



This is a repository copy of *'Proper' pro-nun-fha- fhun in eighteenth-century English: ECEP as a new tool for the study of historical phonology and dialectology.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/110307/>

Version: Accepted Version

---

**Article:**

Yanez-Bouza, N., Beal, J., Sen, R. [orcid.org/0000-0002-1558-7938](https://orcid.org/0000-0002-1558-7938) et al. (1 more author) (2018) 'Proper' pro-nun-fha- fhun in eighteenth-century English: ECEP as a new tool for the study of historical phonology and dialectology. *Digital Scholarship in the Humanities*, 33 (1). pp. 203-227. ISSN 2055-7671

<https://doi.org/10.1093/lc/fqx003>

---

This is a pre-copyedited, author-produced version of an article accepted for publication in *Digital Scholarship in the Humanities* following peer review. The version of record Yáñez-Bouza, N. et al; 'Proper' pro-nun-fha-fhun in Eighteenth-Century English: ECEP as a New Tool for the Study of Historical Phonology and Dialectology, *Digital Scholarship in the Humanities*, fqx003 is available online at: <https://doi.org/10.1093/lc/fqx003>

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# 'Proper' *pro-nun-fha-fhun*<sup>1</sup> in Eighteenth-century English: ECEP as a New Tool for the Study of Historical Phonology and Dialectology

## 1. Introduction

In recent years English historical linguists have voiced complaints about the scholarly neglect of the Late Modern English period (1700–1900). While grammar and the prescriptive grammatical tradition have received increasing attention over the last couple of decades (Beal *et al.*, 2008; Tieken-Boon van Ostade, 2008), there is still relatively little research on the phonology of Late Modern English, and of the eighteenth century in particular; as Beal (1999: 13) points out, '[w]here interest is shown in the eighteenth century, phonology is neglected, and where interest is shown in the history of English phonology, the eighteenth century is neglected'. There remains an urgent need for new studies of historical phonology in general and of eighteenth-century phonology in particular. One reason for this lack of research could be the idiosyncratic notation systems used by eighteenth-century authors, which make it difficult to search and interpret phonological evidence. Yet the value of pronouncing dictionaries as rich and reliable evidence of lexical diffusion as well as of sound variation and change in eighteenth-century pronunciation has been observed in studies such as Beal (1999) and Jones (2006).

With this in mind, we have constructed a new electronic, searchable database of *Eighteenth-Century English Phonology* (ECEP). The purpose of this paper is to present ECEP as a tool to facilitate research on the social, regional and lexical distribution of phonological variants in eighteenth-century English, thereby meeting the demands of the growing research community in Late Modern English generally (Mugglestone, 2003; Hickey, 2010) and in historical phonology in particular (Honeybone and Salmons, 2015). Taking Wells' (1982: 127–68) lexical sets for comparing the vowel systems of present-day varieties of English as its reference, the database provides unicode IPA transcriptions for the relevant segment of each word given as an example of its lexical set or subset in Wells' account of standard lexical sets, as documented in eighteenth-century pronouncing dictionaries (e.g. Thomas Sheridan's *A General Dictionary of the English Language*, 1780). Wells' subsets provide important points of comparison where varieties of English differ as to the distribution of variants within the lexical set. For example, whilst the lexical set BATH is defined as 'comprising those words whose citation form contains the stressed vowel /æ/ in GenAm, but /ɑ:/ in RP' (Wells 1982: 133), subset (59b) consists of words that are sometimes pronounced with /æ/ 'in accents which otherwise have broad BATH (Australian, West Indian)' (1982: 134–5), whilst those in subset (59c) 'typically have the PALM vowel in the otherwise flat-BATH accents of the north of England' (1982: 135). We have retained the subsets in the structure of ECEP in order to determine whether these major distinctions between accent systems in present-day English have precursors in the variation present across our eighteenth-century data sources. We have included all of the examples provided by Wells which occur in these sources in order to reveal any further patterns of lexical distribution. Since Wells' lexical sets are designed only to investigate vowel systems, and some consonantal changes are also of interest in the study of Late Modern English phonology, we have constructed consonantal sets of phonological interest and extracted relevant data in the same way as for Wells' vowel sets.

---

<sup>1</sup> Sheridan's (1780) transcription (diacritics omitted) for the word 'pronunciation', showing palatalized /j/ instead of /si/ at the start of the third and fourth syllables.

This paper describes the structure and contents of the ECEP database, and reports on the three-step method of compilation: (i) the selection of primary sources (Section 2); (ii) the process of data input and annotation (Section 3); and (iii) the design of the web-based interface (Section 4). The context for the development of this new tool is provided in Section 2, which gives an overview of the phonology of eighteenth-century English and of the value of pronouncing dictionaries as evidence for eighteenth-century pronunciation. Section 5 describes two pilot case studies which demonstrate the value of ECEP for the study of English historical phonology.

## 2. Background

### 2.1 *The Phonology of Eighteenth-century English*

Since Charles Jones described the eighteenth and nineteenth centuries as the ‘Cinderellas of English historical linguistic study’ (1989: 279), there has been considerable progress in Late Modern English<sup>2</sup> studies. Much of this progress has been made possible by the availability of corpora such as ARCHER (*A Representative Corpus of Historical English Registers*), which have enabled searches across large datasets for the complex patterns of variation and change which characterize this period. However, despite the monographs by Beal (1999) and Jones (2006), research on the phonology of this period has been less prolific than that in other areas such as morphosyntax, pragmatics and language ideology. One reason for this relative neglect of eighteenth-century phonology is the lack of accessible primary source material: as argued by Beal (2012b), the corpus revolution which has energised other areas of Late Modern English studies has so far had little effect on phonology. The ECEP project aims to redress this.

Some scholars have actually suggested that the phonology of eighteenth- and nineteenth-century English is not worthy of their attention. Bloomfield and Newmark (1963: 288), for example, state that changes in the language between the eighteenth century and the present day are ‘due to matters of style and rhetoric [...] rather than to differences in phonology, grammar or vocabulary’. Bloomfield and Newmark go on to state that historical linguists are less interested in style and rhetoric, a statement which no longer rings true given the recent development of historical pragmatics and historical sociolinguistics. Strang notes that ‘some short histories of English give the impression that change in pronunciation stopped dead in the eighteenth c[entury], a development which would be quite inexplicable for a language in everyday use’ (1970: 78). MacMahon (1998), after summarising the views of earlier scholars who claimed that there had been little change in this period, states that ‘there is other evidence to show that the pronunciation of English more than 150 years ago was *noticeably* different, for reasons mainly of phonotactics (structure and lexical incidence) from what it is today’ (1998: 374, original italics). Whilst both Strang and MacMahon assert that changes in pronunciation have taken place since 1700, both make the point that these more recent changes are less systemic than those occurring in earlier periods. Beal argues that this opposition is to some extent ‘an illusion created by the different types of evidence available for the earlier and later periods’ and goes on to invoke the saying ‘can’t see the wood for the trees’: in her opinion, ‘[i]t is a matter of perspective: at a distance, a forest appears as a monolithic block, but, the closer you get to the forest, the more you notice the variation between individual trees’ (Beal, 2004: 125). Not only are we closer in time to the eighteenth century than to the Middle or Early Modern English periods, but the amount of detailed

---

<sup>2</sup> Late Modern English is generally agreed to cover what historians would term the ‘long’ eighteenth and nineteenth centuries. See Beal (2004; 2012a), Tiekens-Boon van Ostade (2009) for more detailed definitions.

information on the pronunciation of more recent English makes us more aware of the range of variation. Moreover, some of the changes occurring in this period are still ongoing and/or are reflected in variation between varieties of English today. Examples of these changes are the distribution of ‘long’ /ɑ:/, /ɔ:/ and ‘short’ /a/, /ɒ/ variants in the BATH and CLOTH lexical sets respectively (see Beal and Condorelli, 2014 for an account of the latter); the ‘North-South divide’ in the absence or presence of the phoneme /ʌ/ (Beal, 2012c); and the ongoing palatalization of alveolar consonants preceding earlier /ju:/ (see Section 5 below).

Since many of the phonological changes taking place in the eighteenth century involve shifts in lexical incidence, sources of evidence used to investigate these changes need to be lexically rich. As we shall see in the next section, the sources chosen for inclusion in ECEP are ideal for these purposes, as they provide evidence for the entire lexicon.

## 2.2 Phonology Sources in ECEP

Written evidence for the historical pronunciation of English can be divided into direct and indirect types. Evidence that is indirect involves sources whose authors were not overtly commenting on or describing pronunciation, but which give clues about it. Typical sources of indirect evidence are rhymes, puns and non-standard spellings. Direct evidence, on the other hand, comes from authors who deliberately set out to describe (or prescribe) the pronunciation of their time. In reconstructing the pronunciation of earlier periods of English, we have to rely mainly on indirect evidence, but from the sixteenth century onwards, direct evidence becomes increasingly available as interest in spelling reform and in phonetics increases. Texts such as Christopher Cooper’s (1687) *The English Teacher* provide detailed and sophisticated descriptions of the sounds of English, lists of homophones and near-homophones and even metalinguistic comments on the social and/or geographic distribution of variants, but exemplify their descriptions with a very restricted number of lexical tokens. However, from the middle of the eighteenth century dictionaries are published in which the pronunciation of every word is described, and these provide the source material for ECEP.

To illustrate the quantitative difference between orthoepistic works such as Cooper’s (1687) and eighteenth-century pronouncing dictionaries, let us consider the distribution of long and short variants of ME *ǣ* in the BATH and START sets. Cooper provides important early evidence for the lengthening of the vowel in these sets, which allows us to identify some of the phonetic environments in which the change first occurs. First he tells us of ‘the vowel a lingual’ that ‘in these *can, pass by, a* is short; in *cast, past, for passed, it is long*’ (1687: 4). Then he goes on to discuss the contexts in which the vowel ‘is pronounced long in its own sound’ (that is, /ɑ:/ rather than /e:/), these being ‘before *nch* and *s* when another Consonant follows, and before *r* unless *sh* follows’ (1687: 34). Cooper provides a list of words in which this lengthened *a* occurs: *barge, blast, carking, carp, cast, dart, flasket, gasp, grant, lance, mask, path, tart*. These words have been chosen to provide the same pre-vocalic environments as words which exemplify ‘a short’ (/a/) and ‘a slender’ (/e:/): thus *bar* with short /a/ is contrasted with *barge* pronounced with /ɑ:/ and *bare* with /e:/. From Cooper’s evidence we can piece together an account of the environments in which early lengthening occurs, but we have no way of knowing whether the examples chosen represent all the words in which orthographic <a> occurs in the given environments. For instance, Cooper provides *path* as an example of a word with /ɑ:/, but does not specify whether the same vowel would be used in other words with this post-vocalic environment, such as *bath, lath*, etc. By contrast, ECEP contains 127 words from the BATH set and 28 words from the START set. This will allow users to trace variation between /a/ and /ɑ:/ across a much larger subset of the lexicon and to identify differences in the transcriptions of authors from different places writing at different times within the eighteenth century (see Beal, 1999: 105–18 for further discussion of this sound change).

