

Article

'ALIEN' VS. EDITOR: WORLD ENGLISH IN THE *OXFORD ENGLISH DICTIONARY*, POLICIES, PRACTICES, AND OUTCOMES 1884–2020

David-Antoine Williams 

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

Abstract

This article discusses the changing ways in which the *Oxford English Dictionary* has recorded the vocabularies of 'World English'—English as spoken outside of the British Isles—from the first to the present edition. Based on direct analyses of the coded text of multiple editions, it documents and compares the practices of successive editors, taking into account various contextual factors, such as editorial principles and policies, institutional resources, and historical language development. Significant attention is given to labelling practices, including the notorious 'tramline' mark of the First Edition and Second Supplement, designating 'alien' vocabulary; the evolution of the notion of 'regional' English within the dictionary; and the contributions of technology to the art of lexicography. The final section details changes in policy and methods in the current revision and expansion, evaluating both its practices vis-à-vis its predecessors, and the picture it gives us of the current state of World English.

Keywords: dialect, lexical borrowing, digital dictionary, online dictionary, etymology, historical lexicography, Afrikaans, Australian languages, American languages, English

1. Introduction

Yogurt was a foodstuff sufficiently appreciated in the Anglosphere in the early decades of the twentieth century that a fascicle of the *New English Dictionary* (OED1), published in 1921, could describe it as 'common in many English-speaking countries'. Incongruously, the word *yogurt* was apparently sufficiently exotic to have been deemed 'alien or not fully naturalized' by that volume's editors, with the result that the lemma ||YOGURT appeared with the prefatory upright parallel bars normally designating such terms. Sometimes called 'tramlines' (or 'tram-lines') by OED editorial staff and later commentators, the use of this

symbol has a spotty history. While James Murray and his fellow OED1 editors thought the inclusion of such ‘alien’ words significant enough not only to mark, but also to count up and discuss in the prefaces to each volume (along with, perhaps for analogous reasons, words marked ‘obsolete’), when it came to publishing the First Supplement (SUP1) in 1933, William Craigie and C. T. Onions, who had also edited volumes of OED1 (Onions was responsible for ||YOGURT), abandoned the system as too subjective and impracticable (Ogilvie 2013: 161).¹ Robert Burchfield reinstated tramlines for new and revised entries in the Second Supplement (SUP2, four volumes 1972–1986)—though he un-tramlined (and demajusculed) YOGURT and a number of other lemmas now deemed naturalized. Burchfield’s decisions carried through to the amalgamated 1989 Second Edition of the OED (OED2), though the editors of that volume did not themselves apply the mark to the relatively small amount of new material they added. In the ongoing wholesale revision and expansion project known officially as the Third Edition, or *OED Online* (OED3), tramlines have been removed from all entries. Use restrictions are now indicated via ‘Category’ labels, including subject (knowledge domain) and regional designations.

What does it mean to include a word in an English dictionary—let alone the ‘definitive record of the English language’, as, perhaps not unjustifiably, OED now bills itself (oed.com)—while at the same time labelling it ‘not fully naturalized’? Beginning with this question, in this article I give a history of OED’s policies and practices regarding those English words in use predominantly outside the British Isles, to think about what ‘the English language’ has meant in Oxford over more than 130 years of evolving language, and an evolving English dictionary, and the tensions this has produced between notions of ‘definitiveness’ and ‘regionality’. The account given here is largely based on analysis of data gathered with custom programs (written in the Python programming language) from the background code of various editions of the Dictionary, allowing for a more detailed and more accurate picture than that which can be derived from the web-based user interface of *OED Online*, or the prototype API currently being tested.² I have therefore included, in addition to conventional endnotes, a series of ‘data notes’ to explain in greater technical detail the data and methods behind the figures presented in the main text.^{D1}

While the picture given in what follows may be of a higher resolution than the adumbrations of previous writers on the topic, it is not my aim here to pursue an independent critical evaluation of OED’s treatment of any particular variety of regional English—much less of *all* varieties—as some have attempted on a case-study basis (e.g., Benson 2001, Ogilvie 2013). To do so would engage multiple additional dimensions of argument, cross-evaluating the outcomes of dozens of regional supplementary and general dictionaries disparate in aim, principles, method, resources, era, and linguistic and social context. Instead, although I contextualize the OED record with reference to relevant other dictionaries where appropriate, I focus mainly on the internal development of OED’s theories and practices, and the results of these for the dictionary text.

2. The trouble with tramlines in OED1

Murray, who taught himself more than twenty languages—each one ‘a new delight no matter what it was, Hebrew or Tongan, Russian or Caffre’ (quoted in Murray 1977: 32)—appreciated that ‘English’ could refer multiply to ‘the English of Scotland and of Ireland, the speech of British Englishmen, and American Englishmen, of Australian Englishmen,

South African Englishmen, and of the Englishmen in India' (quoted in Murray 1977: 193). However, though he understood well enough that English was a various language spoken all across the world (Price 2003: 120), as Edmund Weiner notes, he had no sense that, rather than mere regional variants, these ways of speaking might constitute multiple regional standard Englishes in their own right (Weiner 1987: 31). Thus OED1's representation of English as a 'language of the world [...] naturally took on a form corresponding to the structure of the language and its central and peripheral aspects as they were understood at the height of the British Empire' (Benson 2001: 101).

One aspect of this centric understanding is manifested in Murray's 'General Explanations' to OED1, which outlined his theory of a 'circle of the English language', in which 'foreign' and 'dialectal' usages pointed outwards (along with 'scientific', 'technical', and 'slang') away from the 'common' core (Murray 1888: xvii), with the implication that Standard British English (St BrE) 'is, in fact, the standard version of the language wherever it is spoken and that therefore St BrE and the common core of the language are virtually the same thing' (Weiner 1987: 31). Words in the dictionary were further to be classified according to a vague paradigm of 'citizenship': *Naturals* would include 'all *native* words like *father*, and all fully *naturalized* words like *street*'; *Denizens* those 'naturalized as to use, but not as to *form*, *inflection*, or *pronunciation*'; *Aliens*, the 'names of foreign objects [...] which we require often to use, and for which we have no native equivalents' and *Casuals*, 'foreign words [...] not in habitual use' (Murray 1888: xix). The tramline symbol was explained in the same section as designating 'non-naturalized or partially naturalized' words, which is to say '*Denizens* and *Aliens*, and such *Casuals* as approach, or formerly approached, the positions of these' (Murray 1888: xix).

Following Murray's explanations, when OED's employment of tramlines has received scholarly attention it has generally been within discussions of non-British English. The entry for 'tramlines' in the index to Gilliver 2016 asks us to 'see foreign words and phrases, naturalization of' (620), which leads to a short section on Murray's determinations of various words' 'Anglicity' (68–70), where the examples have French, Spanish, Greek, Singhalese, and Hindi etymologies. Curzan discusses the symbol as 'a critical factor in the OED's overall ability to legitimize loanwords' with examples coming into English from Hindi, Dutch/German, Afrikaans, Hebrew, and Tagalog (Curzan 2000: 107–108). The only mention of tramlines in Brewer 2007 occurs within a short section on 'World English' (197–200). And in Ogilvie 2013, a monograph on 'World Englishes' and the OED which discusses them on more than forty pages, tramlines are mainly considered in relation to specifically non-British, and especially non-European, contexts, in accordance with her focus.³

But there exists a very long list of tramlined OED1 lemmas that do not fit easily within these categories of foreign, non-British, or only partially legitimized vocabulary. It includes, to choose haphazardly, words such as ||AMNESIA, ||SYNECDOCHE, ||TUBERCULOSIS, and ||VAGINA. Yes, ||VAGINA, though it was updated in 1986 by Burchfield (to include references to the *vagina dentata*), was not absolved of its 'alien' status, despite the availability of textual evidence of regular, if not widespread, popular usage,⁴ and OED1's own recording of 'naturalized' English derivatives, such as VAGINAL. The reason, as may be inferred from the other examples adduced above, is that from the beginning tramlines were applied to much of the technical and scientific lexis derived from Classical languages, including prominently the specialized vocabularies of law, grammar, and rhetoric; and, especially, of the various pure and applied scientific disciplines.⁵

If *vagina* and *amnesia* came into English as learned loanwords, and then passed into the popular idiom unacknowledged by the OED, the historical and socio-cultural processes that drove this adoption bear little relation to those that led to the popular loanword ||NULLA-NULLA (the etymology given perfunctorily as ‘Native Australian’), to borrow one of Ogilvie’s early examples (2013: 18). In fact, three categories of tramlined word in OED1 ought to be differentiated: (1) true popular loanwords adopted from contact languages of the Anglosphere, often described as characterizing a ‘World English’ or multiple ‘World Englishes’;⁶ (2) domain-specific vocabulary developed in other languages and imported quasi-contemporaneously into English (terms specific to cultural practices, e.g.); and (3) borrowings from archaic languages and neologisms formed in English from archaic roots for technical and scientific purposes. A word of any of these types may undergo processes of ‘naturalization’ to a greater or lesser degree, whether within or outside the original usage communities; and there are overlaps between the types—Wissenschaft vocabulary formed from Latin and Greek within the German academy in the nineteenth century and then borrowed into English (types 2 and 3), for example, or some cookery or folk-art terms (types 1 and 2). Only type 1, however, can in principle be called ‘non-British English’ or ‘World English’ vocabulary.

