Dataset: Cant Glossaries of Bampfylde-Moore Carew

**Introduction**

The cant (or slang) glossaries after the text of multiple editions of the “memoirs” of Bampfylde-Moore Carew change over time and with each edition. This dataset tracks the changes in the ways words were entered in the glossaries over time and highlights a few of the important characteristics of the data. The works are now in the public domain and so the dataset is made openly available for reuse, augmentation, and recreation by scholars and logophiles.

**Context of the Dataset**

The life of the “King of the Beggars,” Bampfylde-Moore Carew, has been immortalized in the publishing of his memoirs over about 50 editions, 30 of which were published in the 50 years immediately following his death (“Bampfylde-Moore Carew: the early years”)[[1]](#footnote-1). Each edition has its unique differences from preceding works, whether it be a complete change in storyline or simply a decapitalization of the nouns from earlier editions. In one case, the main variation between the 1749 and 1813 versions, the hero is almost always referred to as “Bampfylde” in the former and “Mr. Carew” in the latter. In one edition (1745), Carew’s removals to America are not discussed at all. Another version, a “pocket-book” edition, recounts only 60 pages’ worth of Carew’s stories in an episodic fashion.

Beginning in 1750, each edition includes a glossary after the text, usually entitled something like “A Dictionary of the Cant Language,” [[2]](#footnote-2)(Coleman 2004)*.* Surprisingly, the glossary and the memoir have no real relation to each other. The words included in the glossaries are rarely, if ever, incorporated into the main work[[3]](#footnote-3). It is not known with any surety why the “dictionary” is included if it does not serve as a reference for words used in the text. Perhaps it is to add a *bona fide* feel to the entire work; or it may be that the editor found the idea to add a list of words from a rakish and risqué society would appeal to readers and help sell copies.

Mr. Carew is often vaunted as being “King of the Gipseys,” and the heading of the glossary in many editions reads, “As the Language of the Community of the Gipsies is very expressive, and different from all others, we think we shall do a Pleasure to the Curious by annexing a short Specimen of it.” Despite these assertions (and the outdated and offensive terminology), the Romani influence upon the words included in the glossary is negligible if not nonexistent. The chosen inclusions seem to be more mendicant terms than specifically Romani ones.

The glossary is as malleable as the text. The words contained therein change meaning, spelling, punctuation, capitalization, and are transposed and spliced together. The entries are chaotic, often not completely in alphabetical order after the first letter. Each editor seems eager to both continue and to add their own flairs to Carew’s story, as well as to the cant dictionary included in the back. Whether from haste or a desire to “correct” or improve upon earlier editions, each version of the glossary has at least minor changes in spelling, printing, capitalization, or punctuation.

The editions of the memoir that were investigated for this dataset were, by year: 1765, 1768, 1779, 1782, 1785, 1788, 1793, 1798, 1813, 1827, 1835, ca. 1850, and ca. 1880[[4]](#footnote-4). The 1931 reprint, *The King of the Beggars* by C. H. Wilkinson, contains a version of *The Life and Adventures of Bampfylde-Moore Carew* and *Apology for the Life of Bampfylde-Moore Carew*, but no glossary for either. Words in the dataset were entered by hand and double- and triple-checked down to capitalization and punctuation marks, so errors included are printers’ errors and not due to OCR mischaracterization. Each column is an edition, delineated by year, and rows are used to track entry changes over time. Blank fields indicate that an entry does not appear in the glossary for that edition.

**Outside Resources**

The origin for entries included in early glossaries appears to be the *New Canting Dictionary* (Coleman 2004). However, if a cant glossary from Carew’s memoirs prior to 1813 is compared with a glossary published after 1813, there is quite a noticeable difference in the number and choice of words included. Entries such as *Clickman toad* and *Dimber damber* appear in the 1813 edition; and although the next two or three printings do not retain all of these entries, by around 1850 they have been re-included in the glossaries. Most of these “new” entries come from Grose’s *Dictionary of the Vulgar Tongue*, 1811. The editor of the 1813 version seems to have become tired with his task of deciding which words of Grose’s to insert into the original and most frequently published gloss, because most of the entries pulled from Grose are in the first part of the alphabet. After the entries beginning with the letter D, there is a marked decrease in the addition to the glossary of new words from the *Dictionary of the Vulgar Tongue*. The letter A contains 11 new entries; B, 30; C, 27; D, 13; E, 2; F, 5; G, 6; H, none; I, none; J, 1; K, none; L, 2; M, none; N, none; and so on for the rest of the list.

