Salazar, Danica. "Documenting World Englishes in the *Oxford English Dictionary*: Past Perspectives, Present Developments, and Future Directions." *Research Developments in World Englishes*. Ed. Alexander Onysko. London: Bloomsbury Academic, 2021. 271–294. *Bloomsbury Collections*. Web. 26 Apr. 2023. <<http://dx.doi.org/10.5040/9781350167087.ch-013>>.

Downloaded from Bloomsbury Collections, www.bloomsburycollections.com, 26 April 2023, 13:44 UTC.

Copyright © Alexander Onysko and Contributors 2021. Released under a CC BY-NC-ND licence (<https://creativecommons.org/licenses/by-nc-nd/3.0/>). You may share this work for non-commercial purposes only, provided you give attribution to the copyright holder and the publisher, and provide a link to the Creative Commons licence.

Documenting World Englishes in the *Oxford English Dictionary*: Past Perspectives, Present Developments, and Future Directions

Danica Salazar

1. Introduction

In 2018, the *Oxford English Dictionary* (*OED*) celebrated its 90th birthday—ninety years since its ten-volume first edition (*OED1*) was published in 1928. Throughout this time, and in the forty-nine years between 1879 and 1928 when James Murray and subsequent chief editors of the dictionary led a team of editors in the gargantuan task of producing instalment after alphabetical instalment of this ambitious and unprecedented lexicographical project, the *OED* has been held up as a landmark work of linguistic scholarship, and until now it is recognized as the undisputed authority on the lexical history of the English language. The *OED* is an historical dictionary of unparalleled size and scope, showing not only the current meanings of hundreds of thousands of words, but also tracing their chronological evolution through millions of quotations taken from written examples of authentic language use.

The *OED* is presently undergoing its first thoroughgoing revision and update since it was first published in full over ninety years ago. This endeavor employs over seventy editors based in Oxford and New York. The dictionary's third and latest edition is accessible through *OED Online* (oed.com), where updates are published quarterly. This makes the *OED* a work in progress, a hybrid text which combines unrevised entries from previous editions with fully revised and newly added entries. One of the key components of the dictionary's revision project is improving and enlarging its coverage of world Englishes, particularly of those varieties used outside of the United Kingdom and the United States.

This chapter begins with a chronological overview of how the *OED* has covered world Englishes throughout its long history. It will then discuss the important steps that the *OED* is currently taking towards providing a more balanced, representative, and authentic treatment of words used in a wider range of Anglophone settings, and conclude with some remarks on the future of world Englishes in the dictionary.

It should also be clarified at this point what is meant when we talk of world Englishes in the *OED*. In the *OED*, *world Englishes* is taken to mean those localized or indigenized

varieties of English spoken throughout the world by people of diverse cultural backgrounds in a wide range of sociolinguistic contexts. Although, strictly speaking, all the different varieties of English currently spoken can be considered world varieties of English, in the *OED*, the term *world Englishes* only refers to varieties other than standard British and American English, which are the two major varieties that have always been considered core English by most dictionaries. All other varieties—whether they are first-language regional varieties such as Canadian English or Australian English, or second-language regional varieties such as Nigerian English, Singapore English, or Indian English, or even British dialectal varieties such as Scottish English, Manx English, or Welsh English, or sociolectal varieties such as African American English—are within the scope of world Englishes in the *OED*. Pidgins and creoles based on English are not included, since the *OED* does not consider them varieties of English but rather as separate languages that can be a source of borrowings into English. For instance, the word *sefi* is recorded in the *OED* not as a Nigerian Pidgin English word, but as a Nigerian English word that has been borrowed from Nigerian Pidgin English, though the pidgin word itself may have ultimately come from the English word *safe*.

2. Putting the *OED*'s Coverage of World Englishes in its Historical Context

Fishman (1995: 34) urges us to “interpret dictionaries in context and see them as both resultant of and constructive of their contexts.” If we are to critique the *OED* in this way, we can examine it as a product of its very particular British origins. What has been the place of what we would now call world English vocabulary in what is, still today, considered by many people as a quintessentially British institution, and where does it figure in the dictionary's present and future?

The *OED* was born in the Victorian era, a time when the homogenization and standardization of language was viewed as the main function of a dictionary, in line with the predominant scientific thinking of the period. Benson (2001: 4) argues that the “English dictionary is a historically specific form of discourse embedded within broader discourses that represent knowledge of the world in terms of metaphors of centre and periphery,” with the center generally being understood as standard British English and the periphery as everything else outside this standard. The *OED*, despite the undeniable rigor of its historical methodology, cannot wholly escape criticisms for this perceived ethnocentrism.

James Murray himself, in his *General Explanations* that prefaces the first fascicle of *A New English Dictionary on Historical Principles* (*NED*), the very first incarnation of the *OED*, uses the center-periphery metaphor in describing the scope of English. He presents a diagram (Figure 13.1) in which he identifies a core of English words whose “Anglicity is unquestioned” (Murray 1884: xvii). Outside this core is what we would consider today to be the bulk of world Englishes lexis—dialectal and foreign words—words beyond what was then recognized as Standard British English.

