

@ing in the #OED: Twitter/X as Citation Evidence in the *Oxford English Dictionary*, 2018–2025

David-Antoine Williams 

St Jerome's University in the University of Waterloo (david.williams@uwaterloo.ca)

Abstract

Between 2018 and 2025, the social media platform Twitter, and its successor X, emerged as the most important source of language evidence in updates and revisions to the *Oxford English Dictionary* (OED). At the start of 2026, Twitter/X accounted for over 5,200 evidentiary quotations in the dictionary, including first evidence for 213 new words and senses. Based on detailed analysis of the underlying XML-encoded data of several historical iterations of the OED published between 2017 and the present, this article studies the dictionary's recent turn towards Twitter/X as a major source of lexical evidence in updates to the OED since 2018, for words and senses both new and old, remarkable and un-. It begins with brief history of the OED's incorporation of Twitter/X evidence over the last decade, noting ways in which that evidence differs from that of other major sources; then it discusses particularities of the Twitter/X corpus and factors affecting its suitability as evidence in historical lexicography; and finally it discusses some of the unique aspects of new and current English that Twitter/X has allowed the OED to document, before offering some concluding caveats.

Keywords: lexicography; *Oxford English Dictionary*; Twitter; X.com; social media; regional English; slang; neologisms; new words; language change

Introduction

On 11 February 2020 at 15:09 GMT the World Health Organization's @WHO Twitter account relayed news that the thing theretofore referred to as the '2019 novel coronavirus' (among other names) would now officially be called 'COVID-19'.¹ Two hours later, at 17:11 GMT, a British physician specializing in fungal infections wrote on Twitter, with ironic but predictive imprecision: 'Names matter. The novel #Coronavirus has how been named #Covid'² (Figure 1).

It was 'Covid', rather than 'COVID-19' or another of the half-dozen related forms, that would become the lingua franca appellation of the virus, and the disease it caused, which the *Oxford English Dictionary* (OED) recognized with uncharacteristic speed, in July 2020, in the second of two special updates that year, citing Twitter as the earliest known use of this term (the usage of 'Covid' to refer to the ensuing period of emergency measures has not yet been recognized).

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDeriv licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.



Figure 1. 20 February 2020 tweet by @DrNeilStone (Internet Archive instance) <<https://web.archive.org/web/20200211201831/https://twitter.com/DrNeilStone/status/1227278889103372293>>.

In all, sixty-two new Covid-related words and senses were published in the OED in extraordinary updates in April and July 2020. Clearly the magnitude of the event and its reverberations throughout the language warranted this diversion from the programme of lexicographical work that has been unfolding at the OED since the Third Edition began to publish quarterly updates in 2000. With only two or five months of contemporary currency, these were among the most swiftly documented new lexis in the history of the dictionary,³ casting aside the rule of thumb that at least ten years of current use be required before a word can become eligible for inclusion as an OED lemma (Simpson 2013). All of the most recent evidence was thus necessarily drawn from publication forms with short lead times, even when the new headword had a long history of theretofore undocumented previous use in other contexts (e.g. HYDROXYCHLOROQUINE, PPE, SOCIAL DISTANCING). For the true contemporary neologisms, like COVID-19 and its derivatives, Twitter supplied five of the eleven quotations given as citation evidence, including the earliest citations for each.

By the time Covid sequestered OED lexicographers behind their home computers—with much of the world’s population similarly confined to quarters, much of it with devices to hand—Twitter had already become an important source in documenting current English, including several neologisms, most of them particular to the platform itself. In the lockdowns Twitter became a primary textual connection point to the language’s attempts to give words to the circumstances. As one OED editor said, ‘To be able to witness the point at which this term [COVID-19] came into being had a profound effect on me’ (McPherson et al., 2020).⁴

As this article goes to press, at the start of 2026, the OED is witness to 213 words and senses first attested on Twitter/X—from social media jargon (e.g. CATFISHING, LIKE (*n.*²)), to more general computing and technological vocabulary (CRYPTOCURRENCY, ZOOM (*v.*²)), to regionalisms (ACEGIRL, FROCKSUIT), colloquialisms (AMAZEBALLS, WTAF), and topical buzzwords (SHRINKFLATION, SOCIAL DISTANCE)—with most additions published in quarterly updates to the dictionary since the start of 2018. Additionally, and dwarfing these in number, Twitter supplies 5,000 more citations of recent use in entries for words as unremarkable as BREAK (*v.*) (giving the most recent citations in twenty-one separate senses and uses) and DIE, (*v.*¹) (seven quotations in seven senses), and as exotic as APPENDICATION and ST. ANTHONY’S TURNIP.

In this article I study the OED’s recent turn towards Twitter, lately renamed X, as a major source of lexical evidence in revisions, additions, and updates to the OED since 2018,⁵ for words and senses both new and old, remarkable and un-. I begin with brief history of the OED’s incorporation of Twitter/X evidence over the last decade, noting ways in which that evidence differs from that of other major sources; then I discuss particularities of the Twitter/X corpus and factors affecting its suitability as evidence in historical lexicography; and finally I discuss some of the unique aspects of current English that Twitter/X has allowed the OED to document, before offering some concluding caveats. Throughout the article, the data analysed derive mainly from a set of full-text XML instances of the OED published

between May 2017 and December 2025, and their associated publication histories and citation data.⁶ These have been supplemented using the web interface of the most current *OED Online* (updated December 2025), and with datasets compiled from Twitter and X.

OED's twitter/X evolution

When TWEET (*v.*) and TWEET (*n.*) first appeared in the OED, in the June 2013 update, the event earned a brief public mention by Chief Editor John Simpson, who noted the exceptional rapidity (seven years) with which these words had earned their inclusion in the dictionary (Simpson 2013). Simpson's update bulletin was noticed in the news media worldwide, often in terms that implied institutional legitimization of the social media platform and the vocabulary it was spawning, or spreading.⁷ At the time, however, OED lexicographers were not themselves looking at Twitter for citation evidence, with the result that the first uses of these terms were incorrectly attributed to an anonymous Yahoo Groups user (for the transitive verb, on 28 April 2007), and the blogger Neville Hobson (a.k.a. 'NevOn', for both the intransitive verb, on 15 March 2007, and the noun, on 11 March 2007).⁸ If inclusion of a new sense in under ten years from first use is exceptional, revision within a single year is even more so, but this is what happened to all three senses of TWEET when OED lexicographers got hold of Twitter itself: new citations of 'Twitter postings' from 15 January 2007, 27 December 2006, and 23 November 2006, respectively, quietly replaced the blog and Groups citations. Thus the OED's first published citations of Twitter were antedatings of words coined on Twitter.

By mid-2015, the OED counted twenty-four Twitter citations in the same number of senses, twenty of them first attestations, mostly of words, like TWEET, with particular or exclusive applications on Twitter (e.g., HASHTAG, RETWEET), or on social media and the Internet more broadly (e.g., FAP, LIKE (*v.*), NSFW). Although citation records indicate that Twitter may very occasionally have been consulted to collect examples of regional usages as early as September 2011 (for FISH CAKE, published 2022), February 2012 (for FRESH (*n.*), published 2022), and April 2014 (for TELL PIE, published 2015), only in the Spring of 2017 did OED lexicographers begin to access Twitter regularly as a major source of general language evidence, with large numbers of quotations coming through to publication starting in the following year. As of the June 2017 update there were only 53 citations of Twitter published in the dictionary, including twenty-four first attestations; by the end of 2018 the number of quotations had increased tenfold to 457, illustrating 428 senses, with only thirty-nine first attestations.

Since then, the OED has increased its citations of Twitter, and its successor X,⁹ tenfold again (Figure 2), to a level greater than any other source of language evidence added to the dictionary over the period. In revisions and additions published between the start of 2018 and the end of 2025, Twitter has been the most cited source of new quotations all told, and individually in each year since 2019 (Figure 2).¹⁰ This is especially striking given the relative recency of the source: Twitter only began life in 2006, whereas the next nine top sources are all newspapers in their second or third century of publication (prominently the *Times* (London), the *New York Times*, the *Daily Telegraph*, and the *Guardian*), and with excellent digital archival preservation.

