and lots of examples; many of them also have paragraphs that discuss usage issues in depth.

People of course also look up words that we don't yet define, and we can make note of those words and monitor their development. We don't aim to be at the vanguard of neology or slang, as our criteria make clear, but these "failed lookups," as we call them, are useful nonetheless. In addition to the various articles on usage and punctuation at Merriam-Webster.com we have a regular

feature called “words we’re watching.” `_Fleek_` was addressed there years before the term met our entry criteria.

The internet as source of information—backing for our research into the meanings of words, and illumination of what dictionary users look for—is just one piece of what the internet has offered lexicography, though.

Arguably far more significant in the daily practice of lexicography is the shift from print to an electronic platform. The most challenging defining stricture—the need to keep entries as concise as possible to keep a book portable and affordable—is gone. We can now include whatever and as much information as we think will be useful to the reader.

The entries for `_gender_` and `_sex_`, for example, now include a lengthy note explaining the historical and current uses of both. The note addresses their earliest synonymous use—when `_gender_` expanded from its grammatical use to join `_sex_` in referring to either of the two primary biological forms of a species; and then the senses each acquired in the 20th century—the “sexual intercourse” meaning of `_sex_`—now its most common meaning, and `_gender’s_` use to refer to the behavioral, cultural, or psychological traits typically associated with one sex, as in “gender roles,” and later `_gender’s_` use in the compound terms `_gender identity_` and `_gender expression_`, and its eventual use as a synonym of `_gender identity_`. The note also addresses the prescribed distinctions between the words—`_sex_` for biological forms, `_gender_` for psychological and sociocultural traits—and the fact that the prescribed distinctions tend to be observed in technical and medical contexts, but are frequently not observed in journalistic and general use.

This kind of information offers the reader perspective and nuance that formerly was impossible for us to provide.

Such information is also a response to the other side of the dictionary-internet coin: Not only does the internet provide us with information about how words are used and what words people look for, but it also makes it possible for dictionary users to actively engage with lexicographers and the dictionary itself. That usage note on `_gender_` and `_sex_` is in some ways a reply to queries and complaints and proclamations about our definitions for the terms. It was clear from what we

were hearing from dictionary users that a note like that was needed if clarity of the terms' meanings and functions is what we were offering.

We've always received correspondence from dictionary users—I read you some emails earlier but we receive letters too; in the basement of the Merriam-Webster building there is a tall and narrow room, called the vault—it's about 4 x 15 with a 15' ceiling—that is full of file drawers, many containing letters from correspondents (we throw away nothing)—but Merriam-Webster is no longer a silent logo with a post office box and email address. We tweet, and people tweet back; we announce that we've entered new words and people take to social media to applaud or excoriate.

Anonymity used to be guaranteed the lexicographer—our recently retired Director of Defining Steve Perrault had a hand in many thousands of definitions consulted by millions of dictionary users daily and oversaw major changes to our practices but his name is unknown by nonlexicographers (which doesn't bother him one whit, I'm sure). But anonymity for the lexicographer can no longer be taken for granted. Nowadays you might find yourself on a podcast, or get drafted to appear in videos that go on a website, where commenters will voice opinions on your hair or on your evolving choice of eyewear.

The interactions between the public and the dictionary don't always go smoothly. It's not infrequently that we're accused of using our Word of the Day to throw shade at someone or as a comment on some current news item; in truth it's all coincidence—that daily vocabulary feature is prepared weeks in advance and very much aims to avoid controversy. Our Twitter feed does not have the same excuse but coincidence also sometimes gets mistaken for a jab there.

We've learned some important lessons from these interactions over the years. Merriam-Webster first declared a Word of the Year in 2003—it was `_democracy_`. (The American Dialect Society started the whole Word of the Year thing in 1990, with the word `_bushlips_`—a comment on the president at the time; in 2003 the members of the society chose `_metrosexual_`.) That 2003 word was based on lookups at Merriam-Webster.com. But in 2006 Web.2.0 was all the rage—dictionary readers were dictionary users and user participation was key. We opened the Word of the Year selection to users, and they chose Stephen Colbert's coinage `_truthiness_`. We did the same thing in 2007 and users chose `_w00t_`.

(Neither of these was a word we even defined at the time.) In 2008 Merriam-Webster went back to its traditional data-driven Word of the Year procedures, and a lucky editor narrowly escaped having to write a press release about Wordy-McWordface.

We'll be announcing our 2021 Word of the Year soon—2020's word was _pandemic_—and the selection will say something about what people ask of a dictionary, and about what a dictionary offers.

Merriam-Webster's dictionaries still make the same bold offer they've made for a long time now: to tell the truth about words. The changed relationship between the dictionary and the dictionary user that the internet has wrought means that that truth is closer to a real catalog of the language than it ever has been. It used to be that the lexicographer's evidence was mainly what had been permitted a platform—and a platform that came at a significant price—a price reflected in a newspaper's distribution, a magazine's subscription rate, a publisher's bestsellers list. But the internet so dramatically lowers that price, and so dramatically democratizes the dissemination of the written word—and even the published, edited written word—that dictionaries can now see the language as many more of its speakers actually use it. The change is subtle and significant—and the result is a dictionary that's truer than it's ever been before.