Figure 13.1 Murray's Model of the English lexicon (1884)

Before ascribing the apparent ethnocentrism of Murray's model to his deliberate neglect of non-British cultures, it is important to note the contemporary evidence that shows Murray having staunchly defended the place of foreign and dialectal words in the *OED* (Ogilvie 2012). His correspondence from this period shows how much resistance he faced from his subeditors and superiors at Oxford University Press, as well as from consultants and reviewers, when he decided to include words such as the Philippine *abaca*, the South African *aardvark*, and the Indian *amah* in the first fascicle. The presence of these words in the *OED* serves as proof of Murray's commitment to including this type of vocabulary in the dictionary, and it is indeed during his decades-long tenure as Chief Editor that a number of Anglo-Indian words (*bungalow*, *chit*, *dal*), Southeast Asian terms (*gong*, *rattan*, *sarong*, *wat*), and other lexical items of similarly foreign origin (*chop-stick*, *ginseng*, *meerkat*, *okra*) made their way into the *OED* (Salazar 2018a).

What is more meaningful to observe about these words is the way that these non-British items were presented entirely through the lens of Victorian Britain, as a response to the needs and interests of the British Empire. A good example is the *OED* entry for the aforementioned *abaca* (Figure 13.2), which appears on page 5 of the 1884 first fascicle.

Despite being a name for a plant from halfway across the world from Britain, *abaca* managed to enter the *OED* in the late nineteenth century because of the strong natural fiber yielded by the plant and the objects that can be made from it: paper, rope, matting, hats. It is obvious that *abaca*'s Philippine origin is peripheral to the entry, and that the word made it into the dictionary by virtue of its significance to British industry, being

Figure 13.2 Entry for *Abaca*, n. in the First Fascicle of *A New English Dictionary on Historical Principles* (1884, 5)

a raw material used in making everyday objects. The selection of words of foreign origin in the *NED* fascicles, published between 1884 and 1928, and in the first edition of the *OED*, published in 1928, reflects the interests and concerns of the British of the late nineteenth and early twentieth centuries: the native flora and fauna of their colonies in Asia, Africa, and the Caribbean; exotic local terms that British travellers and settlers pick up in these overseas territories; the names of the vast array of products that they import from every corner of the globe (Salazar 2018a). These words were all welcomed into the dictionary, but their alien status was marked in the fascicles and in the complete *OED1* by two parallel lines beside the headword (||), a symbol known by the in-house term *tramline*.

The tramline method of classification was one that even Murray admitted was inherently subjective, and it was perhaps a desire to avoid inconsistencies in the use of this symbol that the Chief Editors of the 1933 Supplement to the *OED*, Charles Talbot Onions and William Craigie, decided to stop using them altogether (Brewer 2007), as they opened the dictionary's doors to even more foreignisms (*kumquat*, *bushveld*, *impala*, *safari*, *tango*, to name a few).

It should also be considered that the editorial choices made by Murray, Craigie, Onions, and their staff were significantly limited by the chiefly British and American sources they had access to in Oxford during their time. These limitations led to a certain bias in the *OED* towards loanwords relating to flora, fauna, ethnic groups, and other items of mostly historical and anthropological interest. Bolton and Butler (2008) call this type of vocabulary “Webster words,” since they observed their prevalence not only in the *OED* but also in various editions of *Webster's*.

The mid- and late twentieth century brought new ideas and new theoretical frameworks in mainstream linguistics that led to a new conception of dialectal variation in English. Linguistic variation came to be seen as a natural consequence of cultural diversity, and it came to be understood that a language such as English that had been transplanted into new environments tended to adapt to these environments and that national or regional varieties of languages served to express national or regional identities. In a post-imperialist world where the globalization of English was

an accepted fact, regional Englishes spoken throughout the world began to be studied as varieties in their own right, and became known collectively as world Englishes (Kachru 1986; McArthur 1987; Görlach 1990).

The *OED* responded to this paradigm shift by looking further into the written English of Anglophone regions outside the British Isles in the four-volume Supplement edited by Robert Burchfield. The Burchfield Supplements, published in the 1970s and 80s, added to the *OED* a wider variety of words that expressed various aspects of life in their places of origin (*kebaya, gamelan, sambal, nasi goreng, satay, jeepney*). However, these Supplements also marked a return to the use of tramlines to distinguish foreign words, a policy that was once again dropped in the twenty-volume Second Edition (*OED2*) that came out in 1989.

In a 1987 article, Edmund Weiner, one of the Chief Editors of the *OED*'s 1989 Second Edition, acknowledged the ad hoc treatment of regional English items in *OED1* and its subsequent supplements, and discussed possible ways of applying a more articulated policy in *OED2*. Finally, in 2000, with the launch of the Third Edition (*OED3*), then Chief Editor John Simpson explicitly recognized the status of English as a world language, and the relegation of British English from being the only standard to being one of many to be covered by the dictionary:

From its base in Britain, the English language has expanded over the centuries to become a world language, in which individual varieties share a common core of words but develop their own individual characteristics. [...] The English of the British Isles now becomes one (or indeed several) of these varieties, whereas previously standard British English may have been regarded as the dominant form of English.

Simpson 2000

The *Financial Times* put it succinctly in a 2016 piece when it said, “the *OED* finds itself in the curious position of being a national institution called upon, almost by default, to assume the role of a global one.”

