



This is a repository copy of *Towards a corpus of eighteenth-century English phonology*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/95378/>

Version: Submitted Version

---

**Book Section:**

Beal, J.C. and Sen, R. (2014) *Towards a corpus of eighteenth-century English phonology*. In: Vandelanotte, L., Davidse, K., Gentens, C. and Kimps, D., (eds.) *Recent Advances in Corpus Linguistics: Developing and Exploiting Corpora*. Brill/Rodopi, Amsterdam, pp. 31-54. ISBN 9789042038714

---

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**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/>

Beal, Joan & Ranjan Sen (in press, 2014). 'Towards a corpus of eighteenth-century English phonology', in Kristin Davidse, Caroline Gentens, Ditte Kimps & Lieven Vandelanotte (eds.), *Recent Advances in Corpus Linguistics: Developing and Exploiting Corpora*. In the series *Language and Computers: Studies in Practical Linguistics* (series editors: Christian Mair, Charles F. Meyer, Nelleke Oostdijk).

# Towards a Corpus of Eighteenth-Century English Phonology

---

*Joan C. Beal*

*Ranjan Sen*

*University of Sheffield*

## **Abstract**

This paper gives an account of plans for constructing a searchable database of eighteenth-century English phonology, an area which has hitherto received little attention from corpus linguistics. The project draws on a sample of eighteenth-century primary sources to construct a searchable database which will eventually provide visualisations of the distribution of phonological variants in time, space and social class.

The project incorporates data from pronouncing dictionaries published in the second half of the 18th century, recoded in the form of SAMPA transcriptions of as many of the approximately 1,700 words used to exemplify John Wells' Standard Lexical Sets as appear in the pronouncing dictionaries chosen, together with supplementary sets chosen to represent consonantal variants such as /hw/~w/ in WHICH, etc. The use of these sets and their associated keywords is standard practice in studies of variation and change in English, and including the full range of example words allows for differences in lexical distribution between the dictionaries, and between these and the contemporary accents described by Wells. Although all these dictionaries purported to describe the 'best' English, they were compiled by authors from different parts of the English-speaking world (mainly different regions of England, Scotland and Ireland but including some from North America) and so can provide evidence for geographical diffusion of innovations. (Beal 1999, Jones 2006).

The entries will be tagged according to the main lexical set to which they belong. Thus, a researcher interested in the distribution of words in Wells's (1982) PRICE and CHOICE sets will be able to find how each of the example words from these sets was transcribed in each of the 18th-century pronouncing dictionaries included in the database. There will also be links to descriptive and prescriptive comments included in the primary sources. The database will also include metadata providing background information on the

dictionaries, such as place of publication, birthplace, occupation and social class of author, and bibliographical references to published work referring to these dictionaries.

This paper provides an account of the design of this database and presents the results of a pilot study demonstrating how such a database can be used to answer questions concerning the chronological, social, geographical and phonological distribution of variation between /hw/ ~/w/ ~ /h/ in WHICH, WHO, NOWHERE, etc. which is of interest to sociolinguists, dialectologists and historical phonologists.

## 1 Introduction

The ‘corpus revolution’ has transformed the study of English historical linguistics, but, until relatively recently, historical corpora of English have tended to be compiled from Middle and Early Modern English materials, leaving the eighteenth and nineteenth centuries as the ‘Cinderellas of English historical linguistic study’ (Jones 1989: 272). Describing the then newly-compiled *Corpus of Late Modern English Texts* (CLMET), de Smet makes the following comment:

Symptomatic of a certain neglect of anything beyond the 17<sup>th</sup> century is the fact that the *Helsinki Corpus*, until now the most important electronic corpus for the study of the history of English, takes its final cut-off point in 1710. (de Smet 2005: 69).