The 1827 edition is an anomaly, and has been included as a separate sheet in the workbook. There is a marked difference in the 1827 edition of the glossary from other glossaries examined. Nearly 1,230 words are included in the 1827 glossary, and they do not appear in other editions. The other glossaries contain 228-420 words, with one outlier containing a mere 60. Many of these entries are derivations, such as: *Rum, rum degen, rum duke, rum gutters, rum quids, rum chant, rum boose, rum buffer,* etc. with “rum” in this case defined as “good” or “satisfactory.” It is unclear why there is such a radical departure from any preceding glossaries and why no subsequent versions follow the 1827 glossary. Most of the anomalous words find their sources in Grose, or are amalgamations of two or more words into phrases, each individual word of which can be found in Grose (see list above).

Four words are listed as quotations in the Oxford English Dictionary from Bampfylde-Moore Carew’s memoirs: *flick, flicker*, *porker*, and *hand-me-downs*. The three former are quoted from 1750; the latter from 1845. It is unknown why these four entries are the only ones used as quotations in the Oxford English Dictionary; certainly, other words from the glossaries are listed in the OED, and words from the text only are listed as quotations, but there are no other entries from the glossaries that can be traced through the OED. There is one exception: the anomalous 1827 glossary includes the word *pot-hook*, which is listed as a quotation in the OED from 1760. It is possible that the other entries are borrowed from earlier sources, such as the *New Canting Dictionary* (the OED lists as quotations many words glossed in Bampfylde-Moore Carew, such as *barnacled, buck, chiv, cursitor, harman, lag of duds,* etc.) and Grose’s *Dictionary of the Vulgar Tongue* (in which is found *able whackets, autem jet, degen, facer, George,* etc.). It may also be that the editors of the Oxford English Dictionary do not consider Carew’s memoirs a reliable source, because (as stated above) there are so many variations and errors in the glossaries appended to Carew’s text.

Several words vary drastically over the publication history of *Bampfylde-Moore Carew*. For instance, the word *huskylour/husbylour/hasbylour*  is spelled all three of those ways, but retains its original meaning: “a job, or a guinea.” The differences in spelling seem to be due to printer error and possibly difficulty reading the typeface or printing. Another entry, *lad a duds*, changes spelling and punctuation over time. Some of the forms include: *lad a dudds, lad-a-dudds, lad a duds, lad a dads, lage duds,* and *lage o’ duds.* These all have the same definition as well: “a buck of clothes.” The discrepancies among the forms of this word seem to be due to the same factors as the preceding word discussed, but also possibly to an oral shift in the way in the word was pronounced.

The endless variations on the main work of Carew’s memoirs can offer valuable insight into the printing process and how the text and its glossary evolved over time. Many of the changes can be attributed to printer error, but others are conscious choices on the parts of the editors to include new material or alter the existing work. Although the glossary does not relate directly to the text, it is an important part of the books, as well as an intriguing tool to help understand and observe the transformations of the words and entries from the memoir’s first edition.

**Audiences and Re-Use Value**

The reuse value of this project is niche but has larger reverberations for those working in studies of English slang and lexicology. This is one of the first places that slang words were collected and defined, as up to roughly this point, dictionaries were notoriously difficult to keep up-to-date, even regarding words that were not frowned upon by the academics compiling those lexicons. Additionally, it offers a micro-study in word use and morphology over time, similar to Google’s n-gram viewer or HathiTrust’s bookworm, but on a more granular scale (taking into account different spellings and punctuation). It could offer a model to capture entries over time and perform comparative analysis of a lexicon like, for instance, Urban Dictionary[[5]](#footnote-5).On GitHub alone, there are over 50 repositories that match the tags “slang,” “glossaries,” and “cant”[[6]](#footnote-6). This dataset can be read with any of those a user may find useful, as well as with larger datasets like the Oxford English Dictionary, to explore word changes over time and how lexicons develop.