In recent years, the *OED* has undertaken targeted projects to expand its record of several Englishes, publishing particularly large batches of new entries for varieties spoken in Canada, Hong Kong, India, Malaysia, Nigeria, the Philippines, Singapore, and South Africa. The following section will explain the principles and motivations behind the dictionary's current efforts in documenting world Englishes.

3. Current Policies and Practices on World Englishes in the *OED*

The lexicographical documentation of world Englishes is a complex undertaking made even more challenging by the fact that most people still view dictionaries as representations in paper and ink, or in screens and pixels, of a monolithic ideal of a standard language. This conceptualization of a dictionary as the definitive arbiter on correct usage puts into question the necessity, or even the propriety, of including in it those regional, colloquial, and non-standard words that characterize world Englishes lexis (Dolezal 2006).

However, as Seargeant (2011) suggests, the very linguistic ideology which has made dictionaries a powerful tool for language standardization can itself be used in changing long-held prejudices against emerging varieties of English. Adding world Englishes words to a dictionary, especially one as well-known and respected as the *OED*, serves as a recognition of their status within the English lexicon, and gives the people who use them a sense of belonging to a distinct language community with its own idiosyncratic vocabulary (Dolezal 2006). As Butler (1997) puts it, “being able to say that your words are in a dictionary [...] in general has more effect on the public perception of the validity of a variety than many words spent in linguistic theorizing.” The *OED*’s reputation as a linguistic authority can turn it into an instrument for promoting diversity; by recognizing the plurality of centers of English, it can push varieties formerly in the periphery towards the center of language scholarship (Salazar 2015).

The inclusion of lexical items from world Englishes in the *OED* also gives them their rightful place in the historical record of English and contributes significantly to World Englishes research by providing crucial documentary evidence of their chronological development. Just like all other *OED* entries, entries from world Englishes have undergone a rigorous and exhaustive research process that identifies their earliest use in written texts and distills all available information on how their meaning and usage have evolved over time. All this historical dictionary data can be used for further academic inquiry on the lexis of world Englishes, and serve as proof of the longevity and stability of these words, and of the varieties of English to which they belong (Lambert 2020).

3.1 More Balanced and Representative Word Selection for World Englishes

In order to fully realize the benefits of greater inclusion of world Englishes, the *OED* has been making necessary adjustments to further broaden its geographical scope and ensure that its entries that fall outside of the lexical core of English are as accurate and reflective of actual usage as possible.

One of the biggest changes in modern lexicography is in the way dictionary evidence is gathered. Unlike with their predecessors in the nineteenth century and early twentieth century, the research done by today’s *OED* editors is no longer restricted to published Anglo-American texts that they have physically on hand. The Internet has given twenty-first-century lexicographers greater access to non-UK and non-US sources of linguistic evidence by making them available electronically. They can also make use of other digital tools and resources designed specifically for linguistic investigation, such as Oxford Languages’ massive in-house language corpora, which contains billions of words and has variety-specific segments from which to generate frequency counts that help *OED* editors identify and prioritize emerging world Englishes words for inclusion. They can also use other corpora of world Englishes, such as the GloWbE corpus (Global Web-based English Corpus, Davies and Fuchs 2015). They can consult the increasingly large body of scholarly work on the lexis of world

Englishes, as well as collaborate with an international network of consultants working on various world Englishes. They are also able to launch crowdsourcing initiatives in order to ask for word suggestions or antedatings from the very speakers of world Englishes, as it did with a crowdsourced batch of Philippine English words it published in October 2018 (Salazar 2018b), or as it does regularly through the *OED Appeals* website (<https://public.oed.com/appeals/>).

All of these new research tools and resources have allowed the *OED* to focus its attention on more contemporary lexis, and mechanisms of lexical innovation beyond simple borrowing. As Dolezal (2006: 702) observes, “a language variety is not the sum of its ‘exotic’ parts,” and the *OED* is now more able to provide coverage of world Englishes that not merely amounts to a listing of loanwords referring to native plants, animals, and cultural artefacts, but encompasses all forms of lexical innovation, including a diverse range of strategies of word creation such as semantic change, fossilization, loan translation, loan blending, conversion from one part of speech to another, affixation, clipping, initialism, and compounding (Table 13.1).

New sources and methods for collecting lexical data have also enabled the *OED* to add more words from semantic fields that are of particular interest to actual speakers of world Englishes, not just to British or American lexicographers or linguistic anthropologists looking at a variety from an outsider perspective. The dictionary has recently included more words from such culturally specific domains as food and drink, places to eat, kinship terms, and informal modes of transport (Table 13.2).

These lexical innovations illustrate how users of world Englishes, who usually speak English in addition to one or more other languages, modify English in order to express in this adopted tongue their own local traditions, values, and social norms. For instance, Asian speech etiquette is characterized by a complex system of kinship terms and forms of address in which age, gender, status, and family relationships are marked by a highly specific vocabulary with no direct equivalents in English, forming a lexical gap that is filled by borrowing such words from local languages or adapting existing English words (Salazar 2017).

Now more than before, words from world Englishes are considered for inclusion in the *OED* on the basis of their salience and currency in a particular geographical area or language community, regardless of their diffusion beyond this area or community. Lexical items that may be thought of as peripheral in Standard British or American English and form no part of the vocabulary of British or American English speakers are also given space in the dictionary and defined according to local usage. For example, the *calamansi* (Figure 13.3), known in the West by the name *calamondin* as a decorative citrus plant with orange fruits, is to Filipinos a green fruit that is as essential to their cooking as the lemon is in many other parts of the world. *Calamansi* has therefore been incorporated in the *OED*, with a small-type note acknowledging its significance in Filipino cookery. Another example is *HDB* (Figure 13.4), an initialism that is little known outside of Singapore, but is ubiquitous in the city, being the name of the government-built high-rise flats that most Singaporeans live in.