Although the *Helsinki Corpus* (Rissanen et al. 1991) does indeed stop at 1710, there are now several corpora of English texts from the eighteenth and/ or nineteenth centuries. The *Penn Parsed Corpus of Modern British English* (PPCMBE), released in 2010, takes up where the *Helsinki Corpus* left off and covers the period 1700-1914, whilst the *Corpus of Historical American English* (COHA) includes nineteenth-century American English texts and the *Corpus of Oz Early English* (COOEE) (Fritz 2007) is compiled from English texts written in Australia, New Zealand and Norfolk Island between 1788 and 1900. ARCHER covers the period 1650-1990 and includes material from nine genres and both British and American English. Other Corpora, such as the *Corpus of Early English Correspondence Extension* (CEECE), the *Network of Eighteenth-century English Texts* (NEET) (Fitzmaurice 2007) and the Corpus of late Eighteenth-century Prose, concentrate on letters, whilst the *Old Bailey Corpus* has been compiled from the court documents originally digitised for the *Old Bailey Online* project. In addition to these corpora, scholars can now access electronic databases such as *Eighteenth-Century Collections Online* (ECCO), the *Eighteenth-Century English Grammars* (ECEG) database, the Chadwyck-Healey databases of eighteenth- and nineteenth-century fiction and drama and various databases of eighteenth- and nineteenth-century newspapers and periodicals.

It has been pointed out elsewhere (Beal 2012a) that the increasing availability of corpora compiled from texts of this period has revolutionised the study of Late Modern English in the twenty-first century. Denison notes that, ‘in the last two centuries, syntactic change has more often been statistical in nature, with a given construction... either

becoming more or less common generally or in particular registers' (1998: 93). Since statistically-based studies require large amounts of comparable data, it is not surprising that Late Modern English scholarship has followed in the wake of Late Modern English corpora. The first decade of this century has seen the publication of three monographs dealing with the whole of this period (Beal 2004, Jones 2006, Tieken-Boon van Ostade 2009), as well as volumes dedicated to the eighteenth (Görlach 2001, Hickey (ed.) 2010) and nineteenth centuries (Smutterberg 2005, Kytö et al 2006). Furthermore, a series of conferences on Late Modern English, which began in Edinburgh in 2001, will have its fifth meeting in Bergamo in 2013.

Whilst the above discussion seems to indicate that Late Modern English scholarship is in a healthy state, it has been argued (Beal 2012a) that phonology has been the poor relation in the Late Modern English family, largely due to the readier availability of corpora for the study of syntax and pragmatics. Although two monographs on Late Modern English pronunciation have been published (Beal 1999, Jones 2006), papers dealing with phonology have been in the minority in all the Late Modern English conferences held to date (see Beal 2012a: 22 for an analysis of the contents of publications from these conferences). The tendency for electronic corpora to be more useful for research in areas such as syntax and pragmatics is not confined to historical corpora. Anderson and Corbett point out that 'most accessible online corpora focus on the printed word, even if occasionally these words have been annotated to show their pronunciation' (2009: 124). Nevertheless, several corpora of twentieth-century English pronunciation are now available, including the *Diachronic Corpus of Tyneside English* (DECTE), the *Phonologie d'Anglais Contemporain* (PAC) corpus, the *Scottish Corpus of Texts and Speech* (SCOTS) and the *Intonational Variation in English* (IViE) corpus, all of which allow the user to search sound files. Of course, sound files of eighteenth- and most nineteenth-century speech are simply not available, so a corpus of historical English phonology would have to be based on printed information. In the next section, we outline the nature of the evidence available for eighteenth-century English phonology and discuss its usefulness and suitability for corpus construction.

## **2 Evidence for eighteenth-century English phonology**

Evidence for the pronunciation of English (or any other language) in historical periods preceding the invention of sound recording can be divided into two major categories: direct and indirect evidence (Beal 2012b, 63-4). Direct evidence consists of metalinguistic comments and linguistic descriptions from grammarians, lexicographers, orthoëpists and others who are overtly and intentionally providing this information, whilst indirect evidence is pieced together from clues provided in rhymes, puns and spellings by authors who were almost certainly unaware that they were leaving phonological information for future historical linguists. Thus, as Beal points out:

Shakespeare rhymed *war* with *jar* and *warm* with *harm* in *Venus and Adonis* (ll. 98/100 and 193/ 195 respectively) because he was writing within a tradition which demanded end-rhymes and because those words fitted in with the theme of his poem, not because he wished to record for posterity the fact that /w/ had not yet exerted a rounding influence on the following /a/. (1999: 37).

