
Opening up the Oxford English Dictionary: what an enhanced legacy dataset can tell us about language, lexicography, and literature

David-Antoine Williams

david.williams@utwaterloo.ca

St Jerome's University in the University of Waterloo
Canada

Overview

This paper will discuss recent research carried out in the context of an OMRI-funded project: “The Life of Words: Poetry and the OED.” It will discuss the processes and methods developed for enhancing and manipulating a large legacy dataset—the Second Edition of the Oxford English Dictionary—and will present analyses and applications pertaining to lexicography, lexicology, and traditional literary studies. Styled as an “opening up” of latent information in a previously closed system, and with implications in digital humanities, linguistics, lexicography, lexicology, literary criticism, and poetics, the paper will address the conference theme of “Access” in the sense of tapping into knowledge that has previously been inaccessible.

Context

The Oxford English Dictionary (OED) is widely considered to be the greatest philological and lexicographical achievement in English. As the first fascicles of the OED were being prepared, editor James Murray professed his dedication to “The perfection of the Dictionary in its data” (1880: 129, orig. emph.). The “data” of the work is its 2.43 million quotations, a significant portion of them from poetic and other literary texts, which both shape and illustrate the various sense definitions of roughly 600,000 English words and word forms. Conversely, since its publication, poets have relied on the OED to guide their deployments and arrangements of English words in poems. This reciprocal intertextuality has led to two striking facts which have yet to be fully explored: 1) that the OED’s definitions of English words depend to

a significant degree on poetic language, which is striking because by any standard account, poetic usage tends away from the denotative or definitional and towards the connotative and metaphorical; and 2) that much English poetry of the last hundred years contains a philological, etymological, and lexicographical dimension, informed by the OED.

Although the Second Edition of the OED (Murray et al., 1989) was among the first large reference works to be prepared for public and academic communities in digitized, marked-up form, and despite the current and ongoing revision of the dictionary (Simpson et. al., 2000-), no version has ever been marked-up with additional metadata. Information we might expect to find in a modern text dataset, such as author gender, genre of quotation, and type of publication, is not included in the OED. Attempts to incorporate such information into studies of linguistic, literary and cultural questions have until now have therefore been limited to “case-study” or “sampling” methodologies.

“The Life of Words: Poetry and the OED” addresses this by working directly with legacy versions of the electronic OED, enhancing these with appropriate metadata about the quotation evidence. With the enhanced dataset, alongside modern large text corpora, we then generate quantitative and qualitative assessments in two broad fields of inquiry: 1) What has been the influence of poetry on the English language’s most comprehensive lexicographical work? and 2) What influence has the OED had on English-language poetry? To take a modern turn on Murray, our concern now is the perfection of the dictionary in its data, metadata, and comparative data.

Paper Outline

The preamble to this paper will briefly introduce the project, discussing its background, methods employed, and current stage of development, and offering some observations regarding the use of dictionaries in general, the OED in particular, and specifically the “opened-up”, marked-up and directly accessible enhanced OED, as “evidence” for interpretation in a number of scholarly domains, a methodological topic which has received recent attention in dictionary studies, literary studies, and linguistics (e.g. Coleman 2013a, 2013b; Coleman and Ogilvie 2009; Hoffman 2004). The bulk of the presentation will be devoted to an exploration of the enhanced OED, demonstrating some top-level findings relevant to current scholarship in three fields. In the history of lexicography, recently there has been much discussion on the interpretation of OED quotation

evidence as a complex indicator of both the prestige of certain kinds of writing over time, and the particular judgements, biases, and practices of the nineteenth-century philologists who compiled the First Edition (1884-1928) (Ogilvie 2013; Brewer 2012, 2010, 2009, 2007; Considine 2009; Mugglestone 2005, 2000; Willinsky 1994). In the first main part of the paper, I give an overall quantitative assessment of the generic make-up of OED quotations, comparing the First and Second Editions, and discuss the implications of this for literary and for cultural history. Next, at the intersection of lexicology and literary studies, I offer a re-assessment of claims surrounding the linguistic inventiveness of canonical authors such as Shakespeare and Milton (Goodland 2011, 2010; Brewer 2013, Crystal 2000; Gray 1989; Schafer 1980) based on benchmarks for the period and genre of their various works. Finally, in the realm of literary criticism, I demonstrate a number of ways that information embedded in the OED can be harnessed to detect literary tropes such as allusion and etymological wordplay, either in poems that have been directly influenced by the OED, and those for which the evidence is less conclusive.