Table 13.1 Examples of world Englishes words recently added to the *OED*, created using different processes of word formation

Word	Variety	Definition	Process of word formation	Date of publication in <i>OED</i>	Date of first quotation
sarmie, noun	South African English	A sandwich.	Affixation	2018	1970
agric, adjective	Nigerian English	Designating an improved or genetically modified variety of crop or breed of livestock.	Clipping	2019	1992
timepass, noun	Indian English	An activity or interest that serves to pass the time agreeably but unproductively; an idle distraction.	Compounding	2017	1982
comfort room, noun	Philippine English	A room in a public building or workplace furnished with amenities such as facilities for resting, personal hygiene, and storage of personal items (now rare); (later) a public toilet.	Fossilization	2015	1886
HDB, noun	Singapore English	Housing and Development Board; used chiefly attributively with reference to public housing estates built and managed by the Singapore government.	Initialism	2016	1961
add oil, phrase	Hong Kong English	Expressing encouragement, incitement, or support: go on! go for it!	Loan translation	2017	1980
chakka jam, noun	Indian English	The blocking of a road or the deliberate creation of a traffic jam as a form of civilian protest; a blockade.	Loan blending	2017	1972
blur, adjective	Singapore English	Slow in understanding; unaware, ignorant, confused.	Part of speech conversion	2016	1977
colony, noun	South Asian	A housing estate constructed by an employer or organization for its workers. In later use also: any housing estate or residential community, esp. one surrounded by a closed perimeter of fences and walls.	Semantic change	2017	1886

Table 13.2 Examples of world Englishes words recently added to the *OED* from different culturally specific semantic fields

Semantic field	Word	Variety	Definition	Date of publication in <i>OED</i>	Date of first quotation
Food and drink	bunny chow	South African English	A South African dish consisting of a hollowed-out loaf of bread (or part of a loaf) filled with curry, typically sold as takeaway food; a serving of this.	2018	1972
	char siu	Hong Kong, Singapore, Malaysian English	In Cantonese cookery: roast pork marinated in a sweet and savoury sauce, typically served sliced into thin strips.	2016	1952
	chilli crab	Singapore, Malaysian English	A dish originating in Singapore but also popular in Malaysia, consisting of crab cooked in a sweet and spicy gravy containing red chillies and tomato.	2016	1963
	laksa	Singapore, Malaysian English	In Peranakan cookery: any of various types of spicy noodle soup made with meat, vegetables, or fish.	2019	1846
	milk tea	Hong Kong English	Any of various drinks made with tea and milk or cream; esp. a drink originating in Hong Kong, made with black tea and evaporated or condensed milk.	2016	1897
	yum cha	In Chinese contexts	A meal eaten in the morning or early afternoon, typically consisting of dim sum and hot tea.	2016	1936
Places to eat	buka	Nigerian English	A roadside restaurant or street stall with a seating area, selling cooked food at low prices.	2019	1972
	carinderia	Philippine English	A food stall with a small seating area, typically in a market or at a roadside, selling cooked food at low prices.	2018	1910
	dai pai dong	Hong Kong English	A traditional licenced street stall, typically with a small seating area, selling cooked food at low prices; (now more generally) any food stall of this type.	2016	1983
	dhaba	Indian English	A roadside food stall or restaurant.	2017	1886
	hawker centre	Singapore, Malaysian English	A food market at which individual vendors sell cooked food from small stalls, with a shared seating area for customers.	2016	1966

(Continued)

Table 13.2 *Continued*

Semantic field	Word	Variety	Definition	Date of publication in OED	Date of first quotation
Kinship terms or forms of address	ate	Philippine English	An elder sister. Also used as a respectful title or form of address for an older woman.	2016	1937
	bapu	Indian English	One's father; a father. Often as a familiar form of address.	2017	1930
	cousin sister	South Asian, Australian English	(a) South Asian: a female cousin of one's own generation; (b) Australian (in Australian Aboriginal usage): a closely related female relative of one's own generation.	2017	1881
	lola	Philippine English	One's grandmother. Also used as a respectful title or form of address for an elderly woman.	2016	1934
	mamak	Malaysian English	A maternal uncle or elder brother; the male head of a household. Also used as a respectful title or form of address for an older man.	2016	1884
Modes of transport	danfo	Nigerian English	A yellow minibus that carries passengers for a fare as part of an informal transport system in Lagos, the most populous city in Nigeria.	2019	1973
	jugaad	South Asian	A makeshift automobile constructed from inexpensive materials.	2017	1995
	okada	Nigerian English	A motorcycle which passengers can use as a taxi service.	2019	1993