The qualitative value of evidence from eighteenth-century pronouncing dictionaries has been disputed in the past. Dobson, albeit writing at a time when many of the sources used in the *Eighteenth Century Collections Online*<sup>3</sup> were unknown or inaccessible, stated that ‘the eighteenth century produced no writers to compare with the spelling reformers who are our main source up to 1644 (Hodges) or with the phoneticians who, beginning with Robinson (1617) carry us on from 1653 (Wallis) to 1687 (Cooper’s *English Teacher*)’ (Dobson, 1957: 311). Others have taken issue with the prescriptivism of eighteenth-century authors. John Walker, the most successful and influential of these, is often singled out for criticism on this account. Sheldon (1947: 146) writes that ‘Walker satisfies the temper of his time [...] and its demand for linguistic regulation and reform’, whilst Holmberg (1964: 41) accuses Walker of being ‘influenced by the spelling’. It is true that all the pronouncing dictionaries used for ECEP were written with the aim of providing their readers with a guide to what the authors considered ‘correct’ pronunciation, but the same could be said of the many twentieth- and twenty-first century dictionaries which transcribe the pronunciation of English words in RP and/or General American. Recent scholars such as Agha (2003), Beal (2003), Ranson (2012) and Trapateau (2016) have rehabilitated Walker’s reputation as a phonetician by taking his work on its own terms as an important and highly informative source of information on the prestigious metropolitan pronunciation which was the precursor of RP. Walker’s (1791) *Critical Pronouncing Dictionary* is the major source of metalinguistic comments in ECEP, many of which provide valuable sociolinguistic information (see Section 3.2 below for further discussion of metalinguistic comments). Other sources used in the compilation of ECEP provide accounts of what was considered ‘correct’ pronunciation in the provinces.

The sources included in ECEP include the earliest available editions of all the accessible pronouncing dictionaries of English printed in eighteenth-century Britain.<sup>4</sup> At the time of writing, these are as follows:

- Buchanan (1757) *Linguae Britannicae Vera Pronuntiatio*.
- Johnston (1764) *A Pronouncing and Spelling Dictionary*.
- Kenrick (1773) *A New Dictionary of the English Language*.
- Perry (1775) *The Royal Standard English Dictionary*.
- Spence (1775) *The Grand Repository of the English Language*.
- Sheridan (1780) *A General Dictionary of the English Language*.
- Burn (1786) *A Pronouncing Dictionary of the English Language*.
- Scott (1786) *A New Spelling, Pronouncing and Explanatory Dictionary of the English Language*.
- Walker (1791) *A Critical Pronouncing Dictionary and Expositor of the English Language*.
- Jones (1797, 1798) *Sheridan Improved. A General Pronouncing and Explanatory Dictionary of the English Language*. 2nd and 3rd editions.

Buchanan (1757) is the first true pronouncing dictionary of English, in the sense that every word is transcribed. It was decided to include two editions of Jones’s dictionary because the third edition demonstrates significant changes in which Jones distances himself from Sheridan, most noticeably in recognising a distinction between long and short vowels in the BATH and START sets. In future, we intend to augment ECEP with data from later editions and from other sources, but those listed above provide evidence across the second half of the eighteenth century from authors of varying geographical provenance – one Irishman (Sheridan), four Scotsmen (Buchanan, Perry, Burn, Scott), one northern author from Newcastle (Spence), three authors from the London area (Kenrick, Jones, Walker), and one author of uncertain origin but

<sup>3</sup> <http://gale.cengage.co.uk/product-highlights/history/eighteenth-century-collections-online.aspx>

<sup>4</sup> We intend to include early American pronouncing dictionaries in later versions of ECEP.

who lived and worked in the south-east county of Kent (Johnston). It is important to state that ECEP is not intended to be a database of dialectal pronunciation, but it does reflect the variation between the ‘received’ speech of London and of the equivalent in provincial centres such as Edinburgh and Newcastle, as well as providing evidence for change over the course of the eighteenth century.

### 3. Data Annotation

Once the pronouncing dictionaries had been selected, the next step in the compilation of ECEP was the process of data input and annotation. This section reports on the design and contents of the database, including the methodological principles adopted.

#### 3.1 Database Design

ECEP has been built in MS Access format as a relational database constructed with a variety of integrated tables. The data have been systematically annotated and thematically grouped in three major categories: phonology data, source metadata and author metadata. Details for each category are set out in Table 1.

**Table 1 Design of the ECEP database**

(Meta)Data	Fields
Phonology	Lexical set, Lexical subset, Keyword, IPA, IPA variants, Example word frequency, Metalinguistic comments, Metalinguistic attitude, Metalinguistic label, Compilers’ notes
Source	Type of work, Title, Year of publication (of the edition consulted), Edition, Place of publication, Imprint (printers, booksellers), Price, Physical description, Paratext, Audience (age, gender, social class, instruction, specific purpose), References consulted, Compilers’ notes
Author	Name, Life dates, Gender, Social class, Place of birth, Places of residence, Occupation, Other biographical details, Works by this author in ECEP

The metadata for the dictionaries have been drawn from the original sources, such as the title-pages and prefaces to works, and also from the literature (e.g. Alston, 1966; Beal, 1999). The metadata for the authors come principally from the *Oxford Dictionary of National Biography*.

Regarding the phonology data, the starting point for drawing up the list of words for ECEP was John Wells’ (1982) *Accents of English*, in particular his list of Standard Lexical Sets for the vowel system in varieties of present-day English (1982: 119–20, 127–68). Our aim was for ECEP to incorporate data from the selected pronouncing dictionaries in the form of IPA transcriptions so that the historical data documented in the database could be easily compared to present-day studies; this was necessary because, as mentioned above, the notation systems used by eighteenth-century authors were often idiosyncratic and difficult to interpret (see Section 3.2 and Appendix III). The use of Wells’ lexical sets and their associated example words is standard practice in studies of variation and change in present-day English. Including the full range of example words allows for differences in lexical distribution between the primary sources, and also between these and the contemporary accents described by Wells. For instance, a scholar interested in the distribution of words related to the STRUT-FOOT split would be able to find how each of the words provided as examples for Wells’ sets is

transcribed in each of the eighteenth-century sources documented in the database, and how phonological variants are perceived at the time in the context of the standardization of English (e.g. correct, vulgar, improper, etc.).

Wells (1982: 119–20) explains that ‘[t]he use of one vowel or another in particular words (lexical items) can be illustrated by tabulating their occurrence’ in the set of keywords presented in Table 2 in small caps, so that each of them ‘stands for a large number of words which behave the same way in respect of the incidence of vowels in different accents’; the latter are referred to in this paper as *example words*. Overall his list contains twenty-four lexical sets for stressed vowels and three sets for unstressed vowels; this makes 1,737 example words in total, distributed in sixty-one subsets. The sets KIT, DRESS, TRAP, LOT, STRUT, FOOT, CLOTH, concern short vowels; the sets BATH, NURSE, FLEECE, PALM, THOUGHT, GOOSE, START, NORTH, FORCE refer to long vowels;<sup>5</sup> the sets FACE, GOAT, PRICE, CHOICE, MOUTH, NEAR, SQUARE, CURE include diphthongs; and the sets *happy*, *letter*, *comma* represent unstressed vowels.<sup>6</sup>

**Table 2 Wells’ (1982: 127–68) lexical sets in ECEP (sorted as in Wells)**

SET	SUBSET	EXAMPLE WORD
<b>Short Vowels</b>		
KIT	--	bit, drink
DRESS	--	bed, deaf
TRAP	--	back, thank
LOT	--	box, sock
STRUT	--	blood, cut
FOOT	--	bush, full
<b>Long Vowels and Diphthongs</b>		
BATH	BATH_a	ask, castle
	BATH_b	branch, enhance
	BATH_c	banana, calf
	BATH_f	blasphemy, plastic
CLOTH	CLOTH_a	broth, cough
	CLOTH_b	coffee_1, moth
	CLOTH_c	coroner, florin
NURSE	--	birth, nerve
FLEECE	FLEECE_a	agree, cheese
	FLEECE_b	bead, deceive
	FLEECE_c	machine, police
FACE	FACE_a	age, safe
	FACE_b	day, faith
	FACE_c	break, great
PALM	PALM_a	calm, father
	PALM_b	bravado , innamorato
	PALM_f	almond, sultana
THOUGHT	THOUGHT_a	fall, sought
	THOUGHT_b	false, fault

<sup>5</sup> The sets are categorized as long- or short-vowel sets according to their pronunciation in RP.

<sup>6</sup> Practical notes. (a) The codes \_a, \_b etc. in Wells’ lexical subsets are preserved as in his original list, except for \_f, which Wells codes with an apostrophe and often refers to as an ‘appendix’ to the original set. (b) The codes \_1 and \_2 in some of Wells’ example words are used when the same word appears in more than one subset; the number indicates the syllable that is relevant in each particular case, as in *coffee\_1* for CLOTH\_b and *coffee\_2* for *happy\_b*.

GOAT	GOAT_a	boat, holy
	GOAT_b	grow, know
GOOSE	GOOSE_a	choose, shoot
	GOOSE_b	blue, few
PRICE	PRICE_a	arrive, try
	PRICE_b	fight, high
CHOICE	CHOICE_a	boy, noise
	CHOICE_b	join, spoil
	CHOICE_c	groin, hoist
MOUTH	--	down, mountain
NEAR	NEAR_a	beer, near
	NEAR_b	beard, fierce
	NEAR_c	hero, period
	NEAR_f	idea, real
SQUARE	SQUARE_a	air, pear
	SQUARE_b	scarce
	SQUARE_c	dairy, rarity
START	START_a	far, start
	START_b	bark, party
	START_c	tiara
NORTH	NORTH_a	for, war
	NORTH_b	assort, mortal
	NORTH_c	aura, Taurus
FORCE	FORCE_a	adore, door
	FORCE_bi	deport, forth
	FORCE_bii	coarse, fourth
	FORCE_c	aurora, glorious
CURE	CURE_ai	amour, tour
	CURE_aii	endure_vw, pure
	CURE_b	gourd, tournament
	CURE_ci	boorish
	CURE_cii	bureau, curious
<b>Weak Vowels</b>		
<i>happy</i>	<i>happy_a</i>	baby, city
	<i>happy_b</i>	coffee_2, vanity
<i>letter</i>	--	better, razor
<i>comma</i>	--	opera, saliva

To these sets for the study of the vowel system in general we have added five supplementary sets for the study of the consonant system in eighteenth-century English, including ten subsets and a total of 204 example words.<sup>7</sup> The sets DEUCE, FEATURE and SURE address the process of palatalization, dealing with stress patterns (subsets \_a for stressed syllable, \_b for post-stress syllable, \_c for pre-stress syllable), and the pre-/j/ phoneme (/t, d, s, z/ in each set). The set HEIR relates to the presence or absence of initial /h/, and the set WHALE to the pronunciation of 'wh'. See Table 3 and Appendix I for details.<sup>8</sup>

<sup>7</sup> More consonant sets may be added in due course.