On an equal chronological basis, considering only those new quotations dated 2006 or later, Twitter/X is far and away the most cited source in updates and revisions between January 2018 and December 2025 (Table 1b)—4.6 times the next-most cited source of all such recent quotations (*Times*), and five times the next-most cited in newly documented words and senses (*New York Times*).¹¹ Notably, of all the newly documented words and senses added to the dictionary since 2018, 434 had senses first attested 2006 or later, of which 184 had their first uses traced to Twitter/X, or 42 per cent of all recent new words and senses (occurring in 4 per cent of all Twitter/X citations). Of all the post-2006 evidence published in the OED in this time period—totalling almost 70,000 quotations from 27,000 different sources—7 per cent comes from Twitter/X.

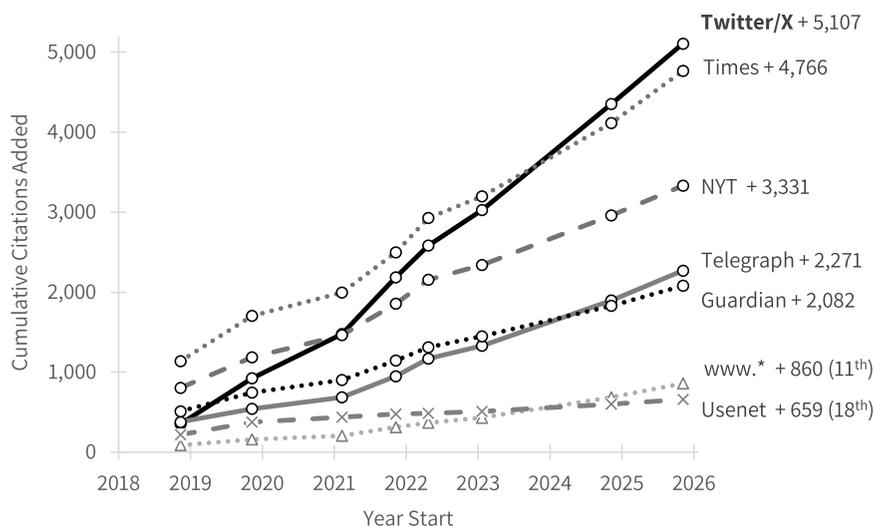


Figure 2. Top five most prevalent sources of OED quotations added in updates between January 2018 and December 2025 (January 2018 baseline). Two additional digital-only sources (Usenet, and all websites) are included on an aggregated basis, with their rank.

In addition to being more recent, and more weighted towards first evidence, the profile of Twitter/X citations is distinct from other major sources in a number of ways. The proportion occurring in senses marked for regionality and for register, for instance, is 36 and 30 per cent, respectively, about twice the overall averages for these subsets, whereas other highly cited sources are significantly below average (Table 1). Individually this overweighting holds for the most common regional and register labels, excepting *historical* and *archaic*, and is especially striking for the regional labels *African American English* and *Caribbean English* (these sometimes overlapping), cited at four times and eight times the overall rate, respectively; and for the register labels *colloquial and slang*, *offensive, derogatory*, and *humorous*, cited at roughly 3–4 times the average (Table 1a).¹²

The twitterverse as a lexicographical corpus

Even if Twitter/X has made possible more extensive and quicker documentation of certain kinds of restricted English lexis, the sheer volume of Twitter quotations suggests that much or even most of the evidence cited there is not categorically to be distinguished from run-of-the-mill attestations selected for new and revised senses in the dictionary. As Jesse Sheidlower has detailed, the basic desiderata for a quotation paragraph in a historical dictionary such as the OED are chronological and connotative coverage: quotations should establish (as far as possible) the earliest published use, as well as a relatively recent usage, and as many in between as needed to show development or continuity, and they should illustrate range and shades of meaning and usage not fully capturable in the definition. Oddball and unillustrative quotations are to be avoided, as well as uninteresting ones, as far as possible: ‘They should be fun to read, from authors a reader would want to track down’ (Sheidlower 2011: 206). A further consideration, which must be weighed against the others, is the time and resource investment by a dictionary project with many millions of quotations yet to research: all other things being equal, the easily findable evidence is preferable to the recondite.

On these criteria, Twitter/X has many desirable properties in a corpus: it is large and diverse, with hundreds of millions of accounts producing hundreds of billions of words of text every year, from devices located in every corner of the world; it is historical (albeit recent),

Table 1. Profile of OED quotations added between January 2018 and December 2025 by marked attribute, Top 5 Sources. Regional labels have been consolidated.

	All Sources		Twitter/X		Times		NYT		Telegraph		Guardian	
	400,531 (100%)	5,107 (100%)	4,773 (100%)	580 (10%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
Marked Regional	61,301 (15%)	1,850 (36%)	465 (19%)	580 (10%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
North American English	32,668 (8%)	947 (19%)	168 (14%)	553 (17%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
Britain and Ireland	29,332 (7%)	740 (14%)	327 (14%)	21 (1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
Australasian English	8,479 (2%)	204 (4%)	47 (4%)	13 (<1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
Caribbean English	1,513 (<1%)	158 (3%)	3 (3%)	1 (<1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
African English	1,925 (<1%)	99 (2%)	10 (2%)	5 (<1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
South-East Asian English	318 (<1%)	26 (1%)	1 (1%)	1 (<1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
Indian English	837 (<1%)	18 (<1%)	7 (<1%)	2 (<1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
Marked for Register	67,666 (17%)	1,551 (30%)	607 (24%)	385 (13%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
colloquial and slang	28,109 (7%)	1,219 (24%)	217 (24%)	248 (7%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
derogatory	3,511 (1%)	167 (3%)	26 (3%)	27 (1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
offensive	3,041 (1%)	149 (3%)	32 (3%)	22 (1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
disparaging	3,294 (1%)	91 (2%)	38 (2%)	26 (1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
historical	26,115 (7%)	63 (1%)	287 (1%)	85 (3%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
humorous	1,242 (<1%)	57 (1%)	10 (1%)	11 (<1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
archaic	6,213 (2%)	42 (1%)	40 (1%)	8 (1%)	3,326 (100%)	2,268 (100%)	2,086 (100%)	292 (14%)	2,268 (100%)	2,086 (100%)	292 (14%)	
b. Qs Dated 2006+	68,917 (100%)	5,106 (100%)	1,120 (100%)	1,207 (100%)	1,131 (100%)	1,048 (100%)	1,048 (100%)	120 (11%)	1,131 (100%)	1,048 (100%)	120 (11%)	
Marked Regional	11,890 (17%)	1,850 (36%)	129 (36%)	167 (12%)	1,131 (100%)	1,048 (100%)	1,048 (100%)	120 (11%)	1,131 (100%)	1,048 (100%)	120 (11%)	
Marked for Register	11,506 (17%)	1,551 (30%)	139 (30%)	105 (12%)	1,131 (100%)	1,048 (100%)	1,048 (100%)	120 (11%)	1,131 (100%)	1,048 (100%)	120 (11%)	
In New Senses	33,454 (49%)	3,300 (65%)	441 (65%)	641 (39%)	1,131 (100%)	1,048 (100%)	1,048 (100%)	120 (11%)	1,131 (100%)	1,048 (100%)	120 (11%)	
In Revised Senses	35,463 (51%)	1,806 (35%)	679 (35%)	566 (61%)	1,131 (100%)	1,048 (100%)	1,048 (100%)	120 (11%)	1,131 (100%)	1,048 (100%)	120 (11%)	
First in Sense	435 (1%)	184 (4%)	3 (4%)	4 (<1%)	1,131 (100%)	1,048 (100%)	1,048 (100%)	120 (11%)	1,131 (100%)	1,048 (100%)	120 (11%)	
Last in Sense	64,127 (93%)	4,512 (88%)	1,059 (88%)	1,095 (95%)	1,131 (100%)	1,048 (100%)	1,048 (100%)	120 (11%)	1,131 (100%)	1,048 (100%)	120 (11%)	

with posts going back to 2006 continuously hosted on the live site, and sometimes preserved on publicly-accessible archives such as Internet Archive (a.k.a. ‘The Wayback Machine’), as well as some privately preserved archives; and it is freely accessible, with few barriers to posting and accessing posts.