Figure 13.3 OED3 definition for *calamansi, n.*, first published in 2018

Figure 13.4 OED3 definition for *HDB, n.*, first published in 2016

3.2 Pronunciation Models and Audio Pronunciations for World Englishes Entries

The *OED* has also formulated specific pronunciation models for several world Englishes (see Table 13.3 for a complete list), which can all be consulted on the *OED* website (<https://public.oed.com/how-to-use-the-oed/key-to-pronunciation/pronunciations-for-world-englishes/>). Each pronunciation model has a dedicated page with a discussion of the rationale for the model, along with major reference sources and potential discrepancies and compromises. These models were developed using pronunciation data drawn from varietal dictionaries, knowledgeable native speakers, and online evidence of natural language use. They form the basis of variety-specific transcriptions in the International Phonetic Alphabet (IPA) for world Englishes entries, accompanied by recorded audio pronunciations. For instance, in the *OED*’s entry for the South African English word *skedonk* (Figure 13.5), there are the IPA transcriptions and audio pronunciations in British and American English that are standard for all non-obsolete *OED* entries, but alongside these is the IPA transcription in South African English, and a blue button that, when clicked, will play a recording of the word being pronounced in a South African accent, as determined by the *OED*’s pronunciation model for South African English.

The audio pronunciations are not synthesized, but are recordings made by actor-phoneticians recruited for each variety, who came to OUP’s recording facilities in

Table 13.3 List of varieties of English for which the *OED* has a Pronunciation Model (as of July 2020)

Australian English	Manx English	South African English
Canadian English	New Zealand English	Welsh English
Caribbean English	Philippine English	West African English
Hong Kong English	Scottish English	
Irish English	Singapore and Malaysian English

Figure 13.5 IPA transcriptions and audio pronunciations for the South African entry *skedonk*, n. in British English, American English, and South African English

Oxford to read each transcription aloud. The *OED*'s review of pronunciations for world Englishes is a work in progress, and pronunciation models for more varieties, along with transcriptions and audio recordings for corresponding entries, will continuously be added to the dictionary by its Pronunciations team.

3.3 More Representative Illustrative Quotations for World Englishes

The *OED* now places utmost importance on ensuring that words from world Englishes in the dictionary are illustrated by quotations from works written by speakers of the varieties in question, and published in their place of origin. Consider this entry for *buka*, a Nigerian English word for a roadside restaurant or food stall (Figure 13.6).

It can be seen that all the quotations for this entry come from Nigerian sources, from its earliest attested use in 1972 in a historical novel written by Nigerian author Kole Omotoso, to a 1991 quotation taken from Nigerian writer Ben Okri's Booker Prize-winning novel *The Famished Road*, to messages posted by Nigerians on the social media platform Twitter.

Figure 13.6 *OED* entry for *buka*, n.

Although, as previously stated, the *OED* now has access to a more geographically diverse range of research resources, it remains true that the amount of lexicographical evidence available for region-specific vocabulary is still very small compared to that for general English. This is compounded by the fact that publications in postcolonial nations continue to be edited following British or American standards, so that lexical innovations in speech may not always make it into the written texts that the *OED* regularly consults (Salazar 2014). It is for this reason that *OED* editors are increasingly turning to less mediated forms of writing, such as tweets and song lyrics, in order to find quotations that more closely approximate the way that speakers of world Englishes talk, and therefore serve as more authentic illustrations of how words from world Englishes are used in real-life contexts.

The *OED* now recognizes that in most places where English is spoken as a second language, code-switching is the dominant form of communication, a fact that is apparent even in the written register. See, for instance, the following antedating for the Philippine English adjective and noun *kilig* that was brought to the dictionary's attention via a recent *OED* Appeal for antedatings (Figure 13.7).

Previously, an *OED* editor might have dismissed this quotation as an unusable Taglish sentence, but the quotation has now been added to the entry, antedating *kilig* by over ten years, from 1994 to 1981. The *OED* considers this as an acceptable quotation written not in Taglish but in Philippine English, as the base language of this sentence is actually English, with Tagalog words inserted. These insertions are not due to the writer's lack of competence or their "non-nativeness"—they are there for discursual/discursive reasons. They were valid choices made by a multilingual speaker from a lexical repertoire that also happens to include words from their other languages; they were choices that the speaker deemed appropriate given their interlocutor and what they were aiming to communicate.

Moving world Englishes from the periphery to the center in the *OED* entails not just documenting their idiosyncratic lexicons. It also means using more sources written in world Englishes to illustrate general English words. Just as works of classic English literature and the most important of British and American newspapers appear regularly in the dictionary's quotation paragraphs to show the development of everyday English

Figure 13.7 Extract from *OED* entry for *kilig*, adj. and n.

words, so should works of postcolonial writers in English, and newspapers read in other parts of the English-speaking world. The *OED* is addressing this by including more world Englishes publications to its Reading Program, so that its team of volunteer readers can add more quotations from sources outside the United Kingdom and the United States to its digital collection of sentences or short extracts.

3.4 More Informative Labels and Usage Notes

The *OED* aims to make its system of geographical labelling clearer and more useful, and to provide more discursive information to accompany these labels. See for instance, this unrevised entry for *jeepney* (Figure 13.8), originally published in the 1976 Supplement. The entry contains really sparse information that does not tell the reader much about the jeepney's significance in Philippine transport.

Compare this to the new entry for the Filipino dish *sisig* (Figure 13.9), which includes, not just discursive metatext relating it to Filipino cuisine, but also an additional small-type note that tells the reader more about the origins of the dish, its possible variations, when it is eaten, and even cross-refers to the Filipino food custom of *pulutan*, for which the *OED* also has an entry (Figure 13.10).