<sup>8</sup> Practical notes. (a) When the same example word appears in a vowel set and in a consonant set, the former is coded with \_vw and the latter with \_cn, for instance *heir\_vw* for SQUARE\_a and *heir\_cn* for HEIR.



Table 3 Consonant lexical sets in ECEP

SET	SUBSET	EXAMPLE WORD
DEUCE	DEUCE_a	/t/ Tuesday
		/d/ due
		/s/ suit
		/z/ resume
	DEUCE_b	/t/ altitude
		/d/ module
		/s/ issue
		/z/ visual
	DEUCE_c	/t/ tumultuous
		/d/ adulation
		/s/ superior
		/z/ --
FEATURE <sup>9</sup>	--	/t/ creature
		/d/ procedure
		/s/ pressure_cn
		/z/ pleasure
SURE	SURE_a	/t/ mature
		/d/ during_cn
		/s/ surety
		/z/ c(a)esura_cn
	SURE_b	/t/ century
		/d/ verdure
		/s/ censure
		/z/ closure
	SURE_c	/t/ maturation
		/d/ duration
		/s/ mensuration
		/z/ --
HEIR	--	honour, humble
WHALE	WHALE_a	when, whine
	WHALE_b	elsewhere, somewhat

Each of the eighteenth-century pronouncing dictionaries in ECEP was examined in order to find Wells' example words for vowels and consonants, and the data were entered according to the following principles:

(b) If the same example word appears in more than one of the consonant sets, *\_deu* stands for DEUCE, *\_ture* for FEATURE, and *\_sure* for SURE, for instance *fissure\_ture* and *fissure\_sure*.

<sup>9</sup> This set consists of words which have schwa in the post-stress syllable in present-day English according to the *Oxford English Dictionary* (as opposed to a full vowel in SURE\_b), following a palatalized consonant in at least one pronunciation variant. These forms presumably arose from pronunciations with /jə/, which appear to have become more widespread in the eighteenth century. These in turn originated in forms with variation between [y:] and [iu] in the final syllable in Middle English. When the final syllable became unstressed, there was variation between 'full' forms with /iu/ and reduced forms with /ə/. The variants with /iu/ could then develop to /ju/ with the subsequent possibility of palatalizing the preceding consonant, whereas those with /ə/ did not lead to palatalization. Subsequent restoration of /j/ in the schwa-forms (with possible palatalization) combined with the reduction of /u/ to schwa in the full forms results in the remarkable variation we see in ECEP between e.g. /tju:/, tʃu:/, /tʃju:/, /tjə/, /tʃə/, /tjə/, and /tə/ in this set.

- a) Wells' lexical sets are designed for the analysis of present-day English. Naturally, the sets include example words that were introduced into the English language in recent times. Given that the scope of ECEP is limited to the phonology of the eighteenth century, we have excluded from the database those lexical items created or borrowed after 1800 (source: *Oxford English Dictionary*, January 2015).
- b) Wells' example words that are not documented in any of the pronouncing dictionaries examined have been excluded.
- c) Proper names and cliticised spellings of the type *don't*, *can't* have been excluded on the grounds that they are unlikely to be considered headwords in dictionaries. Country names appear occasionally in lists, as in Johnston (1764), but some did not exist at the time.<sup>10</sup>
- d) Example words that are documented in at least one pronouncing dictionary are included in the database, and the dictionaries in which an example word does not appear are coded 'NID' (i.e. 'Not In this Dictionary'). For instance, *macaroni* (set *happy\_a*) is missing in all but Perry (1775) and Scott (1786), and *whorl* (set *NURSE*) appears only in Johnston (1764).
- e) If an example word is listed in the dictionary but no pronunciation is provided, it is coded as NoP (i.e. 'No Pronunciation'), such as *cup* (set *FOOT*) in Kenrick (1773).
- f) At times pronouncing dictionaries do not list the precise example word, but they do list or make reference to a related word. In such cases we take note of the latter and add an explanatory note for users. For instance, for *awn* (set *THOUGHT\_a*) we have taken *awning* as the reference in five of the six dictionaries in which it is documented; and for *honourable* and *honesty* (set *HEIR*) we have taken *honour* and *honest* as reference example words in Kenrick's (1773) dictionary.
- g) Example words for which the notation system in the original source is unclear or ambiguous have been coded as Unclear.

Following the above method, ECEP currently lists 1,599 example words for each pronouncing dictionary: 1,395 example words in the vowel sets in 61 subsets, and 204 example words in consonant sets across 10 subsets. This leads to a total of 17,589 items annotated for the study of eighteenth-century English phonology. A summary of the contents of ECEP is set out in Table 4. Appendix II lists Wells' example words that have been excluded from ECEP according to principles a)-c).

**Table 4 ECEP contents**

	Lexical Sets	Subsets	Example words
<b>Vowels – Wells (1982)</b> KIT, DRESS, TRAP, LOT, STRUT, FOOT; BATH, CLOTH, NURSE, FLEECE, PALM, THOUGHT, GOOSE, START, NORTH, FORCE; FACE, GOAT, PRICE, CHOICE, MOUTH, NEAR, SQUARE, CURE; <i>happy</i> , <i>letter</i> , <i>comma</i>	27	61	1,395
<b>Consonants – Supplementary list</b> DEUCE, FEATURE, SURE; HEIR; WHALE	5	10	204
Total in each pronouncing dictionary	32	71	1,599
Total in all pronouncing dictionaries			17,589

<sup>10</sup> The exception to country names is *England*. Note that *Alexander*, *Charles*, *George* and *Morris* are included in ECEP because the dictionary entries refer to derivations which are no longer proper names as such; for instance, *Alexander* refers to the name of the herb.

### 3.2 Database Annotation

The database is designed to address research questions concerning the chronological, social, geographical and phonological distribution of variants such as /hw/~w/~h/ in the WHALE set, BATH broadening or the STRUT-FOOT split, all of which are of interest to sociolinguists, dialectologists and historical phonologists. To this purpose ECEP has been compiled to reflect the inventory of categorically distinct sounds in the way that the eighteenth-century pronouncing dictionaries document them; we avoid second-guessing issues of phonology here. As Beal (1999) has rightly argued with respect to notations for orthographic <a>:

the systems of notation provided in these pronouncing dictionaries tell us about the phonemic inventory of the recommended accent – that is, how many phonemes there are (we can, for instance, easily tell that Sheridan has three sounds—whilst Spence and Walker have four) whilst we can find out about the incidence of those phonemes from the dictionary entries themselves. What we cannot tell from a dictionary such as *The Grand Repository* is the phonetic nature of those phonemes: how do we know that the sound in *father* was [ɑ:] rather than [æ:] or even [ɛ:]? (Beal, 1999: 52)

Bringing together the information from all the pronouncing dictionaries, as we aim to do in ECEP, will help us address Beal’s question. Our method has thus been to translate the idiosyncratic notation systems of the dictionaries into unicode IPA transcriptions, based on the descriptions provided by the authors in the preface or introduction to their works. According to Bert Emsley’s categories of pronouncing dictionaries, eighteenth-century sources are ‘typically’ *diacritic*, so that diacritic marks indicate quality as well as quantity of sounds (cited in Beal, 1999: 80). They all tended to use different types of diacritic marks, though, and Spence’s *Grand Repository* in fact ‘stands apart from all the others both in its purpose and in the means of executing that purpose’ (Beal, 1999: 80) in that it uses a truly phonemic system of notation in which any one symbol always represents the same phoneme and vice versa. For instance, in *A New Dictionary of the English Language* Kenrick (1773) used a notation system based on numbers placed over each syllable, a method which he acknowledges was inspired by ‘the celebrated Mr Sheridan’ (Beal, 1999: 74). In the introduction to the work he gives readers ‘directions for consulting the following dictionary’ (1773: 1–8) and then elaborates on the description of the sounds in the ‘Rhetorical Grammar’ prefixed to it (1773: 1–57). He first provides a table of English sounds for vowels and another for consonants, taking note of spelling variation for the same sound, as shown in Table 5.

**Table 5 Kenrick (1773: v) on ‘the long and short modes of uttering our five vowels’**

A.		<i>barr’d.</i>		<i>bard.</i>
E.		<i>met.</i>		<i>mate.</i>
I.	short in	<i>hit.</i>	long in	<i>heat.</i>
O.		<i>not.</i>		<i>naught.</i>
U.		<i>pull.</i>		<i>pool.</i>

He goes on to explain the notation system with the word *fascination* as an illustrative example:

- (1) The word is next printed, as it is divided into syllables according to a right pronunciation, with figures placed over each syllable, to determine its exact sound, as the figures correspond with those of the above table of sounds: thus FA<sup>11</sup>S-CI<sup>15</sup>-NA<sup>12</sup>-TI<sup>1</sup>ON.]

Now, by referring to the table, we find that the several syllables are to be

pronounced like the words placed over against the numbers 11, 15, 12, 1; by which the quality of the sound, or the power of all the vowels, is exactly determined.

By shewing farther that the consonant *C* in the second syllable is printed in Italicks, it is known, by the table of consonants, that it is here pronounced soft like an *S*. Again, the letters *Tl* in the last syllable being printed also in Italics, it is plain from the same table that they have the usual power of *sh*; so that the word must be pronounced as if it had been printed FA<sup>11</sup>S-SI<sup>15</sup>-NA<sup>12</sup>-SHO<sup>1</sup>N.

(Kenrick, 1773: vii)

Kenrick's system is itself a reference for Perry's (1775) dictionary, which also takes inspiration from Johnston's (1764) method, and in turn is found in Sheridan (1780) in combination with Buchanan's (1757) respelling notations (Beal, 1999: 75, 78). The system in Walker (1791) is 'virtually identical to that devised by Sheridan' (Beal, 1999: 78–9). Walker argues that Sheridan's 'method of conveying the sound of words, by spelling them as they are pronounced, is highly rational and useful', and therefore it 'seemed to complete the idea' of Walker's own dictionary (Walker, 1791: iii). Fig. 1 shows a summary of Sheridan's notation system, where vowels are categorized 'by the titles of First, Second, and Third sounds, according to the order in which they lie, and as they are marked by those figures' (1780: 4), and where consonants are preceded by a vowel (first row) or by 'sounding' the characters so that 'their nature and powers will be expressed in their names' (1780: 5). As an illustrative example from the dictionary entries (see (2)), the example word *whisker* is documented by Sheridan with the consonant cluster *hw* in the first syllable (set WHALE\_a) and with the vowel *u*<sup>1</sup> in the second syllable (set *letter*), that is IPA /ʌr/.