Shakespeare's rhymes thus provide indirect evidence for the unrounded pronunciation, and have been used as such by scholars such as Wyld (1923) and Kökeritz (1953). However, when the orthoëpist Christopher Cooper (1687) provides a separate notation <α> for the vowel in *war*, *warden* and *warm* in a volume whose title page declares that it is 'fitted for the Use of Schools and necessary for all those that desire to Read, Write or Speak our Tongue with Ease and Understanding' (1687: 1) he is deliberately providing this information for his contemporaries and later generations of phonologists can deduce from this that the rounding had taken place by this date in the variety described by Cooper.

The balance of direct and indirect evidence for historical English pronunciation shifts from the Old and Middle English periods, from which direct evidence is very scarce, through the Early Modern (c. 1500-1700) period when, as we can see from the examples above, both kinds of evidence are plentiful, to the Late Modern period, when direct evidence predominates. Standardisation of spelling, increasing literacy and a greater acceptance of eye-rhymes in poetry meant that indirect evidence from this period became scarcer, whilst an increasing awareness of the social value of a 'correct' pronunciation created a market for pronouncing dictionaries and elocution manuals, especially in the second half of the eighteenth century. In his monograph on the eighteenth-century elocutionist Thomas Sheridan, Benzie notes that '[F]ive times as many works on elocution were published between 1760 and 1800 than prior to 1760' (1972: 52). Dobson, whose major work on historical English pronunciation deals with the Early Modern period, dismissed eighteenth-century sources of direct evidence in the following sweeping statement:

The eighteenth century produced no writers to compare either 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)

However, as pointed out by Beal (1999: 47), Dobson was writing at a time when 'the prevailing attitude was... that the study of English philology stopped at 1700' and the ease of access we now have to eighteenth-century texts via *ECCO* was unthinkable. Beal (1999) and Jones (2006) have since made extensive use of eighteenth-century sources to provide detailed accounts of the phonology of this period. Although the purpose of eighteenth-century elocutionists such as John Walker and Thomas Sheridan was undoubtedly prescriptive, Beal and Jones both demonstrate that their work can be taken seriously as providing evidence not only for what was considered the 'correct' pronunciation of their

day, but also for pronunciations that were stigmatised and to be avoided. Furthermore, pronouncing dictionaries such as Sheridan (1780), Walker (1791) and many others from this period, provide descriptions of the recommended pronunciation of every word in the lexicon and thus, as Beal points out ‘invaluable detailed evidence of lexical diffusion’ (1999: 68). As such, this evidence could be of use not only to historians of English, but to scholars in the fields of historical phonology more broadly and of language variation and change. However, as Beal (2007) has pointed out, there has been little use of eighteenth-century evidence by scholars researching the present-day diffusion of sound changes which began in that period. The provision of a searchable database of eighteenth-century phonology would greatly facilitate the use of the ‘past to explain the present’ by researchers who may be unfamiliar with the complexities of eighteenth-century phraseology and notation. In the next sections, we will discuss the problems posed by these sources and propose a solution.