Figure 13.8 *OED* entry for *jeepney, n.*

Figure 13.9 Extract from *OED* entry for *sisig, n.*

Figure 13.10 Extract from *OED* entry for *pulutan, n.*

4. Using *OED Online* to Investigate World Englishes

Lambert (2020) notes how the word “dictionary” often still brings to mind the traditional conceptualization of a heavy book containing an alphabetically ordered list of words, when in fact, a more useful and accurate way of viewing a twenty-first-century dictionary is as an online digital resource, “an amazingly versatile and highly adaptable platform for the presentation of linguistic information,” and, for World Englishes scholars in particular, “a valuable avenue [...] to display research.”

The same can be said specifically of the *OED*. Although the mention of its name conjures images of handsomely bound volumes on a library shelf, the *OED* has always been a pioneer in the process of digitalization that dictionaries have undergone in the past few decades. The ground-breaking computerization of the *OED* began as early as 1984, with the painstaking conversion of the decades-old print dictionary into machine-readable format so that the first edition could be electronically combined with the Burchfield supplements, along with thousands of new entries, to form the twenty-volume Second Edition published in 1989. A few years later, in 1992, the *OED* made history once again when it became available on CD-ROM, and yet again in 2000 with the launch of *OED Online*, which made the *OED* one of the first reference works of its kind to have an online presence (Gilliver 2016).

The digital transformation of the *OED* has revolutionized not only the way the dictionary is compiled, but also how it is used. It has made the *OED* fully searchable, thus freeing users from the limitations imposed by the alphabetical order and enabling them to mine the dictionary’s rich trove of lexical data in ways that can be tailored to various scholarly purposes (Brewer 2004). Simpson (2013) points to how the introduction of the electronic *OED* has shifted the focus of interest from single-word lookups to more general investigations of linguistic trends.

The profound impact of digitalization on the *OED* also extends to world Englishes. There are relatively few studies on the vocabulary of world Englishes, because lexical studies require massive language corpora, which are currently unavailable for varieties other than British and American English. The *OED*, as a massive depository of diachronic linguistic data, is uniquely able to fill this research gap. The digital, online format of its vast and varied content is particularly suited for uncovering patterns of word use that can help inform investigations of linguistic and extralinguistic phenomena that are of interest to World Englishes researchers.

4.1 Advanced Search

Using the Advanced Search function on *OED Online*, users can apply different filters that narrow down searches for world Englishes entries matching a particular criterion or set of criteria. For example, a search for words in the *OED* that originate in Indian subcontinent languages shows that of the 1,469 entries in the dictionary that come from languages such as Bengali, Hindi, Sanskrit, Tamil, and Urdu, 261 (18%) belong to the subject category Religion and Belief (Figure 13.11a). Further limiting the search results to this subject category and examining the list that this brings up reveals the nature and significance of these South Asian words: they are words relating to principles of Buddhist and Hindu philosophy and mythology (e.g., *jiva*, *prajna*, *samsara*, *satya*),

a

b

Figure 13.11a/b Advanced Search results for words from Indian subcontinent languages in the *OED* (11a) / Advanced Search results for words from Indian subcontinent languages in the *OED*, limited to the subject category religion and belief

or to specific practices of these religions such as yoga (e.g., *asana*, *ashram*, *ashtanga*, *hatha yoga*, *vinyasa*) and religious festivals (e.g., *Diwali*, *Holi*, *Kumbh Mela*). Many of these words have gained such currency in English that they have developed extensions of meaning and become part of everyday vocabulary (e.g., *avatar*, *guru*, *juggernaut*, *karma*, *mantra*, *nirvana*, *pundit*). It can be seen from this example how a refined OED search can give insight into the dominant semantic fields in a particular variety of English as recorded lexicographically.

The Advanced Search function can also identify which varieties of English have borrowed from certain substrate languages, and conversely, which substrate languages have had the most lexical influence on a particular variety. For instance, launching an advanced search for words coming from African languages (Figure 13.12a) reveals that unsurprisingly, African varieties have borrowed several words from these languages, but so have Caribbean English and North American English; a similar advanced search for words used in the Caribbean region (Figure 13.12b) shows that Caribbean vocabulary mostly comprises words of English origin and loanwords from European languages such as Spanish and Portuguese, but also includes words that have been borrowed from African languages, Indian languages, and Native American languages, as well as from Caribbean creoles.

Figure 13.12a/b Advanced Search results for words from African languages in the OED (12a) / Advanced Search results for words from the Caribbean in the OED (12b)

4.2 Timelines

Another search feature of *OED Online* is Timelines, which enables a user to generate graphs that show the number of words first recorded by the *OED* within different time periods. Timeline searches can also be limited by various subject, region, and origin categories. An example is the graph in Figure 13.13 illustrating the patterns of borrowing into English from Niger-Congo languages over five centuries, which should be of interest to researchers of various African Englishes who wish to determine the points in time in which there has been the most contact between English and the Niger-Congo languages that form the substrate of many African varieties of English. It can be observed from the graph that borrowing began with the oldest Niger-Congo word recorded by the *OED*, the ethnonym *Kongo*, which is first attested in a 1597 English translation of a book written by Portuguese merchant Duarte Lopes, one of the earliest European accounts of Central Africa; peaked in the late nineteenth century, a period of accelerated British colonial expansion in Africa; and continues to today, with sustained lexical borrowing of Niger-Congo words into English by African speakers who use English alongside their native Niger-Congo languages.