With this distinction in mind it soon becomes clear that tramlines per se serve as a poor marker for isolating World English lemmas, since tramlined words are predominantly of the third (scientific and technical) sort, with the second (knowledge domains developed in other cultures and transmitted into British English) next most prominent, and true World borrowings the least common (Table 1).^{D2}

There are 9,215 tramlined lemmas in OED1, of which 4,992 (54%) contain a Latin or Greek etymology (or both), with French, Italian, and Spanish being the next most common. Only 1,694 tramlined lemmas (18%) have non-Western-European etymologies, i.e. with at least one donor language that is not Germanic (including English), Latin (including Romance), or Greek. That proportion falls again, to 12%, when only non-Indo-European languages are tallied. Only 5%, 475 in total, are explicitly designated as representing regional or ethnolinguistic usages. On the other hand, two-fifths (39%) of all tramlined entries refer somewhere to one or more knowledge domain labels, with three-quarters of these belonging to the technical vocabularies of math, medicine, and the pure and applied sciences—the most common being *Bot.* or the equivalent (in 693 entries), *Path.* (543),

Table 1. Common labels in OED1 tramlined entries

Rank	Etymology	Domain	Region
1	Latin (4,409)	No Domain (5,656)	No Region (8,751)
2	Greek (2,232)	Science & Math (2,635)	South Asia (189)
3	French (1,702)	Academic (450)	Europe (68)
4	Italian (594)	Arts (331)	North America (66)
5	Spanish (427)	Politics (140)	Britain and Ireland (51)
6	Hindi-Urdu (295)	Religion (62)	Africa (34)
7	Arabic (263)	Industry (39)	East Asia (31)
8	Persian (180)	Ritual (32)	Australasia (22)
9	German (170)	Leisure (26)	South America (1)
10	Dutch (119)	Food (18)	

Zool. (488), and *Anat.* (335). Tramlines in OED1, therefore, must be understood to mark out primarily the abstruse vocabulary of learned Western scientific discourse, and secondarily the vocabularies of connoisseurs and aficionados of various kinds. Only a fraction represents the lexis of non-British or World Englishes, or indeed of popular loanwords more broadly.

Not only are tramlined entries in OED1 mostly not World English words, significant numbers of World English words don't carry tramlines. If one wanted to investigate OED1's treatment of English from the Indian subcontinent, for instance, one could isolate 622 entries containing an etymology indigenous to the region (i.e. within either Indo-Aryan or Dravidian language groups), of which only 70% (438) would be marked out with tramlines. Though the remaining 184 include some 'naturalized' words (e.g., ALOE, BANDANNA, RUPEE, YOGA), many more are (or were) arguably still 'alien', or regionally restricted, according to OED1's standards (e.g., CHAMAR, DHOBI, DHURRIE, and LUNKAH—all first attested after 1880).

Even a gathering such as this, based on etymological origin, would exclude an important class of World English lexis, however: new words and senses, local to a region, but formed within or preserved from non-indigenous languages, most prevalently English. For an indication of this one must refer to OED1's regional usage labels—in the previous South Asian English example, *Indian*, *Anglo-Indian*, *S. Asian*, and so on, and cross-compare the etymology. These labels on their own would turn up 189 entries from the former set (i.e. of Indo-Aryan and Dravidian etymologies; 149 of these bear tramlines), plus 123 additional entries of various etymologies (40 with tramlines, mostly from Persian and/or Arabic), including 87 words with European etymologies, 52 of which are, or include, English. These items of South Asian English recorded in OED1 include distinctive uses of otherwise 'common' English words, such as BOX, *n.*² (in *box-wallah*, a pedlar or a shop-keeper) or COUNTRY (attributively, to mean 'native to India, non-European', from 1582), and such regular constructions as WEIGHMENT ('The action of weighing [commodities]', 1878) and COLLECTORATE (an administrative district, 1825).

The significance of tramlines in OED1 is therefore dubious, not only as a lexicographical or linguistic marker, but also as an indicator of the attitudes of editors towards varieties of English spoken outside or imported into the United Kingdom. A better account of these questions would look first and primarily to a cross-comparison of OED's other labels and markers, bearing in mind that these were applied in various ways and with varying consistency over the Dictionary's development. Chief among these markers would be etymologies identifying words originating in or passing through languages other than English (or non-Germanic languages, or non-European, or non-Indo-European, or what have you), and usage labels indicating regionality irrespective of the origin language, capturing retained and extended local senses of English words, as well as regional transmissions from other primary contact languages (e.g., the languages of other historical colonial powers). Although such regional labels may sometimes apply to an entire entry (as for WEIGHMENT), they are just as likely to refer to only one or more senses or subsenses, or attributive or combined usages (as *box-wallah*), so in what follows I refer to senses when discussing regional labels, and entries, lemmas, or words when discussing etymology.⁷

Another major issue facing quantitative analysis, especially in a study comparing the practices of generations of different contributing editors, is the differing purpose, scope, and reach of each Edition, Supplement, and Additional Series. While it is not exactly

meaningless to say that 7% of SUP1 new entries retained in OED2 carried non-European etymologies, versus 1% of OED1 entries, per se it can hardly be taken to represent a divergence in attitudes towards World English, for the plain reason that OED1 had the task of marshalling *all* English lexis from the earliest days up to the early twentieth century, and SUP1 had the very different task of filling in that which OED1 had overlooked, and, for the earlier volumes of OED1 at least, covering the intervening decades. For the same reason, little of comparative value adheres to the fact that the 1% cited above represents 2,508 entries in OED1, while SUP1 only added 657.

Such direct numeric comparisons, whether in raw or percentage terms, might be more reasonably drawn between SUP1 and the additional material collected for SUP2, which had a similar scope (though a broader range). Such comparisons would need to engage not just the theoretical but also the sociohistorical dimension of the dictionary's making (so far as these are distinguishable), however. Between the inception of the OED project and the time of Burchfield's appointment as editor of SUP2 in 1957, OUP transformed from a British academic publisher into a global publishing operation, with established offices and experienced editorial staff in the United States, Canada, South Africa, India, and Australia. Those decades also saw, notably and for related reasons, the emergence of the very notion of 'World English'—the term is first attested in this sense in the same year as Burchfield's appointment (in the older 'Standard English' sense it goes back, in what may be seen as something of a congruity, to 1888, the year of the publication of the first volume of OED1).

In the same period, a number of national and regional supplementary dictionary projects 'on historical principles' were launched—notably in Canada (research beginning 1954; publication 1967), Jamaica (1951; 1967) and the larger Caribbean (early 1970s; 1996), South Africa (1969; 1978 and 1970; 1996), Australia (late 1970s; 1988), and New Zealand (mid-1950s; 1997)—with many of the later titles published by OUP (Benson: 107–108). Clearly these dictionaries were filling a void which was in a sense created by OED1's grand accomplishment, its very aura of comprehensiveness highlighting its many deficiencies to lexicographers at work outside the United Kingdom. At the same time, in their adaptations and extensions of OED's historical method, these projects also underscored the need for a broader outlook in Oxford. Indeed, in addition to calling on overseas OUP staff for advice (Gilliver: 456), at various times and to varying degrees Burchfield engaged the local expertise of several individuals involved in researching the dictionaries cited above, from casual and ad hoc volunteer contributions, to paid reading, to more formal editorial secondments (Gilliver: 456, 461, 475; Dollinger: 152–154; Allsopp 1996: xxiv; Silva 2019).

3. Burchfield's world

Burchfield believed OED1's attitudes towards 'foreign' words and usages too insular. Perhaps with Murray's metaphor of lexical 'citizenship' in mind, he leveled a rare criticism at the work of his predecessors for treating words 'almost like illegal immigrants' (Burchfield 1986: xi). To correct this, his Second Supplement would make 'bold forays into the written English of regions outside the British Isles, particularly into that of North America, Australia, New Zealand, South Africa, India, and Pakistan' (Burchfield 1972: xiv), as the Preface to the first volume announced. This declaration would earn Burchfield a reputation for lexical cosmopolitanism in the popular press as much as among lexicographers.

This reputation has been challenged by Ogilvie (2013: 165–209), who has found via a case study that Burchfield included fewer loanwords and World English terms, proportionally, than did Craigie and Onions in SUP1 (26% versus 31%), and, further, that he omitted 17% of the World English terms they had added (2013: 177). In doing so, Ogilvie writes, Burchfield was going ‘against all OED policy before and since’ (2013: 181), which suggests an exceptional cull directed at this category of word. In fact, however, the aim of SUP2 had always been to subsume and replace SUP1, rather than to extend and sit alongside it (Burchfield 1972: ‘Preface’), and on both editorial and practical grounds, existing material of various kinds was regularly omitted, some of it at Onions’s own suggestion (Gilliver: 454, 487).⁸ It is therefore worth considering whether the omitted items identified by Ogilvie might have been excluded on the basis of other factors than achieved Anglicity (the vast majority had been labeled *U.S.*), or even widespread currency, for instance policies excluding ‘obvious combinations’, as Burchfield called them (1989: 92–93), or those judging the quality and quantity of the quotation evidence.⁹ Indeed, many of the examples listed by Ogilvie (2013: 182) are combinations such as *frog-pond*, *chicken-eater*, or *gift store*. Other omitted entries, especially those for plant and animal names, give no citation evidence, or simply refer to encyclopaedias without quoting them; others still quote only one source.

All are now being reinstated in the revision of the dictionary (Gilliver: 557), which allows for a comparative analysis of their character with respect to etymological and regional labelling, as well as quality of quotation evidence, vis-à-vis those entries that were retained in SUP2 (Table 2).^{D3}

To date 359 omitted SUP1 entries (i.e. not appearing as or within an entry in SUP2) have been restored as independent entries in OED3 (shown in the right column of Table 2; a number of others have been restored as sub-senses and sub-lemmas—these are not analyzed). Of these, only five (1%) have a non-European etymology (ALIF, AMBAN, BELUKAR [SUP1: BLUKAR], IDDAT, and SHIPPO),¹⁰ compared to 10% of SUP1 entries retained in SUP2; and 12% have a regional label in the entry header or the first sense, compared to 14% for the retained entries. By these measures, therefore, the entries retained by Burchfield are more etymologically and regionally non-British than those he suppressed. The greatest factor affecting suppressed as opposed to retained entries is not the etymological or regional labels they carry, but the quotations they adduce: 19% of the restored SUP1 entries carry all new quotation evidence sourced by OED3 editors—i.e. no quotations from SUP1 are retained in the restoration, either because there were none in SUP1 or because they were of poor quality. By contrast, 97% of SUP1 entries retained by Burchfield and revised in OED3 appear there with at least one quotation from SUP1 or SUP2, and 76% carry two or more (versus only 14% of restored entries).