Still, one might ask what has made Twitter/X in particular such a dominant source of OED evidence, even when compared with the Word Wide Web at large, or other stable, open, text-based, multiuser web sites (such as Reddit, for instance, cited only twelve times). The early internet discussion system Usenet is the only other online platform with significant representation in OED citations, with 2,101 citations as of the December 2025 update, but it has been only a moderate and declining source of additional quotations since 2019. Over the same period, the OED has increasingly quoted from individual web sites, but still far less often than from Twitter/X: from 2018 to 2025, 860 new quotations have been added from all web sources in aggregate, the total reaching 1,370 as of December 2025, excluding twitter.com and x.com, and sites archiving older print sources. Individually, further excluding legacy news sites accessed through aggregating services such as Nexis (among which, the top two source are [telegraph.co.uk](https://www.telegraph.co.uk), with 69 citations, and [mirror.co.uk](https://www.mirror.co.uk), with twenty-seven), the most frequently cited websites are [mumsnet.com](https://www.mumsnet.com) (twenty-six citations), [Salon.com](https://www.salon.com) (twenty-one), and *Urban Dictionary* (twenty-one). The remaining online sources include various news publications with a handful of citations each, and a large number of one-off attestations from diverse individual websites.

One public explanation given by OED staff for the turn towards Twitter has been its accessibility and durability—specifically, the lack of an ‘edit’ capability (Dent 2019).¹³ Thus while other social media may also capture a wider variety of language than typically appears in print (or in legacy online media), it may be more difficult to document there in accordance with OED editorial standards. A 2018 ‘OED Appeal’ for ‘youth words’, for example, made a nexus between technical durability and demographics, pointing out that ‘young people’s language today can be particularly elusive’ because ‘newer ephemeral modes of communication (texting, WhatsApp, Snapchat, etc.) make it difficult to monitor and record this kind of vocabulary’ (Allen 2018). Another appeal for local and regional words also highlighted the ephemerality of spoken language, inviting readers to join ‘the conversation on Twitter with the hashtag #wordswhereyouare’ (Gibson 2018). To take a longer view, digitality writ large has doubtless textualized whole worlds of language that would previously never have seen print—or if so only much later, likely intermediately, and fractionally—and as a result never been eligible for OED consideration; but in the shorter term, Twitter/X in particular has represented an accessible and somewhat durable means of diversifying the demographic provenance of OED citations, and thus of the registers and restricted usages the dictionary is able to document.

Among the broad English speaking population, the current citation evidence shows that Twitter/X has indeed diversified substantially the authorship profile of OED citations in a number of ways. Such diversification is a goal which has repeatedly been articulated since the inception of the wholesale revision project known as the Third Edition of the OED (e.g., in Simpson 2000). However, even revised sections of the OED continue to significantly underrepresent certain demographics. For example, in a recent study, the present author showed that female-authored sources are underrepresented compared to male-authored sources, even in the new quotation evidence added since 2000, for all time periods covered, only reaching about 30 per cent for the most recent evidence (Williams 2023).¹⁴ The OED’s citations of Twitter/X, on the other hand, are 45 per cent female-authored, where the author gender is indicated or can be inferred, despite what is by most accounts a male-weighted usership;¹⁵ and just over one per cent of OED citations of Twitter/X—46 quotations—are from accounts identifying the author as nonbinary, genderfluid, or genderqueer, making Twitter/X by far the most productive single source for evidence from these groups (much of it, but not all, for key words in discourses about gender and sexuality).

Table 2. OED quotations added between January 2018 and December 2025 in senses marked for region, top two sources per regional label (percentage of total in regional label category). Some sub-categories in the OED taxonomy have been consolidated.

Regional Label	All Sources	Most Common Source		2nd Most Common	
<i>British English</i>	6,337	Times	198 (3%)	Twitter/X	195 (3%)
<i>English Regional</i>	8,893	Twitter/X	241 (3%)	Engl. Dial. Dict.	148 (2%)
<i>Manx English</i>	229	Twitter/X	21 (9%)	Mona's Herald	17 (7%)
<i>Scottish English</i>	9,881	Twitter/X	177 (2%)	Dict. Older Scot. T.	139 (1%)
<i>Welsh English</i>	381	Twitter/X	27 (7%)	Western Mail	14 (4%)
<i>Irish English</i>	3,611	Twitter/X	79 (2%)	Irish Times	79 (2%)
African Englishes					
<i>East African English</i>	391	Twitter/X	41 (10%)	Daily Nat. (Kenya)	26 (7%)
<i>South African English</i>	1,090	Twitter/X	22 (2%)	Sund. Times S.A.	18 (2%)
<i>West African English</i>	210	Twitter/X	11 (5%)	Afr. News (Ghana)	3 (1%)
<i>Nigerian English</i>	234	Twitter/X	25 (11%)	Nation Nigeria	12 (5%)
<i>Indian English</i>	837	Times of India	69 (8%)	Twitter/X	18 (2%)
South-East Asian Englishes					
<i>Hong Kong English</i>	3	Times	1 (33%)	H.K. Surgeon	1 (33%)
<i>Malaysian English</i>	22	Twitter/X	5 (23%)	Sarah Bishop	5 (23%)
<i>Philippine English</i>	211	Business World	16 (8%)	Daily Inquirer	15 (7%)
<i>Australian English</i>	5,262	Syd. Morn. Her.	156 (3%)	Twitter/X	125 (2%)
<i>New Zealand English</i>	3,217	Twitter/X	79 (2%)	New Zealand Herald	67 (2%)
<i>North American English</i>	7,197	Twitter/X	151 (2%)	New York Times	145 (2%)
<i>U.S. English</i>	23,467	Twitter/X	608 (3%)	New York Times	399 (2%)
<i>African American English</i>	1,144	Twitter/X	156 (14%)	NY Amsterd. News	11 (1%)
<i>Hawaiian English</i>	32	Twitter/X	3 (9%)	Pacific Voices	1 (3%)
<i>Canadian English</i>	828	Twitter/X	29 (4%)	Tor. Globe & Mail	12 (1%)
<i>Caribbean English</i>	1,227	Twitter/X	130 (11%)	Daily Gleaner	25 (2%)
<i>Jamaican English</i>	286	Twitter/X	28 (10%)	Daily Gleaner	23 (8%)

Another recent study by the present author described the OED's remedial documentation of 'World English', and words and senses marked 'regional' more generally, in the Third Edition (Williams 2021). Here Twitter/X has again been the most frequent and important contributor of quotation evidence published in revisions and additions since the start of 2018 (Table 2): for regional Englishes with significant numbers of quotations overall, during this time period the site is the most frequently quoted source for all but three regional categories,¹⁶ often by far; and for the three major regions in which it is not the most cited source—Ireland, India, and Australia—it comes in second (only for Philippine English is Twitter/X not among the top two sources). Notably, Twitter/X has supplied 14 per cent of the evidence in senses labelled *African American English* (155 of 1,144 quotations)¹⁷ and ten per cent or more in each of: *Caribbean English* (130 of 1,225), *Jamaican English* (28 of 286), *East African English* (41 of 392), and *Nigerian English* (25 of 234).

Although Twitter/X is large, easy to access, diverse, searchable, historical within the period, and somewhat durable, each of these advantages has limitations and is susceptible to erosion over time. Posts may be deleted and accounts removed or restricted (by the user), or else suspended (by the platform), making the original text unretrievable unless independently archived; and policies for accessing or posting content on the platform may change according to the discretion of the company's decision makers. For example, searching the Twitter/X historical archive to retrieve a Tweet, while possible via the web, is technically capricious and requires a certain amount of platform expertise to achieve even moderately reliable results; while API (i.e. programmatic) access to historical searches was disabled in 2023 for all but the highest subscriber tiers, for which X charges many thousands of dollars monthly. Thinking at the chronological scale of the OED—not in terms of decades, that is, but centuries—it is well within the realm of possibility that the entire Twitter/X corpus could disappear or become inaccessible to the general public over time.

A future with a pastless Twitter, or Twitterless past, is already somewhat upon us. By the time the OED published its December 2025 update, 143 of the 755 new Twitter/X citations added to the dictionary since January of the same year—19 per cent—were already not retrievable on x.com. Overall, less than two thirds (64 per cent) of the 5,213 Twitter/X quotations cited in the current OED could be retrieved on x.com in December 2025: 333 (6 per cent) referred to posts from restricted accounts, 141 (3 per cent) from suspended accounts, and 921 (18 per cent) from deleted accounts; and a further 484 cited posts (9 per cent), although from active and accessible accounts, had been deleted by the user (or proved resistant to multiple search query formats and attempts).