**Technical Considerations**

Currently, the data is about 70 KB, in the form of an Excel spreadsheet, in an Excel file with multiple worksheets in the book. It is a closed dataset (for now) and has some gaps in the data, whether for missing words that don’t show up until later editions, or entire editions that were not accessible to the compiler at the time of compilation. The file size should not increase by much, so at most, each version of the dataset should be 100 KB. Because the original file contains multiple sheets in the workbook, it is uploaded as one .xlsx file, as well as four separate .csv files for ease of open use. Digital preservation will be done by keeping multiple copies (both on the compiler’s personal computer and online, hosted on and managed by GitHub) checked by checksums and monitoring and transparency of changes made[[7]](#footnote-7).

Storing the data on GitHub will also serve to minimize costs—the possibility of future costs if digitized images are added to the data would be voluntary. Metadata according to Dublin Core have been entered in the file’s accompanying ReadMe file, with LC subject headings and authorities followed. Because the editions consulted are all in the public domain as of this point, the data has been licensed as CC-BY-NC; while the data and its origins are open, this dataset could be used by other scholars and submitted to non-open access journals if necessary, so a CC-BY-SA license was considered too restrictive. Many people will not have access to multiple copies of 18th- and 19th-century texts from which to do research, so the greatest amount of openness in licensing without too much restriction on reuse was balanced in making licensing decisions.

**Future Work**

There are many different directions in which this dataset can be augmented or reused, and the compiler hopes that others will take full advantage of those avenues. The former include: adding to the workbook editions which were not available at the time of compilation; adding definitions to entries in the workbook, which were not the main focus of the original project; continued research on entry etymologies, different dictionaries that influenced the glossaries, and printer patterns and errors; and adding digitized copies of glossary pages to the dataset. Visualization may not be as useful, but perhaps entries could be added to a timeline. It is hoped that this dataset will not only be reusable, but added to by those users to continue to create meaningful, rich data.

**References**

*Angloromani Dictionary*. Romani Project, 2006. <http://romani.humanities.manchester.ac.uk/angloromani/dictionary.html>

"Bampfylde Moore Carew: the early years." *Devon Perspectives*. <http://www.devonperspectives.co.uk/bampfyldemoorecarew.html>

Coleman, Julie. *A History of Cant and Slang Dictionaries, Volume 1: 1567-1785*. New York: Oxford University Press, 2004. 127-41.

Grose, Francis. *1811 Dictionary of the Vulgar Tongue*. London: Printed for R. Hooper, 1811.

*The Oxford English Dictionary*. Oxford University Press. <http://www.oed.com>

1. The “memoirs” are loosely based on his life, and adapted for whatever adventures subsequent contributors felt would titillate readers. This could be considered an early iteration of fan fiction. [↑](#footnote-ref-1)
2. Cant can be understood as jargon or slang of an in-group. Since cant is the wording the primary text uses, it will be used here as well. [↑](#footnote-ref-2)
3. “Word” and “entry” are used interchangeably, depending on what makes the most sense in context. Some entries are multiple words or phrases. [↑](#footnote-ref-3)
4. Ca. dates are approximate as specific printing dates are not given. [↑](#footnote-ref-4)
5. This might be a mistake. [↑](#footnote-ref-5)
6. These include the 1912 Webster’s slang dataset, with which there is surprisingly little overlap! [↑](#footnote-ref-6)
7. The GitHub repository for this data can be found here: [https://GitHub.com/lmjlynn/BMCGlossaries](https://github.com/lmjlynn/BMCGlossaries). [↑](#footnote-ref-7)