None of this is to say that there were no blind spots in Burchfield’s vision of World English. Some of these I discuss below. But his editorship did mark the beginnings of an articulated policy within the Dictionary project to remediate the documentation of non-British Englishes, especially those from what would become known as ‘Outer Circle’ regions (Kachru 1992), i.e. outside settler nations in North America and Australasia. It is perhaps as a corollary, or even a countertendency, to this more outward-looking view, that we should understand Burchfield’s reinstatement of the tramline symbol, cited by Ogilvie as one of several indications that Burchfield was not as open-minded about World Englishes as he is generally given credit for. Although use of the symbol may well contribute

Table 2. Features of SUP1 entries in OED3 (retained in SUP2 vs. omitted in SUP2 and restored in OED3)

Entry Feature	Retained in SUP2	Omitted in SUP2
Entry Etymology		
English or European	82%	95%
English	55%	78%
European	30%	25%
Not English or European	10%	1%
Other & undetermined	11%	4%
Regional Labels in Entry		
No Regional Label	86%	88%
With Regional Label	14%	12%
British and Irish	1%	2%
Non-British/Irish	12%	10%
N. American	9%	9%
Not British or N. American	3%	1%
# Quotations retained in OED3		
0	3%	19%
1	20%	66%
2	21%	8%
3	18%	3%
4	12%	2%
5+	25%	2%

to the ‘marginalization’ of such vocabulary, as Curzan says (2000: 108)—not inconsistently with Murray’s original metaphor—it may also have given Burchfield leeway to include words that would otherwise have found themselves on the other side of the margin—which is to say, off the page. And Burchfield included a good number: of the 23,488 new main lemmas added in SUP2 (not including sublemmas and cross-references), 2,755 (11%) carried tramlines.^{D4} Moreover, SUP2’s use of the symbol follows more closely Murray’s original logic in the General Explanations than did OED1 itself: scientific terms are not typically tramlined, regional usage labels are more frequent within tramlined entries, and when domain labels do appear there, they are typically for words pertaining to the cultural lore of other countries. For example, new tramlined lemmas in SUP2 that contain labelled senses most frequently have *Mus.* (94 times), *S. Afr.* (78), *N.Z.* (38), *Law* (31), or *Philos.* (27). The musical terms, a paradigmatic example of the second type of loanword described in the previous section, come mostly from the French, Italian, and German repertoire (e.g., ||ÉTUDE, ||GRAZIOSO, ||PRALLTRILLER); the philosophical are for the most part a hodgepodge of neo-Latin neologistic compounds (e.g., ||ENS RATIONIS, ||NATURA NATURANS) and German specialist jargon (||DASEIN, ||GEDANKENEXPERIMENT).

And yet, the inherent subjectivity of the practice could produce in Burchfield’s Supplement the same appearance of arbitrariness that had led Craigie to give it up: the new entry ||SABAYON, for instance, with English evidence dating back to 1906, received tramlines, while its etymon ZABAGLIONE, with nearly contemporaneous adoption (1899) did not.

Larger patterns display a similar absence of rationale, not to say rationality: of words with etymons from African languages, 47 are tramlined (e.g., ||INYANGA, ||NGOMA, ||TOKOLOSHE), while 204 are not tramlined (e.g., BOMA, COCOPAN, ZEZE); with the exception of somewhat loosely applied policies regarding ethnonyms, toponyms, and names for flora and fauna, for a large part it is hard to discern from the entries what principles might distinguish them as more or less deserving of tramlines than the others.

At the same time, in about 180 revised entries, Burchfield included instructions to delete tramlines, including for AMNESIA and ANAESTHESIA, MARASCHINO and MAFIA, THESAURUS and TALCUM.^{D5} The largest class of untramlined words are French cultural terms, including BALLET and BOULEVARD, CAFÉ and CASSEROLE, VAUDEVILLE and VINAIGRETTE. As the previous section would suggest, smaller by far is the group of untramlined words with non-European etymologies. As far as I have been able to ascertain, the following list is complete: AVOCADO, CHA, CHAI, COLA, DINGHY, DINGO, DUNGAREE, FUFU, GA, GO, *n.*², GUAR, KOSHER, LAC, MALLEE, MASSASAUGA, PUNDIT, *n.*, SAMURAI, SARONG, SATSUMA, SHARAWAGGI, SHEIKH, SUI, TYCOON, U², YOGA, YOGURT, ZUÑI.

Even if the practice of tramlining was ambiguous, arbitrary, and hard to interpret at scale, such *un*-tramlining is, fairly unambiguously, an editorial statement about the achieved Anglicity of these terms.¹¹ Beyond this, as I argued in the previous section, the measure of inclusion ought to be based upon SUP2's etymologies and regional usage labels, even if these were not always consistently applied (indeed this is one reason why they are best analysed together, as well as with other indicators, such as definition text). Burchfield added 6,550 new lemmas with etymologies other than Latin, Greek, or English (28% of all new lemmas), including 1,972 with at least one non-European etymology (8%).^{D6} Of all the entries added by Burchfield, 2,691 (11%) carried a regional label, whether for the entire entry, or one or more subsenses (most of them pertaining to North America or Australia and/or New Zealand).

Although North American English is, perhaps for obvious reasons, by far the most prevalent non-British English marked out in SUP2, the coverage of Englishes coming out of Africa, Central and East Asia, Australia, and New Zealand is substantial. SUP2's coverage of African Englishes, primarily South African English, is particularly notable, probably due to the assistance early on of expert local informants, who included a co-editor of the historical dictionary of Afrikaans (Gilliver: 461), as well as lexicographers of South African English (Silva 2019). Burchfield added 251 new words with African etymologies and expanded a further 154 entries with additional sense sections. To these numbers could be added a further 205 words with a regional African label but with non-African etymologies, most of them Afrikaans. It is a large total, dwarfing what had been included in either OED1 or SUP1, and rivaling Burchfield's own contributions of South Asian vocabulary. In this area, which had been attended to more closely than others in OED1 and SUP1, Burchfield added a large number of words with South Asian etymologies (325), but very rarely applied a regional label (37 times to these words, and only three times to words with English etymologies). 278 new New Zealand (or 'Australia and New Zealand') sense sections were added in 250 new entries, with 467 regional senses added to 416 existing entries. The majority of the new terms are of English etymology, but a large minority (97, or 39%) are loanwords from Māori (a further 39 Māori words have no regional label—these are mainly toponyms, ethnonyms and names for plants and animals, which typically, though not universally, are not labeled as regional).

Both OED1 and SUP1 had recorded similar numbers of Australian Aboriginal and Māori language vocabulary—61 Australian and 70 Māori in OED1, 35 and 31 in SUP1. Burchfield's Supplement, however, added only 62 Australian Aboriginal words to the existing number, but 136 new Māori words, bringing the total to 144 and 231, respectively.¹² While this disparity better reflects proportionally the fact that, due to various historical and linguistic circumstances, New Zealand English contains something like double the number of indigenous words than Australian English (Ramson 2002: 87–88), it also allows for some perspective on the assessment given by Weiner that 'the *Supplement's* coverage [of non-British Englishes] is so full that it is not far from equivalent to a collection of dictionaries of their contemporary vocabulary' (Weiner 1987: 32). The first edition of the *Australian National Dictionary* (1988), to cite an authority contemporaneous with SUP2, also published by Oxford University Press, and whose editor had consulted on Burchfield's Supplement (Gilliver: 461), listed more than 400 words of Aboriginal origin (Ramson 1988; this is on par with the count of 430 reported in Dixon 2008: 131); the new 2016 edition has more than 550. In a similar vein, the *Dictionary of New Zealand English* contains '700-odd' words of Māori origin (Orsman 1997: viii); a more recent and more comprehensive work has over 1,000 (Macalister 2005).

Clearly Weiner's assessment was greatly exaggerated: OED2, like its predecessors, is not close to the equivalent of a collection of non-British regional supplementary dictionaries, just as it is not equivalent to a collection of British regional dialect dictionaries. What is true, however, is that non-British English vocabulary forms a more substantial and more conspicuous dimension of OED2 compared to OED1, due to the expansive orientation of the editors of SUP1 and SUP2. Tramlines aside, both Supplements have the about the same percentage of words with non-European etymologies (7% and 8%) and about the same percentage of senses marked as regional (14% and 12%). Because it was larger and came later, SUP2 massively increased the store of all such items in the dictionary—much more than SUP1 was able to do vis-à-vis OED1 (Table 3, Table 4). Words with African, Central and East Asian, and Māori derivations were increased by 165%, 127%, and 137%, respectively, over what had been documented by 1933; other non-European loanwords were increased by between 33% and 98%. By contrast the number of words with English or European etymologies was increased by 10%. With respect to regional labels, SUP2 doubled the number of senses with a non-British regional label, vastly increasing the number of North American and Australian senses, and for the first time regularly applying labels to African and New Zealand senses. Even in raw terms, Burchfield's is still the largest contribution when it comes to words with African, Central and East-Asian, Austronesian, Australian Aboriginal, and Māori etymologies, as well as senses with North American, African, Australian, and New Zealand regional usages.