Naturally, less recent posts are less likely to persist than more recent ones. The durability of Twitter/X material degrades over time and is subject to inflection events causing the deletion of posts and accounts, calling into question the lasting value of the platform as a historical language corpus. Even if, as I understand the case to be, OED practice is to preserve screenshots of posts as they are collected by researchers, such an archive is of no practical use to the dictionary reader, for whom the difference between an unverifiable quotation and a fabricated one is only a matter of authority.

As Sheidlower implies (Sheidlower 2011), and as John Considine has made explicit, when it comes to quotations, a ‘dictionary compiled on philological principles needs to supply enough information for a reader to consult the right page of the text in question’ (Considine 2009: 635), and ease of follow-up should be among the dictionary’s desiderata.¹⁸ In an irony typical of the period in question, what Considine perceived to be a new strength of the online dictionary vis-à-vis its printed ancestor—that the wide availability of digital texts would enable robust interaction with the dictionary’s sources—in the case of the OED’s most digital source turns out to be precisely the opposite: much of it no longer exists, or if it does, cannot be accessed; and daily it fritters away.

On archival principles, therefore, Twitter/X may be said to resemble more those undigitized, fragile, and closely protected historical documents sometimes cited by the OED out of necessity. On principles of access and verifiability, a growing share comes to resemble even the evidence fabricated (again, by necessity) by editors of the First Edition (1888–1928), who labelled it ‘Mod.’ (for ‘modern’) based on their own linguistic intuitions, in a glaring exception to the historical principles upon which the dictionary was founded. Such evidence has generally been employed as a measure of last resort, however, and has never been as prominent or routine in the OED as Twitter/X has lately become.¹⁹

And, though very large and inclusive of virtually any text-producing group or subgroup of people on the planet, the population of Twitter is as a whole neither demographically representative, nor stable over time. This creates the potential for skew in the colouring and connotative range of terms quotations will illustrate. To give an example of just one kind of skew, a well-known Pew study of the 2020 American election cycle found that 10 per cent of U.S. accounts generated 92 per cent of tweets, and of those top tweeters 69 per cent skewed Democratic politically (Pew 2020). From a lexicological angle, one might observe specifically in the American political context that scheduled revisions in 2019 of DOTARD, DOTARDAGE, DOTARDISM, DOTARD-LIKE, and DOTARDY are now and for posterity each coloured by the coincidental trending and bending of antiquated ‘dotard’ on Twitter, following a sensational political news event in late 2017.²⁰

Recent studies have indicated that a median shift in political attitudes represented on the platform occurred following its 2022 acquisition and rebranding as ‘X’ (Pew 2024). In addition to the overtly political or socially contentious lexis marked as such in the OED definition or usage note, for terms that might otherwise be taken as neutral, the median bias of the medium as a whole may be relevant in inferring shadings of sense and usage. For example, a sense of the prevailing discourse around Covid and vaccination science on Twitter/X likely should inform the reception of ANTI-SCIENTIFIC as intended by the user quoted in the OED saying, ‘Masks are a handy tool for quickly identifying the anti-scientific among us’ (@mdufay, 11 December 2022; published in OED in 2023).²¹

Other factors influencing the composition of the platform's usership may include future national restrictions on access, or the loosening of these where they are currently enforced, and regional and demographic trends in social media usage, including the rise (and fall) of Twitter-style X alternatives such as Threads, Mastodon, or Bluesky (none of them cited in the OED as of this writing), and their relative popularity among various population segments, often differentiated by age and nationality.

One last aspect that merits consideration, however speculative, when assessing Twitter/X as a historical language corpus, is the substantial and growing amount of non-human-generated language on the platform. Although I am unaware of any specific OED policy on text produced by large language models (ChatGPT, Grok, and the like), *prima facie* such language would seem ineligible for inclusion as usage evidence. Yet it has been known for some time that a large number of Twitter/X accounts are automated 'bots', which in recent years have been harnessed to language generating engines to produce 'slop' (a term not yet documented in the OED, but selected as the Merriam-Webster 'Word of the Year' for 2025, and defined there as 'digital content of low quality that is produced usually in quantity by means of artificial intelligence'). The scale of the slop problem for Twitter/X-based lexicography has not been estimated, but given the general difficulty in screening for bots or generated language, and the prevalence of both bots and human-bot interactions on the platform,²² it is one that is likely to increase. In the longer term it remains to be seen whether social media autoslop differs meaningfully from the artificially generated text that has lately also supplied longer formats such as online news, magazines, and journals, and e-books, and even 'traditional' print publications.

'The discourse' and earliest attestations

Somewhere in a computer that stores draft material for OED revisions there may well be a note to update *DISCOURSE* (*n.*), last revised in 2013, to include the sense dominant on Twitter/X, meaning something like 'widespread commentary or argument on a trending topic, especially if controversial or inflammatory.' If and when such an update is published, it will likely cite a Twitter account as the innovator of this sense.²³ Though the great majority of Twitter/X citations are used for routine contemporary documentation of unremarkable language use, the minority of peculiar, innovative, and otherwise remarkable lexis documented through Twitter/X is nonetheless much larger than for any other contemporary source, and for the rest of the OED quotation corpus taken in aggregate. This explains, among other effects, both the mean difference for marked lexis documented in the tables above, and the high number of earliest attestations attributed to Twitter/X.

In addition to producing 'the discourse' and other Twitter-specific vocabulary (metonymic 'account', as I use it here, is not yet documented in the OED), the discourse has been responsible for a number of important lexical innovations that have been widely diffused into the general culture. As of the December 2025 update the OED included 213 new senses first attested on Twitter/X, 114 of which form new entries (condensed in [Table 3](#)). Almost a third of these new senses (68, or 32 per cent) were published in less than ten years from first attestation.

OED first attestations are in the popular imagination very often misrepresented as attribution of coinage (or of novel extension, if for a later sense), especially if the author in question attracts significant cultural connotations (the pre-eminent example is Shakespeare). Almost all first attestations in the OED entries are not true coinages, but rather the nearest a lexicographical researcher could arrive at the earliest textual occurrence of a word, given the limitations of the lexicographer's time and the availability (and, today, searchability) of texts.

One reason Twitter/X has been responsible for so many recent first attestations is that the discursive spontaneity of its very large and distributed usership makes it a sort of digital listening post for the new language being generated in the anglosphere at large.

Table 3. Words and senses with first attestations in Twitter/X, as of the December 2025 OED update, by primary category. Asterisks* indicate first citations in the entry (first sense). Multiple new senses per entry are compressed.