Figure 13.13 OED Timeline graph for words originating in Niger-Congo languages

4.3 Insights from Individual OED Entries

Even the traditional single-word look-up can still be a source of valuable information for World Englishes research. One simple but fundamental question for lexical studies of world Englishes that can be answered by simply consulting an *OED* entry is whether a word or sense that has been found to be characteristic of a particular variety is indeed a new word or sense that developed within this variety.

Mesthrie (2003) observes that most research in World Englishes has been on the transfer from substrates to New Englishes, and is based on comparisons between contemporary world Englishes corpora and either a British or American English corpus. Mesthrie also cites a study by Davy (2000), who used the *OED* to show that several lexical and grammatical features of African English actually have a long history in British English. Both Mesthrie and Davy question whether current British and American English is always the relevant superstrate to compare to the substrate, and caution against always taking deviations from contemporary British and American English corpora as sufficient evidence for what Davy calls linguistic or normic newness.

One typical example of a word that many people consider to be unique to Philippine English is *comfort room*, which is one of the most commonly used terms in the Philippines to refer to a toilet. However, the *OED* entry for *comfort room* (Figure 13.14) contains evidence of this compound being used in the United States as early as 1886, continuing to be used by Americans during the early years of the American occupation of the Philippines, thus giving an indication of when the term may have been picked up by Filipinos, and then falling out of use in America in the mid-twentieth century and remaining in current use only in the Philippines. The *OED* provides convincing proof that *comfort room* is not a Philippine coinage but rather an archaic usage in American English preserved in contemporary Philippine English.

OED entries also show how spelling conventions have changed for certain words from world Englishes. This is particularly interesting for older loanwords that have been borrowed into English from other languages during a time before firm orthographic rules for such loanwords had been set. The extract from the entry for *chaudhuri*, *n.* (Figure 13.15) lists the nineteen different spellings that *OED* lexicographers have found for this nearly 250-year-old word.

Figure 13.14 *OED* entry for *comfort room*, *n.*

The concentration of large teal and blue bubbles in the upper left side of the graph implies that the majority of the words in this extract are very old Germanic or English words belonging to the common core of the English lexicon shared by many varieties of the language—function words such as articles, prepositions, and conjunctions, but also highly frequent content words such as *see*, *up*, *want*, and *spot*. Words that are specific to Singapore, such as *ang moh*, *aiyoh*, *Ah Beng*, and *lah*, are represented by small yellow dots in the lower right side of the graph, indicating the newness of these words and their relative infrequency in the text. Based on the lexical evidence presented in this graph, it can be argued that although *Sarong Party Girls* provoked considerable discussion upon its publication for being a novel in Singlish, it is not in fact written in Singlish but in Singapore English. The text is largely composed of high-frequency words found in most English writing, but is also characterized by a smaller set of less frequent but highly distinctive and contemporary Singapore English vocabulary that gives life to the novel's characters and places them within the milieu of twenty-first-century Singapore.

While still in beta version, the OED Text Visualizer is only optimized to annotate texts of up to five hundred words written after 1750. A fully optimized version of the tool without these limitations is currently under development, and will be made available soon. With further improvement of the Text Visualizer, it will become an even more effective means of carrying out diachronic lexical analyses of long stretches of actual discourse in world Englishes.

5. The Future of World Englishes in the OED

As the OED carries on with its work of documenting the lexicon of world Englishes, there remains much room for critique, for new ideas for resources and technologies, and for new ways to use the vast amounts of data stored in the dictionary. World Englishes provide particularly fertile ground for innovation in the OED, as they bring up a variety of research questions that lead to productive re-examinations of its practices and policies.

At the time of writing, the OED is preparing for the imminent publication of large batches of new words and revised entries for Bermudian English, Caribbean English, Irish English, Kenyan English, Tanzanian English, Ugandan English, and African American English, while continuing to monitor developments in other varieties of English.

Looking to the future, the OED's goal for world Englishes is to fulfil the needs of its increasingly global audience through wide-ranging, inclusive, and representative coverage of their words and usage. This can only be achieved through collaboration with local partners, and so the OED is introducing several new ways of engaging with the World Englishes research community. There is the OED Researchers Advisory Group (<https://oed-group.oxfordlanguages.com/>), an invitation-only group for academics who use the OED in their research, and are willing to participate in surveys, forum discussions, and other tasks within the RAG interface in order to share their ideas and expertise relating to different aspects of the dictionary's ongoing projects on world Englishes.

There is also *OED Labs*, an initiative which invites scholars of various fields to contribute to the development of new functionalities for the exploration of *OED* content (<https://languages.oup.com/research/oed-labs/>), such as the Text Visualizer described above, and the *OED Researcher API*, a tool that enables researchers to make use of datasets extracted from across the *OED* to support research in a wide variety of disciplines.

With every new development, the *OED* is evolving into something much more than a dictionary. It is now becoming a dynamic and flexible lexical database that can be a starting point or a regular stop in the research journey of scholars seeking to investigate world Englishes lexis. By adding new words, revising existing entries, and developing better tools and methodologies, the *OED* is increasing the quantity and further improving the quality of the lexical data it contains for world Englishes, so that it can be an even bigger, more balanced, more representative, and more reliable dataset that can help researchers answer a variety of questions on how English is used by different communities around the world.