At the same time as one must recognize the growth in OED's documentation of these Englishes under Burchfield, the near absence of regional labels denoting Caribbean, South Asian, and East and South-East Asian usages points to blind spots in his global vision. OED1 had very frequently employed the labels *U.S.* and *Anglo-Indian* (or equivalents), whereas *West Indies* and *East Indies* were employed only a handful of times. More common in OED1 was to indicate those regionalisms within the text of the definition, e.g. for ||VEGA¹, b: 'In the West Indies, a piece of fertile meadowland'. Instead of standardizing the label form, as he had for African and New Zealand senses, across all regions, by and large Burchfield followed the descriptive model for Caribbean, South Asian, and East and South-

Table 3. Entry etymologies per edition, percentage growth over previous total

Category	OED1 #	SUP1 +%	SUP2 +%	OED3 +%
English or European	184,371	4%	10%	10%
English	108,874	4%	10%	13%
European languages	80,153	3%	9%	6%
Latin/Greek	52,631	3%	5%	5%
Romance languages	32,611	3%	8%	5%
French	29,741	2%	5%	3%
Spanish	1,278	13%	29%	21%
Portuguese	351	8%	17%	19%
Germanic languages	8,279	8%	23%	9%
Dutch	2,093	3%	11%	8%
Non-European	2,508	26%	62%	28%
Native American languages	311	24%	40%	26%
Middle Eastern and Afro-Asiatic languages	991	17%	33%	19%
African languages	90	69%	165%	46%
Indian subcontinent languages	669	24%	39%	30%
Central and Eastern Asian languages	302	33%	127%	36%
Austronesian	245	36%	98%	25%
Australian Aboriginal	61	44%	70%	28%
Māori	70	41%	137%	22%

Table 4. Regional senses per edition, percentage growth over previous total

Category	OED1 Total	SUP1 +%	SUP2 +%	OED3 +%	OED3 #
No Regional Label	540,330	3%	12%	25%	783,286
With Regional Label	20,052	15%	40%	133%	75,037
Britain and Ireland	12,259	1%	3%	140%	30,403
North America	6,745	38%	76%	130%	8,908
Other	1,308	27%	115%	149%	37,910
Caribbean	50	10%	60%	1005%	972
Africa	145	58%	154%	179%	1,621
South Asia	409	15%	16%	130%	1,260
E. & S.E. Asia	49	4%	20%	362%	282
Australia	450	24%	126%	99%	2,514
Aus. & N.Z.	150	45%	210%	127%	1,527
New Zealand	65	43%	311%	121%	844

East Asian senses, e.g. for SHOUTER, *n.*², 2b: 'In the West Indies, a member of a Baptist sect influenced by African religious practices'—a continuation of OED1's style in the case of Caribbean and East and South-East Asian words, but a significant departure in the case of South Asian ones. There are something like 80 such definitions for the three regions in SUP2,¹³ which I have not counted in my tallies thus far, as they make no statement about dialecticity per se (cf. the parallel form in BOUZOUKI: 'In Greece, a sort of mandoline'; but *bouzouki*, one observes, refers to a Greek sort of mandoline regardless of where the word is said). This is reflected in OED3's current labeling policy, which distinguishes between the

regional dialecticity of a sense, which takes an italicized label, and the regional frequency of a sense, which does not necessarily.¹⁴

Thus the Caribbean and South, East, and South-East Asian English senses added in Burchfield's Supplement are treated in the same manner as SUP2 might treat an Italian or a Russian word said by an English speaker in Christ Church, Oxford, as opposed to the way it treats a Māori word spoken by an English speaker in Christchurch, New Zealand; i.e. as denoting exotic items from far away places, rather than the standard regional English vocabulary of people in those places.¹⁵ Though context might be seen to mitigate the discrepancy, the question of labelling standards is more than a matter of formatting. Again the case of Caribbean English illustrates well how the availability of a standardized category can come to shape the lexicographical description. Despite the fact that the latest Caribbean lexicography—itsself inspired by OED—was not only available, but was being actively reincorporated into SUP2, only rarely was it being represented as Caribbean per se: Cassidy 1961 and Cassidy and Le Page 1967, for example, are cited forty-two times in SUP2, but only six times in senses under a Caribbean regional label, versus seven times in senses marked *dial.*, and seventeen marked *U.S.* (often paired with *Black*). As we shall see in the next section, part of the behind-the-scenes revision of the dictionary for *OED Online* has involved algorithmically revising labels and adding back regional markers to definitions, so that sense 2b of *SHOUTER* now includes a 'Categories' link with 'Caribbean' listed in the pop-up window, even though the entry itself has not been revised since SUP2, and thus bears no equivalent label in the main text.

4. OED3 – Towards a 'world language' English dictionary^{D7}

What Burchfield articulated as a significant aspect of his enlargement of OED, the current Dictionary's editors have taken as a core principle of revision and expansion (Weiner 1987, Simpson 2000, Price 2003, Salazar 2014). English is described in the Preface to OED3 as 'a world language, in which individual varieties share a common core of words but develop their own individual characteristics', with the effect that the 'English of the British Isles now becomes one (or indeed several)' of these varieties (Simpson 2000). Several special reading programmes and public appeals targeted at varieties of World English have been launched, and since 2014 OED has counted a 'World English Editor' among its senior editorial positions. A number of updates since then have focused on global varieties of the language, notably the Englishes of Singapore and Hong Kong (March 2016), the Philippines (June 2015 and October 2018), India (September 2017), South Africa (December 2018), and Nigeria (January 2020). The June 2016 update saw the integration of twelve pronunciation models for several types of World English (Sangster 2016); OED3 now has fifteen such models (Sangster 2020). As a result of this editorial activity, global English has been a returning topic of discussion in the release notes and blog posts covering OED's quarterly updates, reflecting the high level of interest and preoccupation within the Dictionary project itself and, presumably, among the dictionary's usership.

Consistent with this orientation, the user interface to *OED Online* in operation since 2011 allows one to filter search results by etymological origin¹⁶ and by regional category. The addition of these categories, which have been determined automatically based on labels and definition text, represents a significant change in the presentation of the dictionary. Categories appear in pop-up windows attached to individual senses via a link; entries

bearing such links can be perused or searched through a dedicated browsing page (currently at <https://oed.com/browsecategory/>) or with the 'Advanced Search' function. The actual regional labels appearing in the dictionary text, however, are no longer searchable.

Though it has the benefit of implementing a standard geographical taxonomy, there are several theoretical and technical problems with the regional category feature as it currently is implemented. For one, the categorization scheme is neither traditional nor fully rational. The subdivision of United States English, for instance, comprises a mix of idiosyncratic geographical divisions (i.e. Northern, Eastern, Southern, Western, rather than the more usual North-East, mid-West, West Coast [Pacific], etc.) plus one ethno-linguistic category (African-American) and one conventional dialectical region (United States Midland). The decentering of the British perspective is not very evident either: there are twenty-one subcategories for 'Britain and Ireland' (including one each for Orkney, Shetland, and the Isle of Man), yet only one main category for each of 'South-East Asia' and 'India'. The former is simply overly broad, frustrating searches for, e.g., Philippine or Singaporean English in particular. The latter is both too narrow and too broad, leaving without a proper category those South Asian Englishes spoken outside the nation of India, as well as the regional variations within. The automatic classification is also far from perfect: the algorithm has classified some strings of definition text (e.g., 'In the West Indies') but not others (e.g., 'In Guyana').

These are flaws which may be addressed as part of a future technical update, as was a longstanding systematic miscategorization of all references to 'South Asia' (as opposed to 'India') within the 'South-East Asia' class, corrected during the course of 2019. More fundamental however is the question of whether the application of text-recognition algorithms can justifiably stand in for lexicographical judgement in this regard. It is evident, for instance, that in many cases, words and senses have been categorized as 'regional' because they refer to regions, rather than because they represent actual regional usages. ANTILLEAN, e.g., is categorized as 'Caribbean', even though 'Antillean' is not an exclusively or even a particularly Antillean word. Many names for flora and fauna have been similarly categorized, as have names of local customs, apparel, and foodstuffs. CASHEW, *n.*², e.g., is likewise categorized as 'Caribbean', because definition 1, from 1888, contains '[...] cultivated in the West Indies', which even at the time it was written was encyclopaedic information pertaining to cashews, not lexicological information pertaining to *cashew*.

While it is debatable whether such designations represent errors or simply the institution of a broader categorization method, the ironic result of replacing label searches with category searches has been that it is impossible at present to search *OED Online* for actual regional usages as determined by lexicographers—these must be sorted manually from the larger class of words either from or pertaining to (sometimes quite large) regional areas. At the same time, it is also the case that, both in the search panels and within entries themselves, OED3 presents its regional categories as an integral part of the lexicographical document, continuous with its editorial emphasis on World Englishes.¹⁷ Therefore in what follows, unless otherwise specified, I count 'regional' senses as they are presented, i.e. either bearing labels or in categories, or both.

So far, as Table 3 indicates, OED3 has increased only modestly the number of entries bearing non-European etymologies, +28% versus +62% in SUP2, though still these items are being added at a greater rate than English or other European derived words (+10%), and the project is but at its mid-way point, with 48% of entries still unrevised. In its labelling of

regional senses, however, the changes are more dramatic, even if the figures are inflated somewhat by the automatic application of regional categories (Table 4, Table 5). One fifth (21%) of the new senses, subsenses, and sublemmas added since 1989 carry (or fall under) a regional marker, increasing the total number by +133%, with new senses in new entries slightly more likely to be marked as regional (23%) compared to new senses in revised entries (19%). In the amalgamated OED2, by contrast, only 5% of senses carried a regional label (the large majority American or Scottish usages); in OED1 it was 4% (the Supplements labelled at an intermediate rate, and 14% and 12%, respectively). After the first edition, in which Scotticisms outnumbered Americanisms, North American English is the largest category in every revision and edition. It makes up 66% of newly added regional senses in OED3 (14% of all new senses), versus 7% of OED1's regional senses (<1% of all senses). Thus, despite the overall increase in regional labelling, North American English makes up only a slightly larger percentage of the OED3 expansion than it did of SUP1 (12% of all new senses) and SUP2 (9% of all new senses), approximately 1.1 and 1.5 times their rates, respectively.^{D8}

If North American usages are the most conspicuous elements of the OED3 expansion in raw terms, relative to the previous Supplements two other areas stand out. The first is the newly systematic use of Caribbean and East and South-East Asian English labels and categories,¹⁸ of which there are only a handful in all of OED2. Beyond this, as far as non-British varieties of English go, things have been relatively stable: like North American words, African, South Asian, Australian and New Zealand words, while contributing significant numbers of regional senses, do so at roughly the same ratios as the Supplements (an exception is South Asian senses vis-à-vis SUP2, which are 3.7 times the proportion, though they are only 1.2 times more prevalent than in SUP1). The second main driver of increased regionality in OED3 is, counter-intuitively perhaps, the labelling of British and Irish senses: these make up 14 and 12 times the proportion of all new entries as they did in SUP1 and SUP2, respectively; closer to the increase in Caribbean (17 and 9 times) and East and South-East Asian (19 and 14 times) than, e.g., the Australian (1.4 and 0.8 times) or North American. While clearly these do not represent World English usages, they may be seen to reflect OED3's efforts to decenter the British perspective, treating British English as a regional standard English unto itself, as well as more focused attention on regional varieties within British English (as highlighted in Sofield 2018).