General Tech. & Computing	4D printing*; binge view*; butt dial*; chargeless; cypypasta response*; crossplay; cryptocurrency; drone; droning; jailbreak (<i>n.</i> & <i>v.</i>); jailbreaker; jailbreaking; jailbroken*; mukbang*; non-fungible token*; phabler*; prompt engineer*; prompt engineering*; ride-hailing*; selfie stick*; stablecoin; tap (<i>v.</i> ²); unmute; Zoom (<i>v.</i> ²)*
Social Media	at (<i>v.</i>)*; catfishing; chef's kiss (<i>n.</i> & <i>v.</i>); don't at me; direct message (<i>n.</i> & <i>v.</i>)*; DM (<i>n.</i> & <i>v.</i>)*; dotcom; dox*; doxer*; followee; follower; ghost (<i>v.</i>); ghosting; hashtag (<i>n.</i> & <i>v.</i>)*; hashtagged*; hashtagging*; how; hypebeast; like (<i>n.</i> ²); listicle*; livestream; livestreamer*; massive open online course*; [me-too]; MOOC*; mute (<i>v.</i> ³); retweet*; retweet*; retweet*; retweeted*; retweeted*; retweeting*; share (<i>n.</i> ² & <i>v.</i> ²); social; social; social listening*; this; to follow back; trend; trend; trending; tweet (<i>n.</i> & <i>v.</i>)*; tweeted*; tweeter; tweeting*; twitter (<i>v.</i> ⁴)*; Twitterati*; twitterer (<i>v.</i> ²)*; twittering (<i>n.</i> ²)*; twittersphere*; unfollow*; unlike (<i>n.</i> ² & <i>v.</i> ²)*
Regional	acegirl*; agric; ati; bambi; beazy; bias, <i>n.</i> ; boasty; Bongo; bounce (<i>v.</i>); bully van*; catch-up (<i>n.</i>); cuffing season*; daebak; frock suit; gansey; jand*; japa (<i>n.</i> & <i>v.</i>)*; maft; mpango wa kando*; mug (<i>adj.</i> ²)*; oppa; orale; roadman; showrooming*; singeli*; spice bag*; to feel arlarse on; waste youth*; wrinch (<i>v.</i> ²)*
Register	actual; amazeballs; bants*; chill (<i>n.</i>); chonky*; dicksack; dumb hard*; lit; omnishambolic*; shithousery*; simples (<i>int.</i>)*; stan (<i>v.</i>); thot*; thottie*; to bake in; to cock about; to get in the bin; to go dumb hard; top bins*; totes (<i>int.</i> & <i>adv.</i>)*; WTF*; YOLO (<i>adj.</i>)*
Social Discourse	Afro-Latinx*; BIPOC*; body-shame; body-shamer*; cancel culture*; cuck; cuckservative*; deadname (<i>n.</i> & <i>v.</i>)*; deadnaming*; deplatform*; deplatforming*; enby; fat-shame*; fat-shaming (<i>n.</i> & <i>adj.</i>)*; kink-shame*; kink-shamer*; kink-shaming (<i>n.</i> & <i>adj.</i>)*; Latinx*; lumbersexual; mansplainer*; period cup*; period poverty*; sportswash*; sportswashing*; their; them; themselves; they; tone police*; trigger (<i>v.</i>); unvaxxed*; vaxxer*; -washing (<i>comb. forms</i> : brownwashing*, malewashing*, straightwashing*); wokeism*; wokery (<i>n.</i> ²)*; [womxn*]; Xennial (<i>n.</i> & <i>adj.</i>)*
Current Events	Brexitteer*; Brexiter*; C-19*; Covid*; CV-19*; dalgona coffee*; leaver; long Covid*; remainder (<i>n.</i> ²); Remoaner*; shrinkflation*; social distance*
Miscellaneous Other	courtsiding*; cut (<i>n.</i> ²); endgame, (<i>adj.</i>); forever chemical*; gegenpress*; gegenpressing*; glitter bombing*; glitter-bomb*; headcanon*; MC; mic drop; mic drop; [OCD, (<i>adj.</i>)]; over-ears; sharenting; stealthing; ticket splitting; water cremation*

In contrast to news media, Twitter/X imposes no editorial standard and has little enforcement of etiquette, linguistic or otherwise, substantially broadening the sward of new and nonstandard vocabulary being 'published' at any time; and with the lead time to publication compressed virtually to zero, even simultaneously produced new usages will be diffused on Twitter/X days, weeks, or months ahead of traditional media, including online media.

For this reason, Twitter/X may be especially important for early captures of newly formed regional, restricted, and nonstandard words, which in most cases are less likely to occur in the register of traditional low-lead-time media publications. This is observable in the OED record, as the tables above show, often in fairly dramatic ways. To briefly elaborate one example, the dictionary's documentation of Bermudian English since 2011, much of it published in a March 2021 update, added or revised seventeen words and senses. These revisions included eighteen Twitter citations, accounting for 38 per cent of the new citation evidence. Six were earliest attestations (35 per cent of senses), for ACEGIRL, BOASTY (sense 2), GANSEY/GANZEE (sense 4), MUG (*adj.*²), and WRINCH (*v.*², senses 1 and 2). In the latter two entries, six citations in total are given, five of which are tweets, with no print evidence referenced at all; presumably they would not exist in the dictionary today were it not for the platform.

The same principle of broad spontaneous language capture applies on a general scale, and is likely the primary factor in the disproportionate number of Twitter/X citations appearing in senses marked for region and/or register, as well as the large number of first attestations. However, other distinctive aspects of the platform have also contributed to the ‘new words’ phenomenon on Twitter/X, and the successful diffusion of many of these into the broader anglophone culture. These features include the functional affordances and limitations of the medium, the virtual concentration of geographically diffuse communities and cross-pollination between communities, the particular topical areas of trending interest among users, and an intangible Twitter/X ‘style.’

In addition to generating a vocabulary particular to itself (e.g., TWEET (*n.* & *v.*), RETWEET, TWITTERATI, TWITTERSPHERE), some Twitter/X ‘new words’ that originated as specific adaptations to the functional parameters of the platform have been extended to non-Twitter contexts, according to later evidence included in the OED entry—for instance, abbreviations (DM, WTAF), symbolic extensions (HASHTAG (*v.*) from #, AT (*v.*) from @), and gestural forms (THIS (*int.*), CHEF’S KISS (*n.*² & *adj.*)). Other Twitter/X coinages reference broader digital media contexts (DOX, GHOST (*v.*), SELFIE STICK, BUTT DIAL). In each case the OED entry includes a non-digital attestation, documenting the diffusion beyond social media and into the general culture.

Perhaps more unexpected, at least for speakers not habitually on Twitter/X, is the number of first attestations there of terms addressing recent broad social and societal developments. One clear example can be made of the OED’s rich Brexit lexicon, spanning the political and connotative range from BREXITER (2012) and BREXITEER (2013), to REMAINER (2015) and REMOANER (2016), all first attested on Twitter. This is no doubt a reflection of the general prevalence of this topic in all media in the United Kingdom during the period, and the tendency of Twitter/X users to respond textually to current events there as they happen. However, the channel between the general culture and social media goes both ways: while it is true that Twitter/X has been a quicker conduit to text publication than traditional news media, it is also the case that legacy media, online and in print, has played important roles in the formation and diffusion of words first attested on Twitter/X. This is due to the large number of active accounts on the platform which are themselves operated by traditional journalists and news organizations, or politicians and other government officials or offices (five of the seven most-followed Twitter/X accounts quoted in the OED fit these categories). That is, while many words formed to describe and comment upon topical matters occurred first on Twitter/X due to the general user sharing topical thoughts and opinions in real time (as BREXITER occurred in 2012), many were due to engagements between and among other, mainstream publication venues and influential individuals (as the more rhetorical BREXITEER was coined, by a Welsh Labour politician commenting on another politician’s opinion piece in the *Daily Telegraph*).

Other terms of broad social relevance are more directly the result of the peculiar nature of the medium, which forms new communities of exchange among individuals across vast reaches. The term LONG COVID might well not be in use today in media and among medical professionals had it not been tweeted out on 15 May 2020 by @elisaperego78,²⁴ the account belonging to Elisa Perego, an Italian archaeologist advocating for patients with multiphasic symptoms lasting well beyond what was understood to be the normal duration of the illness. For Perego and her interlocutors, Twitter represented a unique community in which physically isolated and geographically distant people could share personal experiences of a poorly understood and difficult-to-study condition;²⁵ and for the OED it represented a reliable archival record of the term’s emergence (compare contemporaneous and initially more popular ‘long-haul Covid’²⁶ or ‘long-term Covid’, neither of them yet documented in the OED). The pathway from Perego’s Twitter coinage, through other social media (chiefly a Facebook Group), into the news media and scientific literature, and on to widespread use,

is swift and well documented: by the Fall of 2020 it was in official use at the UK National Health Service and in testimony before Parliamentary committees (see [Callard and Perego 2021](#): §5).

True to the combative tendencies of ‘the discourse’, many Twitter/X first attestations resonate with contentious issues of the day, as for instance *WOKEISM*, first attested in 2015 on Twitter (the entry added, along with *WOKERY* (*n.*²), in 2023), and marked for register as ‘Chiefly *disparaging*’ by the dictionary. The entry for its etymon *WOKE* (*adj.*²) was published in 2017, when it was perhaps still possible to conflate the term’s various figurative shadings into one sense, meaning ‘in a state of awareness or vigilance’, going back to 1924. ‘Woke’ in the later and now current sense of ‘alert to racial or social discrimination and injustice’, however, in both its approving and disparaging contexts, thrived early on Twitter, as three quotations from 2009 and 2017 attest.