Figure 1 Sheridan's (1780: 5) notation system for vowels and consonants<sup>11</sup>

Vowels.																	
a <sup>3</sup>	a <sup>2</sup>	a <sup>1</sup>	e <sup>3</sup>	o <sup>2</sup>	o <sup>3</sup>	e <sup>1</sup>	i <sup>1</sup>	u <sup>1</sup>									
hall	hat	hate	beer	note	noose	bet	fit	but									
		w			y												
		short o <sup>3</sup> o			short e <sup>3</sup> e												
Consonants.																	
eb	edef	eg	ek	el	em	en	ep	er	es	et	ev	ez	eth*	eth	esh	ezh	in
	h		c		j		q		x								
	ha		ek or ess		edge		qua		eks or egz								

- (2) WHISKER, hwi<sup>1</sup>s'-ku<sup>1</sup>r. s. The hair growing on the cheek unshaven, the mustachio. (Sheridan, 1780: s.v. *whisker*)

Once the correspondence between the dictionaries' systems and the IPA conventions was established (see Appendix III for a sample of two dictionaries), the relevant segment of each example word was transcribed using IPA symbols in an individual entry in the database. The

<sup>11</sup> Here, the symbols <w> and <y> refer to the semivowels /w/ and /j/.

following methodological principles were followed for the interpretation of all pronouncing dictionaries. First, the symbol /ɑ:/, which would be used for the vowel produced by RP speakers in the BATH, PALM, and START sets, has not been used in our IPA transcriptions; rather, we have consistently used /a:/ in line with the general view by historical phonologists that the backing to /ɑ:/ was a later process (e.g. Lass, 1999: 104). This concerns the sets BATH, PALM, START, and variants in FACE, LOT, SQUARE, THOUGHT, TRAP. Second, all the eighteenth-century dictionaries examined describe and/or prescribe a rhotic pronunciation. Since it is therefore a given that orthographic *r* is pronounced in all contexts, we have included post-consonantal /r/ in our transcriptions only when rhoticity is relevant to the pronunciation of the vowel in the example word, namely in the sets CURE, FORCE, *letter*, NEAR, NORTH, NURSE, SQUARE, START. In these sets, historical changes in the pronunciation of the vowels are connected to the presence or loss of rhoticity. The exceptions are the subsets CURE\_ci, CURE\_cii, FORCE\_c, NEAR\_c, NEAR\_f, NORTH\_c, SQUARE\_c, START\_c because the example words in these subsets all have the vowel before /r/ followed by another vowel, as in *boorish*, *curious*, and therefore rhoticity is not an issue.<sup>12</sup> In the sets SURE, FEATURE, HEIR post-consonantal /r/ has not been included in the transcription either, because the relevant segment in these sets is the prevocalic consonant, not the vowel. Third, where current transcription conventions vary, we have chosen the one that most closely corresponds with the descriptions provided by our eighteenth-century sources. For example, in transcribing the vowel of the lexical set FLEECE we have chosen /i:/ rather than /i/ because the majority of our sources describe this as a ‘long’ vowel.

Authors typically provide a single pronunciation; if they comment on variation in the pronunciation of a particular word we document that in a separate column, as shown in Table 6.

**Table 6 Illustrative examples of example words with IPA variants**

Lexical Set	Subset	Example word	IPA	IPA variant	Dictionary
BATH	BATH_a	plant	æ	a:	Walker 1791
CURE	CURE_ai	your	jo:r	jaɹ	Jones <sup>3</sup> 1798
FACE	FACE_a	great	e:	i:	Sheridan 1780
FOOT	FOOT	bosom	u	ʌ	Scott 1786
SURE	SURE_a	sure_cn	sju:	ʃju:	Johnston 1764
SQUARE	SQUARE_a	bear	e:r	i:r	Buchanan 1757
WHALE	WHALE_a	whist	hw	w	Kenrick 1773

In addition, if authors elaborate further on a context in which there is variation, the passage is recorded in the field Metalinguistic Comments. An example of this is the need to explain that a difference in pronunciation implies a difference in meaning, as noted by Buchanan (1757) for the lexical item *bear* (set SQUARE\_a): as a noun meaning ‘A wild beast’ it is pronounced *bēar* (IPA /i:r/), while as a verb meaning ‘To carry’ the pronunciation is *beār* (IPA /e:r/). If the remarks convey prescriptive attitudes towards either variant, this is further annotated in the fields for Attitudes (i.e. positive, negative, neutral) and Labels (e.g. vulgar, improper).<sup>13</sup> Criticism is usually related to pronunciations considered ‘vulgar’, whether in the sense ‘coarse, unrefined’ (OED s.v. *vulgar* II.13.d) and ‘mean; low’ (Johnson, 1755: s.v. *vulgar*, sense 2), or in

<sup>12</sup> There is a peculiar case in which post-consonantal /r/ stands in variation with /l/, namely in *colonel* (set NURSE) with IPA variants /ɹl/ and /lɹ/. Johnston (1764) simply lists the two variants without further comment: ‘cōlonel, cúrnél’, while Kenrick (1773) makes the following remark: ‘It is now generally sounded with only two distinct syllables, *col’nel*, and vulgarly as if written *cur-nel*’, that is IPA /ɒl/ and /lɹ/ respectively. Here we have preserved the /r/ in the transcription.

<sup>13</sup> Beal (1999: 48–58) discusses whether authors of pronouncing dictionaries were ‘good’ phoneticians or not, and how ‘descriptive’ or ‘prescriptive’ their remarks were.

the sense ‘commonly or customarily used by the people of a country; ordinary, vernacular’ (*OED* s.v. *vulgar* 1.3.a), often in phrases such as ‘the vulgar say’ or ‘among the vulgar’ (see also Sundby *et al.*, 1991: 40–2, 52–3). Walker’s entry for *plant* (set BATH\_b) in passage (3) provides an illustrative example of this. For his part, the Irish author Sheridan often comments on variation between English and Irish pronunciation, as in the section on ‘Rules to be observed by the Natives of Ireland in order to attain a just Pronunciation of English’ (1780: 59–62). See, for instance, his passage in (4) about lexical items such as *great* (set FACE\_a), where he warns ‘the gentlemen of Ireland’ to avoid the mistaken pronunciation /i:/ for the ‘just’ pronunciation /e:/ in English. Sheridan emphasizes that ‘[a] strict observation of these few rules [...] will enable the well-educated natives of Ireland to pronounce their words exactly in the same way as the more polished part of the inhabitants of England do’ (1780: 60).

(3) PLANT, pla<sup>4</sup>nt. [IPA /æ/]

↪ There is a coarse pronunciation of this word, chiefly among the vulgar, which rhymes it with *aunt* [i.e. a<sup>2</sup>nt, IPA /a:/]. This pronunciation seems a remnant of that broad sound which was probably given to the *a* before two consonants in all words, but which has been gradually wearing away, and which is now, except in a few words, become a mark of vulgarity. (Walker, 1791: s.v. *plant*; s.v. *aunt*)

(4) The second vowel, *e*, is for the most part sounded *ee* by the English [IPA /i:/], when the accent is upon it; whilst the Irish in most words give it the sound of second *a*<sup>2</sup>, as in *hate* [IPA /e:/]. This sound of *e*<sup>3</sup> [ee] is marked by different combinations of vowels, such as, *ea*, *ei*, *e* final mute, *ee*, and *ie*. [...] The English constantly give this sound [i.e. /i:/] to *ea*, whenever the accent is on the vowel *e*, except in the following words, *gre*<sup>2</sup>*at*, a *pe*<sup>2</sup>*ar*, a *be*<sup>2</sup>*ar*, to *be*<sup>2</sup>*ar*, to *forbe*<sup>2</sup>*ar*, to *swe*<sup>2</sup>*ar*, to *te*<sup>2</sup>*ar*, to *we*<sup>2</sup>*ar*. In all which the *e* has its second sound [e<sup>2</sup>, IPA /e:/]. For want of knowing these exceptions, the gentlemen of Ireland, after some time of residence in London, are apt to fall into the general rule, and pronounce these words as if spelt, *greet*, *beer*, *sweer*, &c. (Sheridan, 1780: 59)

Finally, since word frequency may be an influential factor in the choice of variants or in the development of sound changes such as those arising through lexical diffusion, we have compiled a frequency list with an estimated frequency rate of the lexical item in eighteenth-century British English, based on the data available in the multi-genre historical corpus ARCHER 1650–1999, version 3.2 (535,767 words).

## 4. Web-based Interface

The ECEP database will be made available to users via a web-based application hosted on the website of the Humanities Research Institute, University of Sheffield. Access to ECEP will be free for any user registering at the website. The reference line for citation is as follows:

ECEP = *Eighteenth-Century English Phonology* database, 2015. Compiled by Joan C. Beal, Nuria Yáñez-Bouza, Ranjan Sen and Christine Wallis. The University of Sheffield and Universidade de Vigo. Published by: University of Sheffield.  
<http://hridigital.shef.ac.uk/eighteenth-century-english-phonology>

The online interface has been developed using client-side HTML and Javascript and server-side PHP and MySQL. It displays two layouts – Browse, Search – and offers a download function in CVS file format. The design aims to replicate the MS Access format, and therefore it offers three main blocks of data: the lexical sets plus metadata for works and for authors (see

Table 1). Fig. 2 shows the homepage, from which each of these sections can be accessed (see top row), and from which users can go directly to the pronouncing dictionary they are interested in (see Buchanan 1757 and Burn 1786 in the image). Fig. 3 and Fig. 4 are screenshots of the Browse layouts for Works and Authors, respectively. The Search tool allows users to search in one field or in a combination of fields. Fields which contain a predefined list of values (e.g. lexical sets, example words, author's name) offer an automatic drop-down list menu to facilitate selection, as in the field IPA in Fig. 5. Lexical sets and example words can be searched in the entire database or within a particular work; for instance, Fig. 6 displays a sample of the set BATH, where users can compare the occurrence of the variants /æ/, /a:/ and /ɔ:/.