The most dramatic change in regional labelling between OED2 and OED3 pertains to Caribbean English, with an increase of +1005% labelled senses (the figure would be smaller, though still large, if we were to count OED2's descriptive labels, of the 'In the West Indies' variety, and also discount OED3's algorithmically mis-identified regionalisms, like *CASHEW*). This may be seen in large part as remediating a defect of Burchfield's edition. The other very substantial increase pertains to East and South-East Asian words (+362%), which may additionally be seen to reflect the spread and growth of World English itself since Burchfield's tenure, as well as OED's vastly improved technical capacities, and vastly enlarged information and informant networks (OED3 is now soliciting contributions on Twitter and other 'platforms'), with which to document this growth. With this in mind it is instructive to note another disparity between the English of Inner Circle regions and that of Outer Circle regions. Newly added senses with regional markers designating North American, British and Irish, and Australian usage all are somewhat less likely to have been added in new entries as opposed to in existing ones (ratios all between 0.6 and 0.8:1). By contrast, the new senses describing African, South Asian, and East or South-East Asian

Table 5. Senses by regional label, per edition of origin, % of category (% of total)

Category	OED1		OED2		OED3	
	All Senses		All Senses		All Senses	
No Regional Label	96%	(96%)	95%	(95%)	91%	(91%)
With Regional Label	4%	(4%)	5%	(5%)	9%	(9%)
Britain and Ireland	61%	(2%)	39%	(2%)	41%	(4%)
North American	7%	(1%)	11%	(3%)	51%	(4%)
Not Br./Ir. or N. Am.	7%	(0%)	7%	(1%)	12%	(1%)
Caribbean	4%	(0.0%)	2%	(0.0%)	12%	(0.1%)
Africa	11%	(0.0%)	16%	(0.1%)	20%	(0.2%)
South Asia	31%	(0.1%)	15%	(0.1%)	14%	(0.1%)
E. & S.E. Asia	4%	(0.0%)	2%	(0.0%)	4%	(0.0%)
Australia	34%	(0.1%)	35%	(0.2%)	27%	(0.3%)
Aus. & N.Z.	11%	(0.0%)	19%	(0.1%)	16%	(0.2%)
New Zealand	5%	(0.0%)	11%	(0.1%)	9%	(0.1%)
	OED2		OED2		OED2	
	New Senses		New Senses / Added SUP1		New Senses / Added SUP2	
No Regional Label	88%	(88%)	86%	(86%)	88%	(88%)
With Regional Label	12%	(12%)	14%	(14%)	12%	(12%)
Britain and Ireland	4%	(0%)	2%	(0.3%)	5%	(0%)
North American	78%	(10%)	87%	(12%)	70%	(9%)
Not Br./Ir. or N. Am.	19%	(2%)	12%	(2%)	26%	(2%)
Caribbean	2%	(0.0%)	1%	(0.0%)	3%	(0.0%)
Africa	27%	(0.4%)	24%	(0.4%)	29%	(0.5%)
South Asia	9%	(0.1%)	17%	(0.3%)	6%	(0.1%)
E. & S.E. Asia	1%	(0.0%)	1%	(0.0%)	1%	(0.0%)
Australia	31%	(0.8%)	31%	(0.5%)	31%	(0.9%)
Aus. & N.Z.	17%	(0.5%)	19%	(0.3%)	16%	(0.6%)
New Zealand	13%	(0.3%)	8%	(0.1%)	15%	(0.4%)
	OED3		OED3		OED3	
	New Senses		New Senses / in New Entries		New Senses / in Existing Entries	
No Regional Label	79%	(79%)	77%	(77%)	81%	(81%)
With Regional Label	21%	(21%)	23%	(23%)	19%	(19%)
Britain and Ireland	23%	(5%)	23%	(5%)	23%	(5%)
North American	66%	(14%)	61%	(14%)	69%	(13%)
Not Br./Ir. or N. Am.	14%	(3%)	19%	(4%)	10%	(2%)
Caribbean	14%	(0.4%)	15%	(0.7%)	12%	(0.2%)
Africa	23%	(0.6%)	27%	(1.2%)	16%	(0.3%)
South Asia	13%	(0.4%)	16%	(0.7%)	9%	(0.2%)
E. & S.E. Asia	6%	(0.2%)	8%	(0.4%)	4%	(0.1%)
Australia	26%	(0.7%)	19%	(0.8%)	35%	(0.7%)
Aus. & N.Z.	12%	(0.3%)	8%	(0.3%)	18%	(0.3%)
New Zealand	8%	(0.2%)	8%	(0.3%)	8%	(0.2%)

usages are two to three times as likely to come from newly added words as from existing ones (ratios between 2.3 and 3.0:1). This supports the idea that OED3 is casting a wider net than its forerunners, capturing a broader range of lexical items from these regions. But it is simultaneously the case that the linguistic diversity of these regions, with English often functioning as a minority language, second language, or lingua franca, produces a language with a larger proportion of loanwords, blends, hybridizations, irregular formations, and other neologisms than the Englishes developing as majority languages in settler nations.

This is further supported anecdotally in Salazar's account of the Philippine English added to OED3 as part of the October 2018 update, in which she describes senses added to twenty headwords, with proximate etymologies from English (9), Tagalog (6), Spanish (5), Portuguese (1), Catalan (1), and Hokkien (1). These include the neat hybridization PANCITERIA, 'signifying a Philippine noodle stall (Tagalog *pancit* for 'noodle', ultimately from Hokkien + Spanish *teria* suffix)', and the amusing blend TRAPO—'a derogatory term for a politician' combining 'the two words that make up the English phrase *traditional politician*' but punning on *trapo*, 'the Spanish word for a cleaning cloth, which has also been borrowed into Tagalog' (Salazar 2018). The general picture can be confirmed quantitatively as well, by observing the distribution of etymologies within the set of senses bearing each regional label. The East and South-East Asian group, which includes the English of Singapore, Hong Kong, Malaysia, and the Philippines—most newly-labelled in OED3 and all the subject of special update initiatives since 2012—understandably has local senses with a very broad etymological profile, including Austronesian languages (49%), European languages other than English (25%), English (14%, the lowest of any regional English), East Asian languages (12%), Indian subcontinent languages (6%), and Middle Eastern languages (3%). It also has the second highest rate of blending, with 15% of senses bearing an etymology from more than one of these groups (South Asian English has 18%). By contrast, just 5% of senses with no regional label are within entries with more than one language group in the etymology.

As might be expected, in general local senses in Kachru's Inner Circle regions have etymological profiles that resemble 'common' English, augmented somewhat by local terms from indigenous languages. Exceptions include those senses marked 'New Zealand' only (as distinct from 'Australia and New Zealand') which bear significantly more indigenous etymologies, as well as usages from the Caribbean, which show a broader range of linguistic influences than the rest of Anglophone North America, including a higher (though still small) percentage of words with indigenous American, African, and South Asian origins (e.g., respectively: PIROGUE, BURRU, BAP²), and a higher incidence of words from the languages of other former colonial powers, i.e. Spanish, French, Dutch, and Portuguese (CALLALOO, JAMETTE, KOKER, PIMENTO). In many cases an etymology tells of centuries of global language contact, as MANGOSTEEN, which makes its way to Barbados (where it once described 'a kind of jujube tree') via French and Dutch, having been transmitted to those languages from Malay via Portuguese.

The insidious side of this may equally be highlighted by the various terms more directly associated with colonial racism, originating elsewhere in the Imperial Anglosphere but carrying (or having carried) specific Caribbean regional senses, e.g. DOUGLA (from Hindi), MOKO (perhaps Kalabari), or QUASHIE (Akan). In this respect it is especially indicative that the highly derogatory term COOLIE (probably from Gujarati, via Portuguese) was brought not only to Caribbean shores, but also found its way to the United States and to South Africa, where it refers either to a person of South Asian or East Asian origin. Similarly the offensive term PICCANINNY (also Portuguese) has a number of localized senses spanning the

former colonial world, in each case referring to a different local racialized group: in the Caribbean it refers to a 'child of African origin', in the United States 'an American Indian child', in Australia and New Zealand an 'Aboriginal or Maori child', and in South Africa and Western Africa a 'small black African child'.