Many words in the ‘Social Discourse’ section of [Table 3](#) may come with ‘woke’ or ‘anti-woke’ associations, as they deal principally with questions of racial, sexual, and gender identity, closely tied to progressive politics in the English-speaking West. Among the most consequential of these, in terms both of general diffusion and political effect and backlash, was documented in stealth²⁷ sense additions applied in 2019 to each of *THEIR*, *THEM*, *THEMSELVES*, and *THEY*, as pronouns referring specifically to non-binary or genderqueer persons (singular ‘they’ more generally is another matter). In each case Twitter supplied the first attestation, in threads from 2009 involving users @genderbitch, @thebutchcaucus, @pieskiis, and @FireboltX.

Under *THEY* (I.2.c) and *THEM* (I.4.c), the OED cites as its first attestation a tweet by @thebutchcaucus:

As the preliminary ‘RT’ indicates, @thebutchcaucus’s retweet, and thus the lexical suggestion it makes, was originally composed by an account named @pieskiis, which has since been deleted by the user. Thus a significant element of the original conversation and its contexts is missing from the archival record, presumably since well before an OED researcher acquired the evidence, in October 2019. However, by consulting the yet-undeleted tweet archive of the third mentioned account, a reader can infer that @pieskiis’s quoted tweet is in response to @FireboltX’s earlier musings over possible pronouns to adopt:

Surrounding tweets document a rich metalexical exchange undertaken over two or three days amongst members of a dispersed virtual community sharing tentative self-referencing



11 July 2009 retweet by @thebutchcaucus (Internet Archive instance) <<https://web.archive.org/web/20191010090725/https://twitter.com/thebutchcaucus/status/2584598369>> .

practices. Not only is the earliest textual instance of these terms pinpointed (to within about fifteen hours, as @pieskiis’s original tweet is lost); also captured are a number of alternative options and assessments of their relative strengths and weaknesses, the users’ motivating considerations, and their relevant social contexts, including @FireboltX’s surmise that

‘Especially in a country like India, I am probably the only one using them’ (@FireboltX, 11 July 2009).²⁸



10 July 2009 tweet by @FireboltX, on x.com <<https://x.com/FireboltX/status/2578509022>> (accessed 30 April 2025); no archiving available.

In 2009 ‘pronouns’ were beginning to occur all across Twitter in ways that seemed new to many users, but which the OED has traced to 1977. In this one case it appears that the OED may be *under-citing* Twitter in its documentation of PRONOUN, sense 2 (added in 2021), which leaves a gap of twenty-one years between a somewhat old-fashioned quotation of ‘preferred pronouns’ in 1998 and a highly self-aware quotation from the 2019 *Guide for Supporting Trans & Gender Diverse Students*. While Twitter is cited as the most recent evidence, with a quotation from a somewhat famous tweet by @TheElliotPage (1 December 2020, 1.4 M likes, 256 K retweets/reposts, 32 K replies),²⁹ the transition and diffusion period from restricted to general currency, well documented in real time in the Twitter archive (even if it may have occurred in other quarters as well), is absent from the OED record. ‘My pronouns’ appears to be used in this way by the early summer of 2009, roughly coinciding with the emergence of ‘they/them/theirs’, and ‘preferred pronouns’ unambiguously in September. By the early 2010s these and related phrases had become common topics of ‘the discourse.’ In June 2014 the phrase ‘pronouns in bio’ appeared for the first time in a tweet (as far as has been preserved), and quickly thereafter the practice of listing pronouns—in the ‘they/them’ style anticipated by @pieskiis—became routine among various demographics.

For all its technical innovations and incorporations over its twenty-odd years of life, Twitter/X has remained fundamentally a textual medium. While clearly many of the important societal developments of the century have perceptible, attestable reverberations in the extant tweet record, the medium per se also motivates an innovative textual impulse that can only be called ‘style.’ This is marked by an ironically jocular, linguistically self-aware affectation of casualness, as in the ironically avoidant ‘Hung out with these three chinesians for part of the afternoon. they were making fun of my american-ness’ (CHINESIAN; @mariaedavey, 17 April 2010); or the faux dainty ‘Ooh, just cuz you’re speaking all proper and gentlewoman-like doesn’t make you any less evil’ (GENTLEWOMANLIKE; @Phase2Official, 25 August 2012); or archly morphological ‘mathematical, aerodynamical, scientific and witchcraftical calculations’ (WITCHCRAFTICAL; @harry_maps, 18 March 2020).³⁰

These examples are not in fact first attestations in the OED, though they may as well be neologisms, given the gap of more than three centuries between them and the next-most recent OED attestation. No doubt coined independently of their precedents (respectively, Thomas Heywood, 1607; Richard Burthogge, 1675; and *The Doctrine of Devils*, 1616), these self-aware turns of phrase are to be distinguished from such apparently unstudied independent re-coinages as CRITIC (*v.*), sense 2 (*trans.*) (‘Why do I feel like you always critticking my shit?’, @_KennedyJohnsn, 12 March 2013), previously attested in Alexander Pope, 1744); TITULE (*v.*), apparently a keypad mishap but matching a form of TITLE (*v.*) documented prior to the eighteenth century (‘There’s a bootleg of Waitress on YouTube literally titled

“I’m Sad and Pregnant the Musical” [laughing and crying emojis], @RussellSondheim, 23 August 2017); and LUNATICNESS (“There’s someone out there who just out off lunaticness betted 5-0 against brazil’, @nab_d21aau, 8 July 2014), previously attested in Nathan Bailey, 1727; as much as from the jocularly erudite allusion in ‘At least you didn’t call me a whore-masterly rogue...’ (WHOREMASTERLY; @ewarthale, 9 August 2013), the reference possibly to Shakespeare’s ‘whore-masterly villaine’ (*Troilus*), or else more accurately to an obscure seventeenth century court proceeding with this precise collocation.³¹ CHINESIAN, GENTLEWOMANLIKE, and WITCHCRAFTICAL belong to a linguistically flexible and inventive Twitter/X style, that also produced first instances of AMAZEBALLS (*int.*), OMNISHAMBOLIC, and SHITHOUSERY, each a playful derivation of recently coined or current lexis.

Conclusion

Given that, among the trillions of words produced on Twitter/X since 2006, a lexicographer might well find there an easy recent example of virtually *every* word, it is worth considering what policy limits should restrain the use of citations from this source. Uncontestably, Twitter/X should generally be quoted when it supplies the earliest attestation of a term, or the only contemporary attestation that can be found. But even this broad principle has limits: ‘dotard’ is not now in current use, however amusing its brief 2017 revival may have been; and a great number of spontaneous coinages on Twitter/X will fail to take root in the language as a whole. The spontaneity of the medium may be one of its great advantages to the lexicographer, but it comes with pitfalls: is a single tweet employing ‘lunaticness’, or ‘critic’ as a transitive verb, or ‘titled’ for ‘titled’, a notable lexicological occurrence, or is it a forgettable solecism among the millions that occur in spontaneous speech or typed text every day, saved here on the sole basis of a preexisting archaic usage? Further corroboration of contemporary usage would we welcome in deciding.

There is a more general point to be made however, which while not strictly lexicological has important lexicological and lexicographical implications. A handful of individual sources are quite clearly overprominent in the OED: for example, Shakespeare is quoted in almost 15 per cent of OED senses with evidence from the time of his flourishing; and the *Times* appears in 8 per cent of all senses with attestations after 1900. Although it would take several more decades of quoting X at current rates to arrive at comparable prevalence, the trend is clearly in that direction: already 3.7 per cent of all OED senses that include any post-2006 evidence cite Twitter/X; of those added or revised in 2025, 7.7 per cent do. In the categories of words for which Twitter/X has been particularly useful to OED lexicographers, the percentages are naturally even higher: 8.0 per cent of all senses that are marked for register and include evidence from 2006 or later cite Twitter/X evidence (10.6 per cent of the senses in the 2025 batch of revisions); and 9.5 per cent of senses marked as regional (12.5 per cent in the 2025 batch).

While it is certainly the case that Twitter/X is a paradigm of heteroglossia—it is arguably not one OED source but 5,200-odd individual sources, and counting—still overreliance on any format or medium can produce unintended effects on both the construction of the lexicographical record and its reception. These effects might be archival, given the problem of tweet decay; or tonal, to the degree that Twitter/X conditions certain styles and attitudes; or perceptual, exaggerating the cultural importance of the platform beyond that which it in fact has, or suggesting that certain varieties of English occur mainly in social media rather than in higher status publication formats. Ours is an age of staggering textual surplus, in which enormous numbers of works of all variety are published every year. We may hope that this will be reflected for posterity in the OED record of our English.