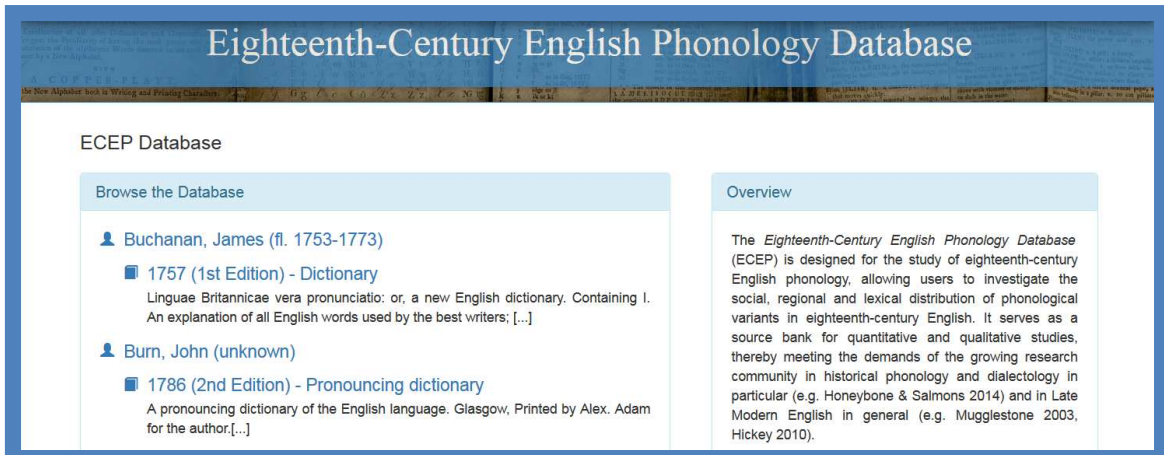


Figure 2 ECEP online interface – homepage

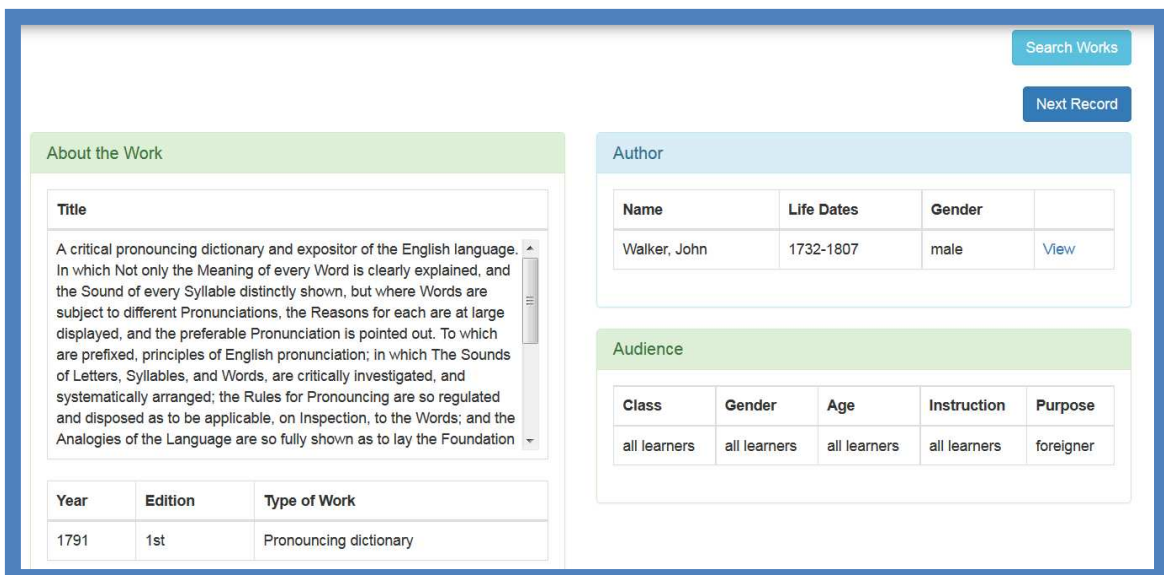


Figure 3 ECEP online interface – Works in Browse layout

[Search Authors](#)

[Previous Record](#) [Next Record](#)

### About the Author

Name	Life Dates	Gender	Class
Kenrick, William	1729/30-1779	male	unknown

Place of Birth	Occupation	Occupation Description
London, Middlesex, England	Miscellaneous, Writing	playwright, poet, author, critic, translator

### Biography

"By William Kenrick, LL.D." (title-page)

"Writer and translator, was the son of Robert Kenerick, staymaker, of Hemel Hempstead, Hertfordshire, and his wife, Mary (1708/9–1789). A member of a Baptist family, Kenrick attended probably either Philip James's Baptist school or Thomas Squire's Quaker school. Apprenticed in 1745 to Thomas Bennett, a mathematical instrument maker in London, Kenrick abandoned his apprenticeship after three years, in favour of a career in literature."

[Author Works](#)

Figure 4 ECEP online interface – Authors in Browse layout

## Eighteenth-Century English Phonology Database

Lexical Search

Lexical Categories	Lexical SubSets	Lexical Keywords	IPA	IPA Variant	Attitudes	Labels
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text" value=" "/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="button" value="v"/>	<input type="button" value="v"/>	<input type="button" value="v"/>	<ul style="list-style-type: none"> <li>ai</li> <li>air</li> <li>ar</li> <li>au</li> </ul>	<input type="button" value="v"/>	<input type="button" value="v"/>	<input type="button" value="v"/>
<input type="button" value="Clear the Search"/>						<input type="button" value="Search Lexical Sets"/>

Figure 5 ECEP online interface – Lexical Sets in Search layout

All Lexical Sets

Show  entries Filter 2:

<input type="button" value="v"/>	Lexical Category	Subset	Keywords	IPA	IPA Variant	MetalexComments	Attitudes	Labels	Notes_exx
<a href="#">View Work</a>	BATH	BATH_c	calf	a:					
<a href="#">View Work</a>	BATH	BATH_c	calf	æ	ɔ:	Sheridan's Irishisms are listed in the 'Rules to be observed by the Natives of Ireland in order to attain a just Pronunciation of English', Appendix pp.59-62.	Negative	Irishism	Headword CALF. Awareness of variation: /æ/ and /ɔ:/, the latter criticised.
<a href="#">View Work</a>	BATH	BATH_c	calf	a:		"ca10lf and sometimes ca10f, as if the l were silent."			Headword CALF. Awareness of variation regarding consonant sounds.

Figure 6 ECEP online interface – Lexical Sets in Browse layout



## 5. Case Studies

In this section we report on two case studies that demonstrate the value of evidence that can be systematically extracted from this database for the analysis of segmental and suprasegmental phonology, in their regional and chronological settings. The results constitute a valuable distillation of the conditioning factors to look out for in a wider range of eighteenth-century evidence, hence a point of departure for further investigation. In this light, these results must be interpreted as indications of patterns rather than definitive analyses; that is, if a sub-set of dictionary writers display a pattern in their choices, it is worth exploring that pattern using all the available evidence to establish whether a conditioning factor in the sound change indeed underlies it.

The first of these studies examined variation in the pronunciation of ‘wh’ (/hw~/w~/h/) in example words of the consonantal set WHALE (Beal and Sen, 2014a; 2014b). In present-day RP, words such as *whale*, *what*, *where* begin with the sound /w/, whilst *who*, *whole* have initial /h/. Eighteenth-century pronouncing dictionaries present evidence, through their orthographic systems, of variation between /hw/ and /w/ for the first set, hence a preserved versus unpreserved contrast in *where/wear*. The fifty example words in this consonantal set were selected on the basis of their occurrence in as many of the sources as possible, and to represent three phonological contexts: (1) thirty-nine example words beginning with the spelling ‘wh’ which are pronounced with /w/ in present-day RP, (2) six example words with initial ‘wh’ which are now pronounced with initial /h/, and (3) five example words with ‘wh’ word internally, which are now all pronounced with internal /w/ (e.g. *somewhere*). The transcriptions of the ‘wh’ segment in each example word found in nine of the eleven pronouncing dictionaries compiled in ECEP are displayed in Table 7.

**Table 7 Transcriptions of the ‘wh’ segment in the WHALE lexical set**

WHALE set	Bu57	Joh64	Ke73	Pe75	Sp75	Sh80	Bur86	Wa91	Jo97
whale	hw	hw	w	w	hw	hw	w	hw	hw
wharf	hw	w	w	w	hw	hw	w	hw	hw
what	hw	hw	w	w	hw	hw	w	hw	hw
wheat	hw	hw	w	w	hw	hw	w	hw	hw
wheedle	hw	hw	w	w	hw	hw	w	hw	hw
wheel	hw	hw	w	w	hw	hw	w	hw	hw
wheeze	hw	hw	w	w	hw	hw	w	hw	hw
whelm	hw	hw	hw	hw	hw	hw	w	hw	hw
whelp	hw	hw	w	w	hw	hw	w	hw	hw
when	NID	hw	w	w	hw	hw	w	hw	hw
whence	NID	hw	w	w	hw	hw	w	hw	hw
where_cn	NID	hw	w	w	hw	hw	w	hw/w	hw
wherry	hw	hw	hw	w	hw	hw	w	hw	hw
whet	hw	hw	w	w	hw	hw	w	hw/w	hw
whether	hw	hw	hw	w	hw	hw	w	hw	hw
whey_cn	hw	hw	w	w	hw	hw	w	hw	hw
which	NID	hw	w	w	hw	hw	w	hw	hw
whiff	hw	hw	w	w	hw	hw	w	hw	hw
whiffle	hw	hw	w	w	hw	hw	w	hw	hw
whig	hw	hw	w	w	hw	hw	w	hw	hw
while	NID	hw	w	w	hw	hw	w	hw/w	hw
whim	hw	hw	w	w	hw	hw	w	hw	hw
whimper	hw	hw	w	w	hw	hw	w	hw	hw
whin	hw	NID	w	w	hw	hw	w	hw	hw
whine	hw	hw	w	w	hw	hw	w	hw	hw
whip	hw	hw	w	w	hw	hw	w	hw	hw
whirl	hw	hw	w	w	hw	hw	w	hw	hw
whisk	hw	hw	hw	hw	hw	hw	w	hw	hw

whisker_cn	hw	hw	hw	hw	hw	hw	w	hw	hw
whisper	hw	hw	hw	hw	hw	hw	w	hw	hw
whist	hw	hw	hw/w	w/hw	hw	hw	w	hw	hw
whistle	hw	hw	hw	w	hw	hw	w	hw	hw
whit	hw	hw	w	w	hw	hw	w	hw	hw
white	hw	hw	w	w	hw	hw	w	hw	hw
whither	hw	hw	w	w	hw	hw	w	hw	hw
whitlow	hw	hw	hw	hw	hw	hw	hw	hw	hw
whitsuntide	hw	hw	hw	w	hw	hw	hw	hw	hw
whiz	hw	hw	hw	hw	hw	hw	hw	hw	hw
who_cn	NID	h	h	h	hw	h	h	h	h
whole	h	h	h	h	hw	h	h	h	h
whom	NID	NID	h	h	hw	h	h	h	h
whoop	hw	h	h	h	hw	h	h	h	h
whore_cn	h	h	h	h	h	h	h	h	h
whose_cn	NID	h	h	h	hw	h	h	h	h
why	NID	hw	w	w	hw	hw	hw	hw	hw
elsewhere	NID	NID	NID	hw	hw	hw	w	hw	hw
nowhere	NID	NID	NID	w	hw	hw	w	hw	hw
overwhelm	hw	hw	NID	hw	hw	hw	hw	hw	hw
somewhat	NID	NID	NID	w	hw	hw	w	hw	hw
somewhere	NID	hw	NID	w	hw	hw	w	hw	hw

This systematic data collection even on such a small scale enabled us to identify patterns in the evidence, along dimensions commonly under investigation in sociolinguistic, historical and phonological research, namely geography, chronology, phonology, lexical factors, and social class. Furthermore, the nature of the data also enabled us to glean ‘direct’ evidence in the form of contemporary commentary on the choices made by the authors. A notable example is that Walker (1791) presents the loss of the /hw ~ w/ contrast as a special case of ‘h-dropping’ in lower-class London English, which was just beginning to attract social stigma in the middle of the eighteenth century (Beal, 1999: 176–8).