Historically speaking, therefore, both from a linguistic and a meta-lexicographical point of view, it is hardly satisfactory to divide World English only according to how it has developed in Inner Circle versus Outer Circle regions, or how much of the vocabulary is of English versus indigenous origin. World Englishes have variously been seeded by colony, settlement, trade, and any number of other vectors by which language spreads itself throughout the world (more recently, e.g., the English Internet). Therefore even those regional senses of English origin are often ambivalently 'local': in addition to post-contact local sense extensions, the category also includes senses preserved from pre-contact British English, sometimes itself of a regional British variety. OED3's newly attentive documentation of Caribbean English may once again be taken as indicative of the variety of causes and influences a language can display: a lingua franca in the Spanish and French Caribbean, in those nations where it is spoken as a mother tongue it typically does so on a spectrum with a local creole. Additionally, those nations themselves, however shaped by British colony, were settled *by* the British but largely *with* non-British people bringing their own languages.

OED3 now contains 972 senses falling under a Caribbean regional designation (811 with labels, 919 in categories), in 731 entries, including about 475 revised senses (many of which carried no label—or category, per force—previously) and about 130 senses 'stealthily' edited by algorithm, or as part of OED3's standardization of OED2's labels.¹⁹ This is still less than what is recorded in more comprehensive supplementary works, such as [Allsopp 1996](#) and [2010](#) combined, or even [Cassidy and Le Page 1967](#). However, OED also differs from these historical dictionaries in that it always aims to document the first written attestation of a sense, making it the only dictionary containing Caribbean English to do so in this way, despite its restricted scope.²⁰ Thus, even taken as a microcosm or as a partial view, it is revelatory to note that 13% of regional Caribbean senses documented in OED3 are attested (in British sources) before 1600—i.e. predating the first lasting British settlements—the largest proportion in any region outside Britain and Ireland.

These preserved terms, which have fallen out of use in general English (or are also dialectal to other regions), include otherwise obsolete senses of many common English words, such as FOOT, *n.* (I.1c: 'The entire leg'), PLAGUE, *n.* (1b: 'A wound') and SAD, *adj.* (3b: 'of a person, orderly and regular in life', also *Scottish*); grammatical functions such as productive ALL (A.3, as in *all-kind*, *all-thing*), transitive LOOK, *v.* (5b) or MUCH, *adj.* used to modify a count-noun (II.3a); and also several (otherwise) archaisms, such as BEFORETIME,²¹ NOSE-HOLE, and NOTHER, *pron.*¹. In each of these cases OED2 had recorded the lemma, and something very near the sense, but had not discovered or did not label the local Caribbean usage, which in most cases could have been located in contemporary lexicological resources, such as [Collymore 1957](#), [Cassidy 1961](#), or [Cassidy and Le Page 1967](#).²² In the case of BEFORETIME, even the readers and editors of OED1 might conceivably have taken notice of it, in the memoirs of the missionary H. M. Waddell, *Twenty-nine Years in West Indies & Central Africa* (1863), which records the word in a conversation he had in Jamaica in 1836 or '37 (OED1 fascicle BATTER–BOZ came out in early 1887).

Waddell's book, quoted throughout [Cassidy and Le Page 1967](#), instead remained unknown to OED until the present revision. Perhaps the biggest change in documentary

practices between OED3 and its predecessors has been the ways in which such documentary evidence is gathered. Previous editors had recourse to a number of reference works—dictionaries, concordances, indexes, encyclopedias, textbooks, and so forth—in which they could actively search out a desired word, in order to obtain references or quotation evidence from which to form an entry. Most of the evidence was collected passively, however, in the form of slips sent by volunteer (later, more commonly, paid) readers. Today much of the evidence brought to bear on revision is not collected in this way, but actively searched, using massive databases of electronic text. The Oxford New Monitor Corpus (a.k.a. Oxford New Words Corpus) is one such database, amounting (at last measure) to nine billion words scraped from the Internet, starting in 2012; the Oxford English Corpus is another, covering 2.5 billion words attested between 2000 and 2006.

These annotated corpora are well suited to the identification of new words for potential inclusion in the dictionary, and also facilitate the documentation of new senses of existing words, though to a lesser extent. They do not, however, help to address historical uses. As Weiner observed at the time, these represented a particular shortcoming in SUP2's coverage of World English, due to Burchfield's general remit to document mostly twentieth-century English. The result, he observed, was a composite dictionary which was 'diachronic but limited spatially to Britain, and international but limited temporally to the present' (Weiner 1987: 32). And yet, even the partial record in OED shows that, for most World Englishes, the period from the beginning of the nineteenth century to the publication of SUP1 is the most productive among comparable spans in terms of the recording of new words and senses (Table 6): 59% of non-British regional senses are first attested during this period, versus 37% of senses with no regional label (and 26% of senses with a British label). Most centered within this timeframe are Australian and New Zealand senses (65%–68%); these Englishes also retain the smallest proportion of pre-1800 senses than any other group (12%–8%). Caribbean and South Asian senses tend to skew somewhat earlier (39% and 36% coming before 1800, respectively), partially a reflection of earlier points of first contact, while North American senses skew slightly later (30% coming after). Most contemporary of all are the East and South-East Asian senses, just under half of which (46%) are first attested after 1933. While a large part of this effect may be historically motivated, in part it must also be due to OED3's recent special interest in the Englishes of this region.

The implication is that there remains much historical World English left to be documented, especially in the period pre-1933. In pursuing this, editors of OED3 have had recourse to recently published or updated national and regional historical English dictionaries (such as those mentioned in the previous section), as well as historical digital archives, such as the National Library of Australia's *Trove*, which includes Australian newspapers from the early nineteenth century to the mid-twentieth, or the various repositories of North American newspapers and periodicals, much of which was scanned from microfilm during the first wave of mass digitization in the 1980s and '90s. These methods, as indispensable as they are in tracing the history of regional words and usages, also introduce a systemic bias that is equally a legacy of colony (as opposed to settlement), in that (1) at each historical remove, the textual record of many World Englishes represents increasingly the English of white colonial and/or settler class, in addition to local vocabulary in the form it was recorded by colonists; and (2), to the extent that they are textual as well as oral, comparably comprehensive digital resources do not yet exist for many varieties of

Table 6. First attestation of OED3 senses by region, % in date range

Category	< 1100	1100–1499	1500–1599	1600–1699	1700–1799	1800–1883	1884–1932	1933–1971	1972–1988	1989–2020
No Regional Label	2%	12%	12%	17%	10%	24%	13%	7%	1%	0%
With Regional Label	1%	9%	8%	6%	9%	26%	19%	14%	5%	2%
Britain and Ireland	3%	19%	18%	11%	13%	20%	6%	4%	3%	1%
North America	0%	2%	2%	3%	6%	29%	27%	21%	7%	2%
Not Br./Ir. or N. Am.	0%	1%	2%	4%	8%	33%	26%	18%	6%	1%
Caribbean	2%	5%	5%	10%	16%	26%	13%	14%	6%	1%
Africa	0%	1%	1%	2%	9%	35%	23%	19%	7%	2%
South Asia	0%	1%	4%	14%	16%	37%	13%	10%	3%	1%
E. & S.E. Asia	—	0%	3%	4%	6%	17%	23%	22%	16%	8%
Australia	0%	1%	1%	1%	3%	33%	32%	22%	6%	1%
Aus. & N.Z.	0%	1%	2%	2%	3%	30%	37%	19%	4%	1%
New Zealand	0%	2%	1%	3%	6%	46%	22%	17%	2%	0%

Table 7. Antedating of regional words in OED3 revised entries

Category	All Revised Entries			Antedated Entries Only					
	% Antedated	% No Change	% Postdated	Mean Antedating	Median Antedating	1st St. Dev.	%10y+	%25y+	%50y+
All Revised Entries > 1500	51%	40%	9%	46	25	58	73%	51%	30%
No Regional Label	50%	40%	9%	47	25	59	74%	51%	30%
With Regional Label	55%	36%	9%	43	23	54	71%	48%	27%
Britain and Ireland	47%	39%	14%	56	32	66	75%	57%	37%
Non-British/Irish	63%	31%	6%	33	19	42	67%	42%	20%
North America	65%	30%	5%	33	19	43	66%	42%	21%
Caribbean	53%	42%	6%	48	21	61	69%	47%	32%
Africa	47%	41%	12%	33	19	43	65%	42%	22%
South Asia	52%	41%	8%	49	32	53	83%	60%	34%
E. & S.E. Asia	54%	40%	6%	38	24	44	74%	49%	24%
Australia	64%	30%	6%	32	17	43	69%	41%	18%
New Zealand	64%	27%	9%	32	17	42	70%	42%	18%

World English. Thus it should not be surprising that OED3 has antedated Outer Circle English words at a far lower rate than Inner Circle English words (Table 7).^{23, D9}

Indeed, by far the highest antedating rate for post-1500 regional vocabulary occurs for words marked as North American (65%), followed by Australian (64%) and New Zealand words (64%). South Asian and African words, by contrast, are antedated 52% and 47% of the time, respectively. While it may seem at first a contradiction that Caribbean words, like South Asian words, have both a low rate of antedating (53%) and a high percentage of very long antedatings (32% being of 50 years or more), this may be explained by the proportion of preserved pre-contact English in the language. While a Caribbean usage is less likely overall to be antedated than an Australian one, when earlier evidence is found, it is more likely than the Australian or the American antedating to be from the *Early English Books Online* database, or some other historical repository of British English, as opposed to local sources preserved in national corpora.

5. Conclusion

All editors of OED conceived of the Dictionary project as both global and comprehensive, if only in theory and only to a point. Those ideals themselves have meant different things to each editorial group, though clearly all viewed their part in the OED project as expansive vis-à-vis their predecessors. The picture each successive Edition and Supplement gives us is therefore to be understood as composite, what Benson calls 'a representation of the English language' (2001: 8), which is equally self-reflective of the ideas and attitudes of the people who assembled it. It is also, naturally, at once both self-reinforcing and self-revising, the idea of 'English' in Oxford a matrix for evaluating documentary evidence from all over the world, which evidence itself takes a role in forming the idea of 'English' in Oxford.