Notes

- 1 @WHO, 2020.02.11 15:09+0, <<https://web.archive.org/web/20200211154640/https://twitter.com/WHO/status/1227248333871173632>>.The reference is to the announcement reported in World Health Organization 2020.

- 2 @DrNeilStone, 2020.02.11 17:11+0, <<https://web.archive.org/web/20200211201831/https://twitter.com/DrNeilStone/status/1227278889103372293>>.
- 3 Besides the Covid-related entries COVID-19 (added April 2020), C-19, CV-19, and COVID (all July 2020), since 2000 OED has only published new senses of *ONESIE* (2b, 2004) and *STATE OF THE STATE* (2012) in the same year as their first evidence. For examples of the few words and senses documented effectively upon coinage (i.e. published in the same year as the earliest evidence) in the First Edition of OED (1888–1928), see Gilliver 2016: 218.
- 4 OED editors held a webinar on 10 September 2020 discussing these circumstances (McPherson 2020). A parallel account of the emergency lexicography of the period, with a focus on work at Merriam-Webster, is given in Fatsis 2021.
- 5 By revisions and additions, in general I mean all changes to the citation evidence in the dictionary, regardless of the way an entry is marked in *OED Online*. Many revisions, additions, and even deletions of evidence, even within newly added senses, do not result in a change of the stated revision status in *OED Online*, which I discuss below in relation to *TWEET* (*n.* & *v.*), and *THEIR*, *THEM*, *THEMSELVES*, and *THEY*. Likewise entries marked as ‘new’ can be, variously, new (i.e. for a word or sense not previously documented), or revised (when the revising editor judges that senses of a word should be separated into independent entries), or even unrevised (as with over 100,000 of the 260,000-plus derivative and compound sub-entries algorithmically reformatted into stand-alone entries, without further revision, in mid-2023).
- 6 All OED data is published by Oxford University Press. In this article I analyze and compare data using several full-text XML instances of the OED published between 2014 and 2026, made available under license by Oxford University Press. As of 1 January 2026, the OED tallies 5,213 citations of Twitter/X, illustrating 5,017 senses under 4,145 headwords. Twitter/X supplies the earliest evidence for 213 new word meanings, of which 114 represent newly documented words (i.e. the first citation in the entry), and 99 are later new sense developments of existing words. A handful of Twitter citations previously in the dictionary have been suppressed.
- 7 Although OED of course does not regard itself as performing such a legitimating function, nevertheless countless headlines in print and online news sources can be cited proclaiming variations of ‘“Tweet” is a real word (in the social media sense): Oxford’ (Dingman 2013).
- 8 This instance of OED is not preserved, but Neville Hobson’s documentation of his own brief lexicological recognition is (Hobson 2013).
- 9 As of the present writing there persists an inconsistency in OED’s attributions to Twitter/X, with a number of posts dated after July 23, 2023 labelled ‘twitter.com’, and a number of earlier tweets labelled ‘x.com’. Presumably a future policy will regularize these attributions.
- 10 In Figure 2, Tables 1 and 2, and the accompanying discussion, I have parsed and compared complete XML code for OED instances published in May 2017, December 2018, December 2019, March and December 2021, March 2023, December 2024, and December 2025. For quotations appearing between May 2017 and December 2018, where an addition date could not be otherwise ascertained via the publication history, the year of addition was assigned to either 2017 or 2018 proportionately at random, introducing a small amount of statistical uncertainty in the final numbers reported here. To give an indication of the effect size, over 1,000 simulations of this allocation process, 95% of tallies for total new Twitter/X quotations added between January 2018 and December 2025 fell between 5,101 and 5,112, with a mean of 5,107 (± 0.16 per cent).
- 11 In mid-2023 OED moved most sub-entries for derivatives and compounds into new stand-alone entries, creating over 262,000 new entries (nearly doubling the number of entries in the dictionary). For this reason I only include data for sense units (and not for entries per se, unless an entry has only one sense) in the historical comparison.

- 12 The statistical cooccurrence of these labels across register and region is not meaningfully significant.
- 13 Since this was published, an ‘edit’ feature has been introduced, with limitations, on [x.com](#).
- 14 According to this study, where an author can be identified, the (June 2022 instance of) OED quotes women about 30 per cent of the time in evidence drawn from the most recent 25 years (i.e. from works published since 1997), whereas, e.g., the Hathi Trust corpus of texts published in that period is about 40 per cent female authored. See the graphs in [Williams 2023](#), p. 739, and surrounding discussion.
- 15 The demographics of the underlying corpus are not easy to estimate, but most recent attempts have indicated a strong male bias in the population, suggesting that OED has quoted proportionally more women from this corpus than would be expected in a random sample. Commercial demographic studies of Twitter/X published on the web by Statista have reported a 64 per cent male global usership as recently as October 2025 ([Statista 2025](#)). The most recent Pew social media factsheet ([Pew 2025](#)) estimates that among American adults, 25 per cent of men had ever used Twitter/X, compared to 16 per cent of women, implying something like 60 per cent male usership, at least among U.S. users (and across all time). Earlier peer reviewed studies independently reported the male percentage of usership variously as 45 per cent ([Sloan et al 2013](#)), 50 per cent ([Chen et al. 2015](#)), 73 per cent ([Yildiz 2017](#)), and 47 per cent ([Vikatos et al. 2017](#)).
- 16 For the purposes of these calculations I have amalgamated a number of highly granular categories. For instance, OED marks out four varieties of ‘Scottish English’ and ten of ‘Regional English’ English. Of the six kinds of ‘U.S. English’ OED differentiates, I keep African American English and Hawaiian English, and amalgamate the rest into the parent category.
- 17 These data may be read against conclusion elsewhere that ‘African American English is the main source of lexical innovation on American Twitter’ (Grieve et al. 2018: 314)
- 18 For Twitter/X citations, including the URL of the post in the citation, whether in text form or as a link (or both), would go a long way towards facilitating follow-up; archiving tweets publicly rather than privately would mitigate the problem of tweet decay (Internet Archive, for example, preserves many older tweets, though for some time now it has failed to archive [x.com](#) posts); and linking to the archive instance rather than [x.com](#) would solve both problems.
- 19 Over the five decades of editing the First Edition of the OED, about 2,550 items of ‘Mod.’ (i.e., fabricated) evidence were added, with the highest concentration in volumes edited by James Murray (between 0.2 and 0.4 per cent of citations, with the exception of Volume I, A–B (1888), which reached 0.6 per cent).
- 20 See news of the period, including [Berlinger 2019](#).
- 21 <<https://x.com/mdufay/status/1601909745618665477>> (accessed 31 December 2025; no archive available).
- 22 Most of the literature investigating automated activity and its detection on Twitter/X precedes both the recent revolution in generative artificial intelligence and the purported purge of bot accounts on X following the 2022 acquisition, as well as the general introduction of the @Grok X account in 2024. A sense of the challenge may be gleaned from [Javed et al. 2024](#). A recent small case study found that ‘X is flooded with platform manipulation of various kinds’ including coordinated bot networks ([Graham and FitzGerald 2023](#)).
- 23 Potentially one of the tweets cited in ‘Discourse (Slang)’ in ‘Know Your Meme’ (website) <<https://web.archive.org/web/20250411040929/https://knowyourmeme.com/memes/discourse-slang>>. See tweets by @glassbeachband (25 June 2016) <<https://x.com/glassbeachband/status/746821980964302848>> (accessed 30 April 2025; not currently accessible, with no archive available); @KrangTNelson (27 November 2017) <<https://>

web.archive.org/web/20200930095159/https://twitter.com/KrangTNelson/status/935189118736371715>.