Three main patterns emerged from the data based on geographical and chronological distribution. Firstly, the London authors prefer /hw/ to /w/ to avoid the proscribed ‘h-dropping’ as discussed by Walker, with the exception of Kenrick (1773), one of the earliest of the group, presumably because the stigmatization of /w/ had not yet fully taken effect by this time. Secondly, two out of the three Scottish authors prefer /w/ (Perry, 1775; Burn, 1786), whereas the earliest, Buchanan (1757), prefers /hw/. Perry and Burn appear to be advising a more London-like pronunciation to avoid the Scottish /hw/, stigmatized due to its regional connotations (Douglas, 1991 [1779]: 141). The /w/ pronunciation could therefore be analysed as a hypercorrect Anglicism, one which is particularly remarkable in the light of the contemporaneous opposite trend in London where /hw/ was proscribed due to ‘h-dropping’. Arguably, this trend was only taking hold in London at the time and had not yet reached the consciousness of the Scottish authors. Thirdly, Spence (1775) from Newcastle in north-east England has near-consistent /hw/, even in words containing a following back, rounded vowel, where other authors have delabialized /h/ e.g. *who*. Along with the fact that Spence is the only author to use a special symbol for the ‘wh’ sound, this could be interpreted as evidence in Spence’s dialect for monosegmental /m/, and not a cluster /hw/, the voiceless counterpart of voiced /w/ which also retained its labial element before back, rounded vowels, e.g. *wound*, *womb*, *wool*, *wood*.

Two lexically based patterns emerged. The first, homophone avoidance, as shown by Buchanan’s (1757) /hw/ for *whoop* ‘a cry’, but /w/ for *whoop* ‘a bird’ and Burn’s (1786) and Perry’s (1775) *Whitsuntide* with /hw/ and *whit* with /w/, is evidence that sensitivity to the contrast remained to a sufficient degree to construct minimal pairs in some regions, notably Scotland where the contrast survives to the present day. The second, onomatopoeia as

illustrated by /hw/ in *whisk*, *whisper* in Kenrick (1773) and Perry (1775), could also be interpreted as evidence for an increased chance of /h/-preservation (perhaps enhanced by considerations of sound symbolism) before a front vowel in precisely these two authors, e.g. *whelm*, and often with a following /s/, e.g. *whisk*, *whiskers*, *whisper*.

Aside from this partial pattern, two main explanations based on phonological context emerged. The first is the unambiguous delabialization to /h/ before any vowel that is higher and more round than /ɔ/ (there is no /h/ in *wharf* in any of the dictionaries), e.g. *who*. Secondly, the realization of word-internal ‘wh’ in Perry (1775) appears to be conditioned by stress, as marked by the author himself, thus stressed-syllable onset /hw/ in *overwhélm*, *elsewhére*, but unstressed-syllable onset /w/ in *sómewhere*, *sómewhat*, *nówhere*.

We therefore repeatedly found that by systematically collating the different types of direct evidence afforded by the eighteenth-century pronouncing dictionaries (sounds and stress, contemporary commentary, geographical and chronological spread), and analysing them in the light of acknowledged influences in sound change, we were able to posit accounts for many of the patterns in a way that only such an orderly approach to the data permitted.

The second case study explored palatalization in eighteenth-century English, i.e. where a postalveolar fricative /ʃ ʒ/ or affricate /tʃ dʒ/ arose from the sequence alveolar /t d s z/ + /j/ + /u:/, as in the word *tune* (Beal and Sen, 2015). The palatalization of alveolar consonants before late Middle English /u:/ is still variable and is diffusing in present-day English. The OED gives several pronunciations for *mature* (e.g. /mə'tʃʊə ~ mətʃʊə/), but provides only unpalatalized (/dj tʃ/) transcriptions for *endure*, *tune*, and *duke*, despite the common occurrence of palatalized (and yod-dropped) variants in many varieties of British English. Extensive variability is not recent in origin, and we can already detect relevant patterns in the eighteenth century from the evidence of a range of pronouncing dictionaries; for instance, Beal (1996; 1999) notes a tendency for northern English and Scottish authors to be more conservative. She concludes that we require ‘a comprehensive survey of the many pronouncing dictionaries and other works on pronunciation’ (1996: 379) to gain more insight into the historical variation patterns underlying present-day English. This study presented results from such a ‘comprehensive survey’ base on the data compiled in ECEP.

The data were divided into two main consonantal lexical sets: DEUCE where there was no /r/ following the vowel, and SURE where an /r/ followed. A third set was FEATURE, where the vowel following the palatalized sequence is schwa in present-day English, and /r/ originally followed the vowel. This division was made after preliminary examination demonstrated a clear difference in the behaviour of the consonantal sequences in these contexts. We were then able to further clarify the nature of the divergence after constructing the database with information from ten dictionaries, and with word-frequency information for the period 1700–1799 from ARCHER 3.2. As mode of illustration, Table 8 displays transcriptions of example words in some of the DEUCE subsets, and Table 9 transcriptions of some of the SURE and FEATURE subsets.

**Table 8 Transcriptions of the DEUCE lexical set in ECEP: subsets DEUCE\_a /t/, DEUCE\_b /t/, DEUCE\_c /t/, DEUCE\_b /s/**

DEUCE_a /t/	Bu57	Joh64	Ke73	Pe75	Sp75	Sh80	Sc86	Wa91	Jo97	Jo98
opportunity	tju:	tju:	tju:	tju:	tju:	tju:	tju:	tju:	tju:	tju:
Tuesday	tju:	tju:	tju:	tju:	tju:	tʃu:	tju:	tju:	tju:	tju:
tumour	tju:	tju:	to:	tju:	tju:	tʃu:	tju:	tju:	tju:	tju:
tube	tju:	tju:	tju:	tju:	tju:	tʃu:	tju:	tju:	tju:	tju:
tutor	tju:	tju:	tju:	tju:	tju:	tʃu:	tju:	tju:	tju:	tju:
tune_cn	tju:	tju:	tju:	tju:	tju:	tʃu:	tju:	tju:	tju:	tju:



maturity_cn	tju:	tju:	tju:	tju:	tju:	tju:	tju:	tju:	tju:	tju:
-------------	------	------	------	------	------	------	------	------	------	------

<b>SURE_c/t/</b>	Bu57	Joh64	Ke73	Pe75	Sp75	Sh80	Sc86	Wa91	Jo97	Jo98
maturation	tju:	NID	tju:	tju:	NID	tju:	tju:	tʃju:	tju:	tʃju:

<b>SURE_a /s/</b>	Bu57	Joh64	Ke73	Pe75	Sp75	Sh80	Sc86	Wa91	Jo97	Jo98
sure_cn	sju:	sju:	ʃju:	ʃju:	ʃu:	ʃu:	sju:	ʃju:	ʃu:	ʃju:
assure_cn	sju:	sju:	ʃju:	sju:	ʃu:	ʃu:	sju:	ʃju:	ʃu:	ʃju:
assurance_cn	sju:	sju:	ʃju:	sju:	ʃu:	ʃu:	sju:	ʃju:	ʃu:	sju:
insure_cn	NID	sju:	NID	NID	NID	NID	sju:	NID	NID	NID
ensure_cn	NID	NID	sju:	ʃju:	NID	NID	NID	ʃju:	sju:	ʃju:
surety	sju:	sju:	ʃju:	ʃju:	ʃu:	ʃu:	sju:	ʃju:	ʃu:	ʃju:
insurance_cn	sju:	sju:	NID	NID	NID	sju:	sju:	NID	NID	NID
unsure	NID	sju:	ʃju:	ʃju:	NID	ʃu:	sju:	ʃju:	NID	ʃju:

<b>SURE_b /z/</b>	Bu57	Joh64	Ke73	Pe75	Sp75	Sh80	Sc86	Wa91	Jo97	Jo98
composure	zju:	zʌ	zju:	zʌ	ʒʊ	zʌ	sjʌ	zju:	zʌ	zju:
seizure	zju:	zʌ	zʌ	NID	zju:	zʌ	sjʌ	zju:	zʌ	zʌ
azure_SURE	zju:	zʌ	zɔ	zʌ	ʒʊ	zʌ	zjʌ	zju:	zʌ	zju:
closure	NID	NID	zʌ	zʌ	zʌ	zʌ	NID	zju:	zʌ	zʌ

<b>FEATURE /z/</b>	Bu57	Joh64	Ke73	Pe75	Sp75	Sh80	Sc86	Wa91	Jo97	Jo98
pleasure	zju:	zʌ	zʌ	zʌ	ʒʊ	zʌ	zjʌ	zju:	zʌ	zʌ
measure_cn	zju:	zʌ	zʌ	zʌ	ʒʊ	zʌ	zjʌ	zju:	zʌ	zʌ
treasure	zju:	zʌ	zʌ	zʌ	ʒʊ	zʌ	zjʌ	zju:	zʌ	zju:
leisure	zju:	NID	zʌ	zʌ	ʒʊ	zʌ	zjʌ	zju:	zʌ	zʌ
azure_FEAT	zju:	zʌ	zɔ	zʌ	ʒʊ	zʌ	zjʌ	zju:	zʌ	zju:
rasure	sju:	sʌ	sʌ	ʃʌ	zju:	ʃʌ	zjʌ	zju:	ʃʌ	zju:

All the pronouncing dictionaries are consistently rhotic, i.e. they report syllable-final /r/ in forms such as *sure*. It was found that there is significantly more palatalization when /r/ follows (SURE/FEATURE) than when it did not (DEUCE), particularly in a post-stress syllable, thus even the resistant Spence (1775) has /z/ in *closure*, *pleasure*. The more frequent of these palatalized forms (e.g. *nature*) seem to be the words which have become lexicalized in present-day English.

The nature of the palatalizing phoneme was also relevant. Palatalization occurred in /sj/ in particular, thus it is near-regular in post-stress DEUCE in Perry (1775), Sheridan (1780), Walker (1791), and Jones (1797, 1798), e.g. /ʃ/ in *issue*. This is arguably because the high tongue position of palatal /j/ shapes frication noise, producing post-alveolar percepts. Furthermore, /sj/ is the only context which palatalizes in a stressed syllable with any regularity, particularly when in a rhotic context, thus *sure*, *surety* with /ʃ/ even in Kenrick (1773), Perry (1775), and Spence (1775). Stress therefore also appears to have been a conditioning factor, with palatalization generally resisted in the onset of a stressed syllable, as noted explicitly by Walker (1791), and more common in post-stress syllables. Pre-stress syllables also show some palatalization, yielding interesting alternations such as /tj/útor but /tʃ/utórial and ma/tj/úre but ma/tʃ/urátion in Walker (1791).

Two other contexts proved to be more conducive to palatalization: word-initial position, thus Sheridan (1780) /tʃ/ in *tune*, but /tj/ in *attune*, and before vowel hiatus, thus /tʃ/ in *punctual*, *sanctuary* in Sheridan (1780), Walker (1791), and Jones (1797; 1798), but mainly /tj/

elsewhere.