When the revision project that is OED3 reaches completion—with any luck in two, perhaps three decades' time—no doubt our picture of the state and development of World English will be of a higher resolution than it is today. There are more words to document, more still to expand and update and antedate. However, it will be important to bear in mind once that detailed picture has been drawn, that it will be a palimpsest, drawn over layers of successive images of English, starting with the first fascicle of OED1, printed in 1884 (where, on p. 5, see ||ABACA). OED1 and SUP1 formed an idea of English, including World Englishes, for millions of readers across the globe for the best part of the twentieth century; SUP2, OED2, and now OED3 for countless more since 1989. The editorial history of the dictionary in this respect is not only a matter of the development of lexicographical theory and practice, therefore, to be evaluated against the latest state of the art. It is, additionally, part-and-parcel of the cultural context from which it draws and to which it contributes. The OED is both a story of English, and, in its own development, a set of chapters in that very story.

7. Data notes

^{D1} In order to parse different editions of the OED's famously intricate patchwork, here I cross-compare the background data of several primary sources: the 1987 Tri-Star CD-ROM publication of OED1 (1E; names of corpora subject to analysis are styled in bold in these notes), public domain PDFs of OED1 and SUP1 kept at archive.org and other

repositories, the SGML-coded text of OED2 (2E), and an XML version of the January 2020 (December 2019) release of OED3 (3Etxt), its bibliography (3Ebib), index of new materials (3Enew), and index of publication origin (3Eorig), supplied under license by Oxford University Press. 2E has been the subject of a metadata enhancement project at St Jerome's University (Waterloo, Ontario) since 2011, and has been annotated to reflect, among other things, the edition that each quotation was added (Williams 2017). From this information I have derived three virtual editions: a virtual OED1 (1Ev), SUP1 (S1v), and SUP2 (S2v). Importantly, SUP1 material not retained in SUP2 is absent from my S1v. A restoration of the original SUP1 to a parseable digital format, underway at the time of writing, was still too preliminary to submit to analysis. For simplicity's sake, here the relatively small number of revisions and additions made for OED2 are amalgamated with SUP2, as are the 1993–1997 Additional Series with OED3. Labels from each edition, including language and regional labels, have been deduplicated and categorized according to OED3's labelling systems. Derived datasets can be provided on request, and enquiries are welcome. All OED data is published by Oxford University Press. I am indebted to several individuals at OUP who have assisted me at various stages of my work on the dictionary, especially James McCracken.

^{D2} Because 1E does not preserve the tramline mark, to produce the data in Table 1 and the following discussion, OED1 tramlined entries were extracted from 1Ev, adding back a list of entries untramlined in OED2 (see Data Note D5, below). In 587 tramlined OED1 entries (6%), no systematically labeled etymology is given. For these, first an attempt was made to follow any cross references within the etymology (e.g., 'from prec.'). If this failed, and there was a corresponding entry in 3Etxt, then the OED3 etymology was used (504 entries). Although in rare cases this may have introduced an anachronism (e.g., if lexicological knowledge changed between editions), in the vast majority it simply represents a precision (e.g., from 'Native name' to an actual language group). English etymologies (e.g., for blends and compound forms) are not counted.

^{D3} Using 3Etxt and 3Eorig, recovered entries were identified if they originated in SUP1 but bore no corresponding 2E entry. Any S1v lemma which was cross-listed anywhere in 2E was deemed to have been sublemmaed (rather than omitted) and thus ignored. Quotations in the remaining 3E entries were then compared for similarity to the text of the corresponding entry in S1v, with likely matches verified manually. For quotation counts of SUP1 entries retained in OED2, old versus new OED3 quotations were identified using 3Enew, and only revised entries were counted (unrevised entries having 100% old quotations).

^{D4} The data analyzed here are drawn from S2v.

^{D5} The list of 180 untramlined lemmas may not be fully complete, as it was reconstructed from the (sometimes poorly OCR'd) scans of OED1 archived at archive.org. The reconstruction was compared to 2E, and then manually verified.

^{D6} These further exclude 'undetermined' etymologies, as well as words formed as eponyms, acronyms, etc. Totals here may be less than what is attributed to SUP2/OED2 in my analysis of existing OED3 entries in the next section (see Data Note D7, below), since the same OED2 entry may be recorded as the 'original entry' of more than one OED3 entry (i.e. where an entry is divided between two or more lemmas).

^{D7} The OED dataset analyzed in this section is the amalgamated current edition, including all revisions up to and including the January 2020 update (such as it is called in the

Release Notes—the dictionary itself records the update as December 2019). At this time just over half of the material in the dictionary had been added or revised since 1989: 126,723 entries had been fully revised, and 20,674 newly added, leaving 136,862 unrevised entries ('new' entries include some instances in which subsections of existing entries became entries unto themselves). New entries account for 34,049 sense sections, with 56,220 new sense sections having been added to existing entries. It should be noted that a certain amount of linguistic bias is 'baked in' to any analysis of a partial OED revision, since entries are not revised randomly: at first, the revision proceeded alphabetically, starting at M and reaching the end of R before this approach was retired in 2006 (Gilliver, 576). Now entries are prioritized for revision or addition based on a number of factors, including evidence of significant semantic development, token frequency in linguistic corpora, and frequency of online lookups, and supplemented by ad hoc prioritization, topical coverage, and special initiatives, including those aiming to expand the coverage of particular varieties of World English.

^{D8} Because the revision of senses in OED3 involves a significant amount of rewriting and reordering, it is not theoretically sound to assign an 'edition of origin' to an OED3 sense section, in the same way one reasonably might with an entry (or with a sense section in OED2 for that matter). Therefore the data presented here are collected separately from 1Ev, S1v, 2E (combining SUP2 and the small number of additions made subsequently) and 3Etxt, with 3Etxt separated out into new, revised, and 'stealthily' revised senses. In each case, a 'sense' here refers to any Arabic-numeral designated sense or miniscule-alpha designated subsense, as well as all sublemmas, such as combined forms. Acronyms and initialisms are ignored. For 1E and 2E (including S1v and S2v) senses, labels are deduplicated and categorized within OED3's taxonomy, and are propagated down the hierarchy of senses, so that if a grouping is labelled, all senses that fall under it assume that region, whether or not the label itself appears in the dictionary text at that level. For OED3 senses, the method employed here combines labels and the regional category tag embedded in the background XML of 3Etxt.

^{D9} These figures treat revised 3Etxt entries where both the entry and its 2E predecessor have earliest attestations occurring after 1500. Dates of the earliest quotations falling in the main sense section in each edition are then compared to arrive at an antedating or postdating. Regional usage labels are assigned to words if either the entire word or the first sense is marked as such—later senses, subsenses, and combinations are ignored. In other words, to use an example from the main body of the article, this avoids treating an antedating of OED1 Box, *n.*² as regional on account of *box-wallah* appearing later on in the entry. Exceptionally, here, words with etymologies unambiguously coinciding with an existing region label are assigned that label, even if the entry does not include one (e.g., New Zealand English for any Māori word).

Endnotes

1. The mark was applied within SUP1 entries, however, to signal when a word could or should be pronounced according to (or in simulation of) the original language's conventions, e.g. AIDE-MÉMOIRE: 'ēi·dmemwāi, ||ēdmemwār'.
2. At the time this was written, and as late as May 2020, a prototype API (Application Program Interface) to query OED3 data directly was available for public testing at

- <https://developer.oxforddictionaries.com/our-data>. As this article was going to press, however, a check of that URL returned a blank page.
3. An earlier article focusing on Burchfield and World Englishes, published in this journal (Ogilvie 2008), appears, revised and enlarged, as Chapter 6 of Ogilvie 2013.
 4. The *Corpus of Historical American English*, for instance, records twenty-three instances in popular magazines such as *Time*, *Good Housekeeping*, *Reader's Digest*, and *The New Republic* up to 1980. There are thirty-three more in fictional works of varying salubrity, all between 1960 and 1980.
 5. Indeed, where even vaguely scientific terms are concerned, there is a clear presumption in favour of technical as opposed to popular usage, as the quotation evidence for *VAGINA*, labeled 'Anat. and Med.', illustrates: all seven quotations for senses referring to animal biology are taken from medical and zoological textbooks and handbooks, despite the evidence available to SUP2 editors noted above. A revised entry was published in OED3 in June 2019.
 6. This is the second sense of 'World English' documented in OED3, corresponding to the standard and standardizing Englishes spoken in Kachru's 'Outer Circle' (1992: 356). An earlier sense, which goes back to OED1, refers to something like the opposite—an 'international variety regarded as acceptable wherever it is spoken in the world'—corresponding to 'World Standard English' as described in McArthur (1998: 97).
 7. With the goal of balancing consistency and variety, here I use 'lemma', 'headword', and 'word' quasi-synonymously, to refer to what might make up an 'entry' in a dictionary such as the OED; and 'sense' to refer to the more finely distinguished semantic units contained therein. In general I have counted attributive and combined forms (but not initialisms or acronyms) along with senses and subsenses; a more detailed examination might wish to make finer distinctions.
 8. It might further be noted that, although the stated policy of OED3 is now that 'The revised text will include all entries (headwords) and meanings, compounds, phrases, derivatives, etc., included in earlier editions of the Dictionary' (Simpson 2000), this is subject to a number of other editorial considerations: headwords can be changed (e.g., JOHN CANOE is now JUNKANOO; MOCO-MOCO now MUCKA-MUCKA); entries merged and lemmas sub-lemmaed (e.g., AIN'T, *v.*¹ now s.v. BE, *v.*; WATER-WITHE now s.v. WATER); sub-lemmas promoted to entries in their own right; sense sections split up or combined. On a much larger scale, quotation evidence from earlier editions is routinely omitted from revised entries.
 9. Burchfield explains in the introduction to the first volume that 'Earlier U.S. examples' in SUP1 were not generally retained, nor were most pre-1820 antedatings, because systematic antedating could not be undertaken for SUP2, and, in the case of the American examples, the work was being accomplished by other historical dictionaries of American English (Burchfield 1972: xv).
 10. The last of these, SHIPPO, was in fact re-added in the 1993 Additional Series; it has not been revised for OED3.
 11. Excepting COLA, GA, MALLEE, MASSASAUGA, SATSUMA, SUI, and ZUÑI, which, as names for plants, animals, and peoples, should not have received tramlines according to OED1's usual practices. SUP2 was the first to document SATSUMA = 'small tangerine'.