- 24 The OED citation is misdated to 20 May of that year (perhaps the day was confused with the year), in an example of a type of error which, although relatively minor, is all too common in the documentation of tweets in the OED, exacerbating the problem of traceability. The original tweet is preserved at <<https://web.archive.org/web/20200515195804/https://twitter.com/elisaperego78/status/1261345351820365824>>.
- 25 Perego’s voluminous Twitter archive, which remains accessible at x.com as of this writing, may be read against the account in [Callard and Perego 2021](#)).
- 26 ‘Long haul Covid’ and ‘long haulers’ appear to derive from the Facebook Group ‘Long Haul Covid Fighters’, created in late April or early May 2020 (<<https://web.archive.org/web/20200628190546/https://twitter.com/katemereditihp/status/1277316840453267456>>), and documented on Twitter as early as 25 May (<<https://web.archive.org/web/20200525100815/https://twitter.com/Zoe00824621/status/1264858601421758471>>). These terms begins to appear in the news media by the end of June, initially in reference to the Facebook Group.
- 27 In an example of revisions occurring without a corresponding change to revision status, as of the current writing these entries continue to show a revision date of 2013, despite the addition of the senses in question in 2019 (the inclusion of a later ‘last modified’ date does not address this discrepancy, as ‘modifications’ can refer to a range of peripheral changes, including sitewide changes).
- 28 <<https://x.com/FireboltX/status/2588090839>> (accessed 31 December 2025; no archive available).
- 29 @TheElliotPage (1 December 2020) <<https://web.archive.org/web/20201201172159/https://twitter.com/TheElliotPage/status/1333820783655837701>>.
- 30 <<https://x.com/mariaedavey/status/12366592353>> (accessed 31 December 2025; no archive available); <<https://x.com/Phase2Official/status/239430483304402945>> (accessed 31 December 2025; no archive available); <https://web.archive.org/web/20200318144546/https://twitter.com/harry_maps/status/1240269032051392513>.
- 31 <https://x.com/_KennedyJohnsn/status/311346568202698752>, <<https://x.com/RussellSondheim/status/900550926234202114>>, <https://x.com/nab_d21aau/status/486614703268179969>, (all accessed 31 December 2025; no archive available). The user @ewarthale has deleted all of their posts on x.com.

References

A. Dictionaries

- Merriam-Webster. 2025. *Merriam-Webster.com Dictionary* <<https://www.merriam-webster.com/dictionary/>>.
- Proffitt, M. *et al.* (eds.). 2000–. *Oxford English Dictionary*, Third Edition. Oxford: Oxford University Press, 1888–2025; instances from May 2017, December 2018, December 2019, March and December 2021, March 2023, December 2024, and December 2025.

B. Other literature

- Allen, A. 2018. ‘Youth Words.’ *OED.com* (19 September). <<https://web.archive.org/web/20180919113026/https://public.oed.com/appeals/youth-words/>>.
- Berlinger, J. 2019. ‘North Korea revives “dotard” insult after Trump’s “Rocket Man” comment “displeased” Kim Jong Un.’ *CNN.com* (4 December). <<https://web.archive.org/web/20191205051005/https://www.cnn.com/2019/12/04/asia/north-korea-kim-jong-un-intl-hnk/index.html>>.
- Callard, F. and E. Perego. 2021. ‘How and why patients made Long Covid’ in *Social Science and Medicine* 268. <<https://doi.org/10.1016/j.socscimed.2020.113426>>.
- Chen, X. *et al.* 2015. ‘A Comparative Study of Demographic Attribute Inference in Twitter.’ *Proceedings of the Ninth International AAAI Conference on Web and Social Media*. 590–93. <<https://doi.org/10.1609/icwsm.v9i1.14656>>.

- Considine, J. 2009. 'Literary Classics in OED Quotation Evidence.' *Review of English Studies* 60.246. 620–38.
- Dent, J. 2019. 'No tears of joy (yet): emoji make their OED debut.' OED Blog (28 June). <<https://web.archive.org/web/20230402030738/https://public.oed.com/blog/no-tears-of-joy-yet-emoji-make-their-oed-debut/>>.
- Dingman, S. 2013. "'Tweet" is a real word (in the social media sense): Oxford.' *The Globe & Mail* (Toronto) (17 June). <<https://web.archive.org/web/20130617152834/https://www.theglobeandmail.com/technology/digital-culture/tweet-is-a-real-word-in-the-social-media-sense-oxford/article12595953/>>.
- Fatsis, S. 2021. 'Thirty-Four Days: Inside Merriam-Webster's Emergency Coronavirus Update.' *Dictionaries: Journal of the Dictionary Society of North America* 42.2. 45–56.
- Gibson, K. 2018. 'Word where you are' OED.com (25 April) <<https://web.archive.org/web/20191003151548/https://public.oed.com/appeals/words-where-you-are/>>.
- Gilliver, P. 2016. *The Making of the Oxford English Dictionary*. Oxford: Oxford University Press, 2016.
- Graham, T. and K. M. FitzGerald. 2023. Bots, Fake News and Election Conspiracies: Disinformation during the Republican Primary Debate and the Trump Interview. *Digital Media Research Center*, Queensland University of Technology. <<https://doi.org/10.5204/rep.eprints.242533>>.
- Hobson, N. 2013. 'Tweet catches on in the OED.' *nevillehobson.com* (15 June). <<https://web.archive.org/web/20130623033608/https://nevillehobson.com/2013/06/15/tweet-catches-on-in-the-oed/>>.
- Javed, D. et al. 2024. 'Towards the future of bot detection: A comprehensive taxonomical review and challenges on Twitter/X.' *Computer Networks* 254. <<https://doi.org/10.1016/j.comnet.2024.110808>>.
- McPherson, F. et al. 2020. 'The language of Covid-19: a special OED update' (10 September 2020). YouTube (22 September) <<https://youtu.be/IMpiCIEOOcU>> and OED.com. <<https://web.archive.org/web/20230701000000/https://www.oed.com/information/using-the-oed/oed-event-recordings/language-updates-and-words-of-special-interest/the-language-of-covid-19-a-special-oed-update/>>.
- Pew Research Center. 2020. 'Differences in How Democrats and Republicans Behave on Twitter' (15 October). <<https://web.archive.org/web/20201015143822/https://www.pewresearch.org/politics/2020/10/15/differences-in-how-democrats-and-republicans-behave-on-twitter/>>.
- Pew Research Center. 2024. 'How Americans Navigate Politics on TikTok, X, Facebook and Instagram' (12 June). <<https://web.archive.org/web/20250129082052/https://www.pewresearch.org/internet/2024/06/12/how-americans-navigate-politics-on-tiktok-x-facebook-and-instagram/>>.
- Pew Research Center. 2025. 'Social Media Fact Sheet' (20 November). <<https://www.pewresearch.org/internet/fact-sheet/social-media/>>.
- Sheidlower, J. 2011. 'How Quotation Paragraphs in Historical Dictionaries Work: *The Oxford English Dictionary*.' M. Adams and A. Curzan (eds.), *Contours of English and English Language Studies*. Ann Arbor: University of Michigan Press. 191–212.
- Simpson, J. 2000. 'Preface to The Third Edition of the OED.' OED.com. <<https://web.archive.org/web/20230724152209/https://www.oed.com/information/about-the-oed/history-of-the-oed/preface-to-the-third-edition-of-the-oed/>>.
- Simpson, J. 2013. 'A heads up for the June 2013 OED release.' OED.com (15 June). <<https://web.archive.org/web/20130617014536/http://public.oed.com/the-oed-today/recent-updates-to-the-oed/june-2013-update/a-heads-up-for-the-june-2013-oed-release/>>.
- Sloan, L. et al. 2013. 'Knowing the Tweeters: Deriving Sociologically Relevant Demographics from Twitter.' *British Sociological Association* 18.3. 74–84.
- Statista. 2025. 'Distribution of X (formerly Twitter) users worldwide as of October 2025, by gender' (25 November). <<https://web.archive.org/web/20251221003917/https://www.statista.com/statistics/828092/distribution-of-users-on-twitter-worldwide-gender/>>.
- Vikatos, P. et al. 2017. 'Linguistic Diversities of Demographic Groups in Twitter.' *Proceedings of HT³17*. 275–84. <<https://doi.org/10.1145/3078714.3078742>>.
- Williams, D.-A. 2021. "'Alien" vs. Editor: World English in the *Oxford English Dictionary*, Policies, Practices, and Outcomes 1884–2020. *International Journal of Lexicography* 34.1. 39–65.
- Williams, D.-A. 2023. 'Women's Words and the Words of Women in the *Oxford English Dictionary*.' *Review of English Studies* 74.316. 730–51.
- World Health Organization. 'Situation Report 22' (2020). <<https://web.archive.org/web/20200213033639/https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf>>.
- Yildiz, D. et al. 2017. 'Using Twitter data for demographic research.' *Demographic Research* 37. 1477–1514.