As with the 'wh' study, chronology, geography, and stigmatization also proved to be relevant factors in accounting for the variation. Palatalization appears to have become increasingly more common over the course of the eighteenth century: there is little in Kenrick (1773), but Sheridan (1780; late in career) is the arch-palatalizer. However, the latter's dictionary was repeatedly singled out for criticism later in the century, as such pronunciations came to be stigmatized (e.g. Jones, 1798: iv). Palatalization consequently became much less common at the end of the century; it is less widespread, but stress-based in Walker (1791; see his principles 376, 450, 459–64), and progressively even less common from Jones' second edition (1797) to his third (1798). In terms of geography, Sheridan's palatalizing tendencies were attributed at the time to his Irish origin; this contemporary explanation requires further scrutiny as there is little evidence that palatalization was common in the Irish English of the time. There is little palatalization in the Scottish sources, with Buchanan (1757; early source), and Scott (1786) notably having no palatalized forms whatsoever. Spence (1775) from Newcastle also has little palatalization. Palatalization in the late-middle part of the eighteenth century may have increased due to the earlier restitution of post-consonantal yod in earlier yod-dropped forms, as in the London-based 'metropolitan pronunciation' criticized by Kenrick (1773). For example, the earlier sources almost all have yod-dropped /t/ in *creature* (Johnston, 1764; Kenrick, 1773; Perry, 1775), but the later ones have /tʃ/ (Sheridan, 1780; Walker, 1791; Jones, 1797; 1798). Furthermore, this observation forms part of a further pattern revealed by the database: there were two chronologically and phonologically distinct yod-droppings. The first, mentioned above, notably occurred in the earlier sources after all phonemes /t d s z/ in *unstressed* syllables before /r/. The second yod-dropping occurred in the later sources in a different context: after any phoneme in a *stressed* syllable. Sheridan (1780) is the earliest to do this in the single example *dual*; Scott (1786) is the most frequent ommitter of stressed yod, mostly in fricative and only in the most frequent words, e.g. *duty*.

## 6. Conclusion

In this paper we have presented a new digital resource for the study of English historical phonology: the *Eighteenth-Century English Phonology* database (ECEP). The database provides IPA transcriptions for the relevant segment of each example word in Wells' (1982) lexical sets for the vowel system of present-day English, and some complementary consonant sets, as documented in a selection of eighteenth-century pronouncing dictionaries. We have described the structure and content of ECEP, while reporting on the methodology of compilation: source selection, data input and annotation, and the web-based interface for users. ECEP is already available, but work will continue with a view to enlarging the database gradually.

Originally designed as a sister to the *Eighteenth-Century English Grammars* database (ECEG, 2010), on the practical side ECEP will help to promote the use of databases as research resources in historical linguistics, beyond or alongside largely available text corpora. In terms of content, ECEP will contribute to English historical phonology, dialectology and sociolinguistics, with a focus on the eighteenth century, but will also be of use for comparative studies with nineteenth-century English or present-day English.

The two case studies outlined in Section 5 demonstrate the potential of ECEP as a resource for investigating the historical phonology of Late Modern English. The database has also been used in studies of the CLOTH lexical set (Beal and Condorelli, 2014) and of the use of labels in the enregisterment of non-standard pronunciation (Beal and Trapateau, in prep.). The availability of this resource will ensure that in the future historical phonology will no longer be the 'poor relation' of Late Modern English studies (Beal, 2012b: 27).

## Funding

This work was supported by the British Academy / Leverhulme Trust [SG–132806] and the Santander Research Mobility Scheme (calls 2012–13 and 2014–15). We are also grateful to the Humanities Research Institute at the University of Sheffield for technical support, in particular Michael Pidd (Digital Director) and Ryan Bloor (Developer). Nuria Yáñez-Bouza would like to thank the Spanish Ministry of Economy, Ramón y Cajal Scheme [RYC-2011-07863] and the research group *Language Variation and Textual Categorisation* at the University of Vigo, the European Regional Development Fund [FFI2013-44065-P], and the Autonomous Government of Galicia [GPC2014/060].

## References

### Primary Sources

- Buchanan, J. (1757). *Linguae Britannicae vera pronuntiatio: or, A new English dictionary*. London.
- Burn, J. (1786). *A pronouncing dictionary of the English language*. Glasgow.
- Johnston, W. (1764). *A pronouncing and spelling dictionary*. London.
- Jones, S. (1797). *Genuine edition. Sheridan improved: A general pronouncing and explanatory dictionary of the English language*, 2nd edition. London
- Jones, S. (1798). *Sheridan improved: A general pronouncing and explanatory dictionary of the English language*, 3rd edition. London
- Kenrick, W. (1773). *A new dictionary of the English language*. London.
- Perry, W. (1775). *The royal standard English dictionary*. Edinburgh.
- Scott, W. (1786). *A new spelling, pronouncing, and explanatory dictionary of the English language*. Edinburgh.
- Sheridan, T. (1780). *A general dictionary of the English language*. London.
- Spence, T. (1775). *The grand repository of the English language*. Newcastle.
- Walker, J. (1791). *A critical pronouncing dictionary and expositor of the English language*. London.

### Secondary Sources

- Agha, A. (2003). The social life of cultural value. *Language and Communication*, 23: 231–73.
- Alston, R. C. (1966). *A bibliography of the English language from the invention of printing to the year 1800*. Vol. V, *The English dictionary*. Leeds: Arnold and Son.
- ARCHER 3.2. *A Representative Corpus of Historical English Registers* version 3.2. 1990–1993/2002/2007/2010/2013. Originally compiled under the supervision of Douglas Biber and Edward Finegan at Northern Arizona University and University of Southern California; modified and expanded by subsequent members of a consortium of universities. Current member universities are Northern Arizona, Southern California, Freiburg, Heidelberg, Helsinki, Uppsala, Michigan, Manchester, Lancaster, Bamberg, Zurich, Trier, Santiago de Compostela and Leicester. [www.manchester.ac.uk/archer](http://www.manchester.ac.uk/archer) (accessed January 2015).
- Beal, J. C. (1996). The Jocks and the Geordies: Modified standards in eighteenth-century pronouncing dictionaries. In Britton, D. (ed.), *English historical linguistics 1994. Papers from the 8th International Conference on English Historical Linguistics, Edinburgh, 19–23 September 1994*. Amsterdam: John Benjamins, pp. 363–82.
- Beal, J. C. (1999). *English pronunciation in the eighteenth century: Thomas Spence's Grand Repository of the English Language (1775)*. Oxford: Clarendon Press.
- Beal, J. C. (2003). John Walker: Prescriptivist or linguistic innovator? In Dossena, M. and Jones, C. (eds.), *Insights into Late Modern English*. Bern: Peter Lang, pp. 83–105.
- Beal, J. C. (2004). *English in Modern times 1700–1945*. London: Hodder Arnold.
- Beal, J. C. (2012a). Late Modern English. In Brinton, L. and Bergs, A. (eds.), *Historical linguistics of English: An international handbook*. Berlin: Mouton de Gruyter, pp. 63–78.
- Beal, J. C. (2012b). 'Can't see the wood for the trees?' Corpora and the study of Late Modern English. In Markus, M., Iyeyri, Y., Heuberger, R. and Chamson, E. (eds.), *Middle and Modern English corpus linguistics: A multi-dimensional approach*. Amsterdam: Benjamins, pp. 13–30.
- Beal, J. C. (2012c). 'By those provincials mispronounced': The STRUT vowel in eighteenth-century pronouncing dictionaries. *Language and History*, 55(1): 5–17.
- Beal, J. C. & Condorelli, M. (2014). Cut from the same CLOTH? Variation and change in the CLOTH



- lexical set. *Token: A Journal of English Linguistics*, 3: 15–36.
- Beal, J. C., Nocera, C. and Sturiale, M., eds. (2008). *Perspectives on prescriptivism*. Bern: Peter Lang.
- Beal, J. C. and Sen, R. (2014a). Towards a corpus of eighteenth-century English phonology. In Davidse, K., Gentens, C., Kimps, D. and Vandelanotte, L. (eds), *Recent advances in corpus linguistics: Developing and exploiting corpora*. Amsterdam: Rodopi, pp. 31–53.
- Beal, J. C. and Sen, R. (2014b). *W(h)o, w(h)en, w(h)ere, and w(h)at?*. The eighteenth-century pronunciation of ‘wh’. Paper presented at the 18th International Conference on English Historical Linguistics (ICEHL 18), Leuven 14–18 July 2014.
- Beal, J. C. and Sen, R. (2015). En[dj]uring [tj]unes or ma[tj]ure [dʒ]ukes? Palatalization in eighteenth-century English: Evidence from the *Eighteenth-Century English Phonology Database*. Paper presented at the 9th Studies in the History of the English Language Conference (SHEL-9), Vancouver 5–7 June 2015.
- Beal, J. C. and Trapateau, N. (In preparation). ‘Named and Shamed’: Labels for non-standard pronunciation in 18<sup>th</sup>-century pronouncing dictionaries. In Brownlees, N., Iamartino, G. and Sturiale, M. (eds.), *Labelling English, English labelled: From the 9th century to Late Modern times*.
- Bloomfield, M. W. and Newmark, L. (1963). *A linguistic introduction to the history of English*. New York: Knopf.
- Cooper, C. (1687). *The English teacher, or, The discovery of the art of teaching and learning the English tongue*. London.
- Dobson, E. J. (1957). *English pronunciation 1500–1700*. Oxford: Clarendon Press.
- Douglas, S. (1991 [1779]). *A treatise on the provincial dialect of Scotland*, edited by C. Jones. Edinburgh: Edinburgh University Press.
- Eighteenth-Century English Grammars* database (ECEG). 2010. Compiled by Nuria Yáñez-Bouza (University of Manchester) and María E. Rodríguez-Gil (University of Las Palmas de Gran Canaria). [www.manchester.ac.uk/eceg](http://www.manchester.ac.uk/eceg) (accessed January 2015).
- Hickey, R., ed. (2010). *Eighteenth-century English. Ideology and change*. Cambridge: Cambridge University Press.
- Holmberg, B. (1964). *On the concept of Standard English and the history of Modern English pronunciation*. Lund: Gleerup.
- Honeybone, P. and Salmons, J., eds. (2015). *The Oxford handbook of historical phonology*. Oxford: Oxford University Press.
- Johnson, S. (1755). *A dictionary of the English language*. London.
- Jones, C. (1989). *A history of English phonology*. London: Longman.
- Jones, C. (2006). *English pronunciation in the eighteenth and nineteenth centuries*. Basingstoke: Palgrave Macmillan.
- Lass, R. (1999). Phonology and morphology. In Lass, R. (ed.), *The Cambridge history of the English language*, Vol. III: 1476–1776. Cambridge: Cambridge University Press, pp. 56–186.
- MacMahon, M. K. C. (1998). Phonology. In Romaine, S. (ed.), *The Cambridge history of the English language*, Vol. IV: 1776–1997. Cambridge: Cambridge University Press, pp. 373–535.
- Mugglestone, L. (2003). *‘Talking proper’: The rise of accent as social symbol*, 2nd edition. Oxford: Clarendon Press.
- Oxford Dictionary of National Biography* (ODNB). [www.odnb.com](http://www.odnb.com) (accessed January 2015).
- Oxford English Dictionary* (OED). [www.oed.com](http://www.oed.com) (accessed January 2015).
- Ranson, R. (2012). Some aspects of Walker’s methodology: A decisive step towards standardization? *Language and History*, 55(1): 18–33.
- Sheldon, E. K. (1947). Walker’s influence of the pronunciation of English. *Proceedings of the Modern Language Association of America*, 62: 130–46.
- Strang, B. M. H. (1970). *A history of English*. London: Methuen.
- Sundby, B., Bjørge, A. K. and Haugland, K. E. (1991). *A dictionary of English normative grammar 1700–1800*. Amsterdam: John Benjamins.