References

- A New English Dictionary on Historical Principles: Part 1: A-Ant.* 1884. Oxford: Oxford University Press.
- Benson, P. (2001), *Ethnocentrism and the English Dictionary*, London and New York: Routledge.
- Bolton, K. and S. Butler (2008), "Lexicography and the description of Philippine English vocabulary," in M. L. S. Bautista and K. Bolton (eds.), *Philippine English: Linguistic and Literary Perspectives*, 175–200, Hong Kong: Hong Kong University Press.
- Brewer, C. (2004), "The electronification of the *Oxford English Dictionary*," *Dictionaries: Journal of the Dictionary Society of North America*, 25: 1–43.
- Brewer, C. (2007), *Treasure-House of The Language: The Living OED*. New Haven and London: Yale University Press.
- Butler, S. (1997), "Selecting South-east Asian words for an Australian dictionary: How to choose in an English not your own," in E. W. Schneider (ed.), *Englishes Around the World Vol. 2. Caribbean, Africa, Asia, Australia. Studies in Honour of Manfred Görlach*, 273–86, Amsterdam: John Benjamins.
- Davies, M. and R. Fuchs (2015), "Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE)," *English World-Wide*, 36 (1): 1–28.
- Davy, J. (2000), "A conservative view of the New Englishes," Paper presented at the *First International Conference on Linguistics in Southern Africa*, January 12–14, University of Cape Town.
- Dolezal, F. (2006), "World Englishes and lexicography," in B. Kachru, Y. Kachru and C. Nelson (eds.), *The Handbook of World Englishes*, 694–708, Oxford: Blackwell.
- Fishman, J. A. (1995), "Dictionaries as culturally constructed and culture-constructing artifacts: The reciprocity view as seen from Yiddish sources," in B. Kachru and H. Kahane (eds.), *Cultures, Ideologies, and the Dictionary: Studies in Honour of Ladislav Zgusta*, 29–34. Tübingen: Max Niemeyer Verlag.
- Gilliver, P. (2016), *The Making of the Oxford English Dictionary*, Oxford: Oxford University Press.

- Görlach, M. (1990), *Studies in the History of the English Language*, Heidelberg: Carl Winter.
- Kachru, B. (1986), "The power and politics of English," *World Englishes*, 5 (2/3): 121–40.
- Lambert, J. (2020), "Lexicography and World Englishes," in D. Schreier, M. Hundt, and E. W. Schneider (eds.), *The Cambridge Handbook of World Englishes*, 408–35, Cambridge: Cambridge University Press.
- McArthur, A. (1987), "The English languages?," *English Today*, 11: 9–13.
- Mesthrie, R. (2003), "The World Englishes paradigm and Contact Linguistics: Refurbishing the foundations," *World Englishes*, 22 (4): 449–61.
- Murray, J. A. H. (1884), "General explanations," in *A New English Dictionary on Historical Principles: Part 1: A–Ant*, xvii–xxiv. Oxford: Oxford University Press.
- Ogilvie, S. (2012), *Words of the World: A Global History of the Oxford English Dictionary*, Cambridge: Cambridge University Press.
- Salazar, D. (2014), "Towards improved coverage of Southeast Asian Englishes in the *Oxford English Dictionary*," *Lexicography*, 1: 95–108.
- Salazar, D. (2015), "The vocabulary of non-dominant varieties of English in the *Oxford English Dictionary*," in R. Muhr, D. Marley, A. Bissoonauth-Bedford and L. Kretzenbacher (eds.), *Pluricentric Languages Worldwide and Pluricentric Theory*, 11–27. Vienna: Peter Lang.
- Salazar, D. (2017), "Release Notes: Indian English," *OED Blog*. Accessed July 26, 2020. <https://public.oed.com/blog/september-2017-update-release-notes-indian-english/>.
- Salazar, D. (2018a), "From 'Abaca' to 'Kilig': World English and the *OED*," *OED Blog*. Accessed July 12, 2020. <https://public.oed.com/blog/abaca-kilig-world-english-oed>.
- Salazar, D. (2018b), "Philippine English in the October 2018 Update," *OED Blog*. Accessed July 26, 2020. <https://public.oed.com/blog/philippine-english-in-the-september-2018-update/>.
- Seargeant, P. (2011), "Lexicography as a philosophy of language," *Language Sciences*, 33: 1–10.
- Simpson, J. (2000), "Preface to the Third Edition of the *OED*," Accessed July 12, 2020. <http://public.oed.com/the-oed-today/preface-to-the-third-edition-of-the-oed>.
- Simpson, J. (2013), "Mediums of access to the *Oxford English Dictionary*," In G. Stickel and T. Varadi (eds.), *Lexical Challenges in a Multilingual Europe: Contributions to the Annual Conference 2012 of EFNIL in Budapest*, 95–110. Frankfurt am Main: Peter Lang.
- Tan, C. L.-L. (2016), *Sarong Party Girls*, New York: William Morrow.
- Weiner, E. (1987), "The new *OED* and World English," *English Today*, 3 (3): 31–4. (Reprinted from *English World-Wide*, 7: 259–66).