Bibliography

- Brewer, C.** (2013). "Shakespeare, word-coining, and the OED" in *Shakespeare Survey* 65: 345-57.
- Brewer, C.** (2012). "Happy Copiousness? OED's Recording of Female Authors of the Eighteenth Century" in *Review of English Studies* 63.258: 86-117.
- Brewer, C.** (2010). "The Use of Literary Quotations in the OED", *Review of English Studies* 61: 93-125.
- Brewer, C.** (2009). "The OED as 'literary instrument': its treatment past and present of the vocabulary of Virginia Woolf" *Notes & Queries* 56: 430-44.
- Brewer, C.** (2007). *Treasure-House of the Language: The Living OED*. New Haven: Yale University Press.
- Coleman, J.** (2013a). "Using Dictionary Evidence to Evaluate Authors' Lexis: John Bunyan and the Oxford English Dictionary" in *Dictionaries The Journal of the Dictionary Society of North America* 34: 66-100.
- Coleman, J.** (2013b). "Forum: Using OED Evidence" in *Dictionaries: The Journal of the Dictionary Society of North America* 34: 1-9.
- Coleman, J. and Ogilvie, S.** (2009). "Forensic Dictionary Analysis: Principles and Practice" in *International Journal of Lexicography* 22.1: 1-22.
- Considine, J.**(2009). "Literary classics in OED quotation evidence" in *Review of English Studies* 60.246: 620-638.
- Crystal, D.** (2000). "Investigating Nonceness: Lexical Innovation and Lexicographical Coverage" in R. Robert Boenig and K. Davis, eds, *Manuscript, Narrative and Lexicon: Essays on Literary and Cultural Transmission in Honor of Whitney F. Bolton*. Lewisburg: Bucknell University Press,; 218-31.
- Goodland, G.** (2011). "'Strange deliveries': Contextualizing Shakespeare's first citations in the OED" in Mireille Ravassat and Jonathan Culpeper, eds, *Stylistics and Shakespeare's Language: Trans disciplinary Approaches* (London: Continuum,); 8-33.
- Goodland, G.** (2010). "The OED and 'single-use' words". <http://ora.ouls.ox.ac.uk/objects/uuid%3A99d462ea-be60-4b60-b7a2-6259a862c500/datastreams/ATTACHMENT03>.
- Hoffman, S.** (2004). "Using the OED Quotations Database as a Corpus—a linguistic appraisal." *ICAME Journal*, 28: 17-30.
- Mugglestone, L.** (2005). *Lost for Words: The Hidden History of the Oxford English Dictionary*. New Haven: Yale University Press.
- Mugglestone, L.** (2000). *Lexicography and the OED: Pioneers in the Untrodden Forest*. Oxford: Oxford University Press.
- Murray, J. A. H.** et. al., eds. (1989). *Oxford English Dictionary*. 2nd edn, compiled by J. A. Simpson and E. S. C. Weiner, 20 vols. Oxford: Oxford University Press.
- Murray, J. A. H.** (1880). 'The President's Annual Address for 1880', in *Transactions of the Philological Society, 1880-1881*. London: Trubner.
- Ogilvie, S.**(2013). *Words of the World: A Global History of the Oxford English Dictionary*. Cambridge: Cambridge University Press.
- Schäfer, J.** (1980). *Documentation in the O.E.D.: Shakespeare and Nashe as Test Cases*. Oxford: Clarendon Press.
- Simpson, J., Weiner, E. S. C. and Proffitt, M.** (2000-). *OED Online*. 3rd edn, rev. J. A. Simpson et al.. Oxford: Oxford University Press.