- Tieken-Boon van Ostade, I., ed. (2008). *Grammars, grammarians and grammar-writing in eighteenth-century England*. Berlin: Mouton de Gruyter.
- Tieken-Boon van Ostade, I. (2009). *An introduction to Late Modern English*. Edinburgh: Edinburgh University Press.
- Trapateau, N. (2016). 'Pedantick', 'polite' or 'vulgar'? A systematic analysis of eighteenth-century normative discourse on pronunciation in John Walker's dictionary (1791). *Language and History* 59(1): 25–36.
- Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

## Appendices

### Appendix I. Consonant sets and example words

SET	SUBSET	EXAMPLE WORD
DEUCE	DEUCE_a	assume, attune, consume, contusion, deuce_cn, dual, dubious, due, duel, duke_cn, duly, dupe_cn, duplicate, duty_cn, exuberant, exude, fiducial, fiduciary, indubitable, obtuse_cn, opportune_a, opportunity, presume, resume, sudatory, sudorous, suicide, suit, suitable, suitor, supine, suture_deu, tube, tuberos, tubular, Tuesday, tulip, tumid, tumour, tumult, tune_cn, tunic, tutor, zeugma
	DEUCE_b	actuary_deu, altitude, amplitude, aptitude, arduous, attitude, bitumen, casual, casualty, consular, consummate, fortitude, fraudulent, glandulous, gradual, incredulous, insulate, issue, latitude, longitude, magnitude, modulate, module, mortuary_deu, opportune_b, punctual, sanctuary, solitude, tissue, visual
	DEUCE_c	adulation, duplicity, insulation, modulation, sudation, sudorific, superb, superior, superlative, supremacy, supreme, tumultuous, tutorial
FEATURE	FEATURE	azure_ture, creature, feature_cn, fissure_ture, future, leisure, measure_cn, nature_cn, ordure_ture, pleasure, pressure_cn, procedure, rasure, suture_ture, torture_cn, treasure
HEIR	HEIR	heir_cn, heiress, herb, herbage, honest_cn, honesty, honour, honourable, hospital, hostler, hour, humble, humorous, humour_cn, humoursome
SURE	SURE_a	assurance_cn, assure_cn, centurion_cn, cesura_caesura_cn, durable, dure, during_cn, endure_cn, ensure_cn, futurity_cn, insurance_cn, insure_cn, mature_cn, maturity_cn, perdure/perdurable, sure_cn, surety, unsure
	SURE_b	actuary_sure, azure_sure, censure, century, closure, composure, fissure_sure, mortuary_sure, ordure_sure, seizure, suture_sure, tonsure, verdure
	SURE_c	duration, duress, induration, maturation, mensuration

<b>WHALE</b>	WHALE_a	whale, wharf, what, wheat, wheedle, wheel, wheeze, whelm, whelp, when, whence, where_cn, wherry, whet, whether, whey_cn, which, whiff, whiffle, whig, while, whim, whimper, whin, whine, whip, whirl, whisk, whisker_cn, whisper, whist, whistle, whit, white, whither, whitlow, whitsuntide, whiz, who_cn, whole, whom, whoop, whore_cn, whose_cn, why
	WHALE_b	elsewhere, nowhere, overwhelm, somewhat, somewhere

#### Appendix II. Example words excluded from Well's lexical sets (alphabetic order by set)

SET	SUBSET	EXAMPLE WORD
<b>BATH</b>	BATH_a	giraffe, Shaftesbury
	BATH_b	commando, Flanders, France, Frances, Francis, ranch, Sandra
	BATH_c	can't, corral, Iran, Iraq, morale, shan't, Slav, Sudan
	BATH_f	Basque, Cleopatra, contralto, Glasgow, graph, intransigent, masturbate, plaque, stance, transept
<b>CHOICE</b>	CHOICE_a	--
	CHOICE_b	--
	CHOICE_c	--
<b>CLOTH</b>	CLOTH_a	Austen, Austin, Australia, Austria, doss, floss
	CLOTH_b	Boston, Gloucester, gong, joss, Ross
	CLOTH_c	Florida, horrify, Laurence_Lawrence, moribund, Norwich, Oregon, tomorrow, Warwick
<b>comma</b>	<i>comma</i>	amoeba_ameba, arena, balsa, Bertha, catalpa, Cinderella, dementia, neuralgia, panda_2, phobia, saga, visa_2, vodka
<b>CURE</b>	CURE_ai	dour, spoor
	CURE_aii	McClure
	CURE_b	Bourbon, bourse, gourmand, gourmet
	CURE_ci	hour, tourism, tourist
	CURE_cii	angostura, anthurium, bravura, Huron, Muriel, neural, neuron_neurone, sulfuric_sulphuric, tellurium, thurible, Truro, Ural, Uriel
<b>DRESS</b>	DRESS	fez, Leicester, rev, Thames
<b>FACE</b>	FACE_a	bouquet, fête
	FACE_b	aitch, beige, raid
	FACE_c	--
<b>FLEECE</b>	FLEECE_a	grebe, Keith, Peter, Sheila
	FLEECE_b	Aesop, anemic_anaemic, Caesar
	FLEECE_c	casino, chic, elite, prestige, ski, trio, unique, visa_1
<b>FOOT</b>	FOOT	shouldn't
<b>FORCE</b>	FORCE_a	chore, crore, galore
	FORCE_bi	Borneo
	FORCE_bii	--
	FORCE_c	angora, boron, Dora, euphoria, fedora, Gregorian, moratorium, moron, Nora_Norah, thorium, torus, Victoria_Victorian
<b>GOAT</b>	GOAT_a	don't, gauche, mauve

	GOAT_b	Owen
<b>GOOSE</b>	GOOSE_a	ghoul, Moog, schooner, smooch, tarboosh, Vancouver
	GOOSE_b	flu, sewage, sleuth
<b>happy</b>	happy_a	birdie, boogie, breathy, budgie, calorie, chilli, corgi, edgy, fluffy, fussy, hibachi, khaki_2, lassie, movie, Nazi_2, prairie_2, salami, sari_2, scampi, sortie, spaghetti, strategy, stymie, talkie, taxi
	happy_b	Chelsea, hockey, Swansea
<b>KIT</b>	KIT	Syria
<b>letter</b>	letter	indicator, liner, ogre, pallor, scorer, Tudor
<b>LOT</b>	LOT	bother, Tom, waffle
<b>MOUTH</b>	MOUTH	MacLeod
<b>NEAR</b>	NEAR_a	--
	NEAR_b	Deirdre
	NEAR_c	diphtheria, eerie, Madeira
	NEAR_f	Colosseum, Crimean, Galatea, Jacobean, Korea, Maria, Sophia, TeDeum
<b>NORTH</b>	NORTH_a	Thor
	NORTH_b	cavort, corm, Dorking, Morgan, Mormon, morph, morpheme, morphia, morphine, orchid, porn, quartz, Thorpe, torque, torso, Warsaw, York
	NORTH_c	aural, Laura, Taurus
<b>NURSE</b>	NURSE	berth, Byrne, Earp, erg, liqueur, masseur, twerp, Worthing
<b>PALM</b>	PALM_a	blah, bra, ma, pa
	PALM_b	Afrikaans, Armagh, Bach, Bahai, baht, Botswana, Brahmin, Brahms, candelabra, couvade, Dada, Dali, façade, guano, Guatemala, guava, ha-ha, iguana, incommunicado, Java, Kahn, Karachi, kava, kraal, laager, lager, legato, llama, Lusaka, mafia, Mahal, maharajah_rajah, maharani_rani, Mahdi, Malawi, Mali, marijuana, Mikado, pizzicato, Pooh-Bah, raj, roulade, salaam, schwa, Shah, Somalia, staccato, Sumatra, swami, Swazi, taj_1, Taj_2, Transvaal, Yokohama, Zhivago
	PALM_f	aubade, bah, Bali, chorale, Colorado, enchilada, finale, Ghana, khaki_1, khan, Koran, lava, locale, Nazi_1, Nevada, nirvana, Pakistan, palaver, panorama, pasha, piranha, plaza, pyjama_pajama, Shan, soprano
<b>PRICE</b>	PRICE_a	bicycle, chi, Christ, Cyprus, eider, Glynde, hi-fi, hybrid, kaleidoscope, tried
	PRICE_b	--
<b>SQUARE</b>	SQUARE_a	Ayr, Eyre
	SQUARE_b	--
	SQUARE_c	aquarium, Dun Laoghaire, Eire, Mary, Pharaoh, prairie_1
<b>START</b>	START_a	bazaar, Saar
	START_b	aardvark
	START_c	aria, Bari, cascara, curare, Mata Hari, safari, Sahara, sari_1, scenario
<b>STRUT</b>	STRUT	Guthrie, mustn't
<b>THOUGHT</b>	THOUGHT_a	auk, Maugham, Paul, Raleigh, taut, Vaughan, Waugh

	THOUGHT_b	--
TRAP	TRAP	jazz, math_maths, panda_1

### Appendix III. IPA transcriptions for Buchanan's (1757) and Walker's (1791) notation systems

Buchanan 1757	IPA		Walker 1791	IPA
ā	/e:/		a1	/e:/
ǣ	/æ/		a2	/a:/
ai	/e:/		a3	/ɔ:/, /ɒ:/
au/aw	/ɔ:/		a4	/æ/
oi	/ai/ <sup>14</sup>		e1	/i:/
ē	/i:/		e2	/ɛ/
ee	/i/		i1	/ai/
ě	/ɛ/		i2	/ɪ/
ī	/ai/		o1	/o:/
ï	/i/		o2	/u:/
ō/oa	/o:/		o3	/ɔ:/
ö	/ɒ/		o4	/ɒ/
oo	/u/		u1	/ju:/
ou	/ɔu/		u2	/ʌ/
ū	/ju:/		u3	/ʊ/
Û	/ʌ/		o3i2	/ɔi/
			o3u3	/aʊ/
			o2u2	/uə/

<sup>14</sup> *oi* and *oy* have a mixed sound which is never varied, and sounds like long (i) (Buchanan, 1757: 11).