12. The individual edition numbers don't sum to the OED2 total, as several SUP1 entries had their etymologies revised in SUP2.
13. I count roughly 20 Caribbean, 30 South Asian, and 30 East and South-East Asian senses constituted this way, but even to do so requires making the very inferences and judgements at issue, in this case by analogy to SUP2's practice with North American or Australian and New Zealand senses.
14. Author's email communication with OED staff. For instance, the vocabulary of certain sports and pastimes may be far more prevalent in the places where those activities are frequently practiced, but that per se does not make them regional.
15. As Gilliver notes, in the second volume of SUP2 Burchfield claimed to have given Caribbean sources 'somewhat more attention' than in the first. He was responding to a critical review in *The Times Literary Supplement*, which found Volume I lacking in its treatment of Caribbean sources and vocabulary (Gilliver, 495). This new attention to sources did not affect SUP2's labelling policy.
16. Importantly, in OED3, an etymology is given for every entry, including unrevised entries, whereas previous editions had left out etymologies deemed too obvious to spell out (Simpson 2000). Also, source languages have been standardized within a hierarchical structure that allows grouping of regions and linguistic families. The taxonomy is intuitive if idiosyncratic, sometimes grouping according to region, sometimes according to linguistic group (e.g., Altaic languages are within a much broader 'Central and Eastern Asian languages' class, and English is a superclass unto itself).
17. In explaining the Categories feature, OED3 says, 'If you want to find all the Japanese borrowings in English [...] this is the function for you' (<https://public.oed.com/how-to-use-the-oed/>).
18. Here and in the data I amalgamate a small number of labels referring to East Asian regions (e.g., 'Chinese English')—not a category in OED3's regional taxonomy—with OED3's South-East Asian category.
19. For instance, where OED2 had labels such as *dial.* or *Colonial*, OED3 has replaced them with region-specific labels, without making other revisions to the entry.
20. Importantly, not only the range of lexemes, but the range of evidence admitted by OED is restricted, as the regional Caribbean dictionaries all include oral evidence from interviewed informants, and OED does not (though it may quote the dictionaries themselves).
21. Used by Gower and appearing in the King James Bible, now it is used in Jamaica. Allsopp 1996 suggests that it was transmitted through Bible teaching (see, e.g., *KJV* Josh 11:10, 20:5; Isa 41:26; Act 8:9), rather than through the vernacular.
22. SUP2 does cite Cassidy 1961 s.v. LOOK, *v.*, but the sense is labelled as generally *dial.*. This is not atypical, as I discuss at the end of the previous section.
23. The question of antedating, though of perennial interest, is not straightforward when comparing OED3 with OED2, since a number of editorial and bibliographical emendations have masked antedatings or produced the illusion of antedating (e.g., when the publication date of a work is emended) or even a postdating (e.g., when an entry is divided, or the appropriateness of the quotation evidence re-evaluated). Nevertheless, with some curation of the data (described in Data Note D9), these factors can be mitigated so as to produce an informative, if partial, picture.

8. References

A. Dictionaries

- Allsopp, R. (ed.). 1996. *Dictionary of Caribbean English Usage*. Oxford University Press.
- Allsopp, R. (ed.). 2010. *New Register of Caribbean English Usage*. Jamaica: University of the West Indies Press.
- Avis, W. S. 1967. *A Dictionary of Canadianisms on Historical Principles*. Toronto: Gage.
- Branford, J. 1978. *A Dictionary of South African English*. Oxford University Press.
- Burchfield, R. W. (ed.). 1972. *Supplement to the Oxford English Dictionary*, vol. 1. Oxford University Press.
- Burchfield, R. W. (ed.). 1972–1986. *Supplement to the Oxford English Dictionary*. Oxford University Press. (SUP2)
- Burchfield, R. W. (ed.). 1986. *Supplement to the Oxford English Dictionary*, vol. 4. Oxford University Press.
- Cassidy, F. G. and R. B. Le Page. 1967. *Dictionary of Jamaican English*. Cambridge University Press.
- Craigie, W. and Onions (eds). 1933. *New English Dictionary on Historical Principles: Introduction, Supplement, and Bibliography*. Oxford: Clarendon Press. (SUP1)
- Collimore, F. A. 1957. *Notes for a Glossary of Words and Phrases of Barbadian Dialect*, 2nd ed. Bridgetown: Advocate.
- Macalister, J. (ed.). 2005. *A Dictionary of Maori Words in New Zealand English*. Oxford University Press.
- Murray, J. A. H. (ed.). 1888. *New English Dictionary on Historical Principles*, vol. 1, A–B. Oxford: Clarendon Press.
- Murray, J. A. H. H. Bradley, W. Craigie, and C. T. Onions, (eds). 1888–1928. *New English Dictionary on Historical Principles*. Oxford: Clarendon Press. (OED1)
- Proffitt, M., J. Simpson., E. Weiner (eds). 2000–. *OED Online*. <https://oed.com>. (OED3)
- Ramson, W. S. 1988. *Australian National Dictionary*. Oxford University Press.
- Silva, P. 1996. *A Dictionary of South African English on Historical Principles*. Oxford University Press.
- Simpson, J. and E. and (eds). Weiner 1989. *Oxford English Dictionary* (Second Edition). Oxford University Press. (OED2)

B. Other literature

- Benson, P. 2001. *Ethnocentrism and the English Dictionary*. London: Routledge.
- Brewer, C. 2007. *Treasure-House of the Language: The Living OED*. Yale University Press.
- Burchfield, R. W. 1989. *Unlocking the English Language*. London: Faber and Faber.
- Cassidy, F. G. 1961. *Jamaica Talk: Three Hundred Years of the English Language in Jamaica*. London: MacMillan.
- Curzan, A. 2000. 'The Compass of the Vocabulary' In Mugglestone, L. (ed.), *Lexicography and the OED: Pioneers in the Untrodden Forest*. Oxford University Press. 96–109.
- Dixon, R. M. W. 2008. 'Australian Aboriginal Words in Dictionaries: A History.' *International Journal of Lexicography* 21.2. 129–152.
- Dollinger, S. 2019. *Creating Canadian English: The Professor, the Mountaineer, and a National Variety of English*. Cambridge University Press.
- Gilliver, P. 2016. *The Making of the Oxford English Dictionary*. Oxford University Press.
- Kachru, B. B. (ed.) 1992. 'Teaching World Englishes' In *The Other Tongue. English Across Cultures*, 2nd edition, Urbana: University of Illinois Press. 355–366.

- McArthur, T. 1998. *The English Languages*. Cambridge University Press.
- Murray, K. M. E. 1997. *Caught in the Web of Words: James Murray and the Oxford English Dictionary*. Yale University Press.
- Ogilvie, S. 2008. 'Rethinking Burchfield and World Englishes.' *International Journal of Lexicography* 21.1. 23–59.
- Ogilvie, S. 2013. *Words of the World: A Global History of the Oxford English Dictionary*. Cambridge University Press.
- Orsman, H. W. (ed.). 1997. 'Introduction', *The Dictionary of New Zealand English: A Dictionary of New Zealandisms on Historical Principles*. Oxford University Press. vii–ix.
- Price, J. 2003. 'The Recording of Vocabulary from the Major Varieties of English in the *Oxford English Dictionary*' In Mair, C. (ed.), *The Politics of English as a World Language: New Horizons in Postcolonial Cultural Studies*. Amsterdam: Rodopi. 119–137.
- Ramson, W. S. 2002. *Lexical Images: The Story of the Australian National Dictionary*. Oxford University Press.
- Salazar, D. 2014. 'Towards Improved Coverage of Southeast Asian Englishes in the *Oxford English Dictionary*.' *Lexicography ASIALEX* 1. 95–108.
- Salazar, D. 2018. 'Philippine English in the October 2018 Update.' <https://public.oed.com/blog/philippine-english-in-the-september-2018-update/>.
- Sangster, C. 2016. 'Release Notes: World English Pronunciations.' *OED Blog*. <https://public.oed.com/blog/june-2016-update-release-notes-world-english-pronunciations/>.
- Silva, P. 2019. 'An Oxford Lexicographer of the 1960s: Penny Silva.' *OED Blog*. <https://public.oed.com/blog/an-oxford-lexicographer-of-the-1960s-penny-silva/>
- Sangster, C. 2020. 'Release Notes: West African English Pronunciations.' *OED Blog*. <https://public.oed.com/blog/release-notes-west-african-english-pronunciations/>.
- Silva, P. 2019. 'An Oxford Lexicographer of the 1960s: Penny Silva.' *OED Blog*. <https://public.oed.com/blog/an-oxford-lexicographer-of-the-1960s-penny-silva/>.
- Simpson, J. 2000. 'Preface to the Third Edition', *OED Online*. <https://public.oed.com/history/oed-editions/preface-to-the-third-edition/>.
- Sofield, C. M. 2018. 'Release Notes: The Many Faces of Energy.' *OED Blog*. <https://public.oed.com/blog/release-notes-the-many-faces-of-energy/>.
- Waddell, H. M. 1863. *Twenty-nine Years in West Indies & Central Africa*. London: Nelson & Sons.
- Williams, D.-A. 2017. 'Getting More Out of the Oxford English Dictionary (by Putting More In).' *Dictionaries: The Journal of the Dictionary Society of North America* 38.2. 106–113.
- Weiner, E. 1987. 'The new OED and World English.' *English Today* 3.3. 31–34.