### 3 Problems arising from eighteenth-century sources

#### 3.1 Problem 1: Annotation

One major obstacle encountered by scholars embarking on research into Late Modern English phonology is the diversity of systems used by eighteenth- (and nineteenth-) century authors to represent the distinct sounds of English. The ubiquity of the International Phonetic Alphabet (IPA) in the twentieth century has made scholars reluctant to decipher earlier systems such as A. J. Ellis’s Palaeotype (see Local 1983, Maguire 2012 for discussion of Ellis’s system). Eighteenth-century authors, like the orthoëpists of the sixteenth and seventeenth centuries, used a variety of methods to convey their recommended pronunciations to their readers. Abercrombie (1981) categorizes the orthographic systems used by these authors into two major schematic types: *new alphabets* and *augmentation of the Roman alphabet*, with the second type further subdivided into schemes using *diacritics* and *extended alphabets*. Although Thomas Spence described his system, illustrated in figure 1, as a ‘New Alphabet’, according to Abercrombie’s scheme this would be categorized as an extended alphabet, based as it is on modification of the letters of the Roman alphabet. Even this was too radical for the majority of eighteenth-century readers, who preferred diacritic systems such as that exemplified by Walker’s ‘Table of the simple and diphthongal vowels’ (figure 2). Here, the conventional spelling is not disrupted unless, as in words like *enough*, the pronunciation deviates considerably from that indicated by the usual values of the orthographic letters. In such cases, authors using diacritic systems would resort to semi-phonetic spelling: Walker represents this word as <e<sup>1</sup>-nu<sup>2</sup>f>. This combination of semi-phonetic spelling and superscripted numbers to indicate separate vowel phonemes was first used in a pronouncing dictionary by Kenrick (1773), though the system had been described by Sheridan in his (1761) *Dissertation on the Causes of the Difficulties which Occur in Learning the English Tongue*. Sheridan went on to use this system in his (1780) *General Dictionary of the English Language*, and its adoption by Walker ensured that this would be the most successful and widespread system

of the eighteenth and early nineteenth centuries. However, each author who uses this system has his own way of representing specific sounds, so that an <a> with superscripted <1> has a different phonetic value in Walker's dictionary, where it represents /e:/ as in *fate*, and in Sheridan's, where it represents /a/ as in *hat*.

INSERT FIGS 1 and 2 HERE

It should be apparent from the above discussion that the existence of such a variety of notation systems would prove an obstacle to the comparison of pronunciations recommended by different authors of the Late Modern period. The researcher must decipher each system and translate each combination of symbol and diacritic into IPA in order to make such comparisons. Those who have undertaken such projects (Beal 1999, Jones 2006, MacMahon 1998) have had to search each source manually to make these comparisons.

Beal (1999) created a searchable database of all the entries in Thomas Spence's *Grand Repository of the English Language* (1775) by recoding them from Spence's 'New Alphabet' into alphanumeric characters, as set out in figure 3. She then used the Oxford Concordance Programme (OCP) to generate lists of words containing specific symbols in specified environments, which, given the phonemic nature of Spence's system, provided all instances of a particular phoneme/ environment in Spence's lexicon. Each word in the list was then looked up in a number of other eighteenth-century pronouncing dictionaries, including Sheridan (1780) and Walker (1791) to yield evidence of variation and change, including lexical diffusion. Beal notes that this task was 'painstaking and time consuming', and, whilst it yielded a great deal of useful information, the study 'barely scratched the surface in terms of the wealth of phonological evidence available in eighteenth-century pronouncing dictionaries' (1999: 183-4).

INSERT FIGURE 3 HERE

The system set out in figure 3 was devised by Beal on an *ad hoc* basis. We propose to use an adaptation of the SAMPA system devised as 'the best robust international collaborative basis for a standard machine-readable encoding of phonetic notation' (<http://www.phon.ucl.ac.uk/home/sampa/>). For each eighteenth-century source, the combinations of character and diacritic denoting specific phonemes will be transliterated into the SAMPA-based equivalent. Figure 4 shows the notations used by Walker (1791), with their equivalents in SAMPA and IPA.

INSERT FIGURE 4 HERE

It is important to note that the notational equivalents in figure 4 are intended to be phonemic: the transliteration of Walker's <a<sup>2</sup>> as SAMPA <A> and IPA /ɑ:/ is not

intended to suggest that the vowel concerned had the same (back) articulation in Walker's time as in present-day RP, simply that it is a separate phoneme from Walker's <a<sup>4</sup>>. The creation of a database of eighteenth-century phonology will inevitably require us to make decisions concerning the attribution of notations in the historical sources to their equivalent SAMPA-based phonemic notations, but all such decisions will be accounted for in the metadata accompanying the database. In the next section, we will discuss the size of this proposed database.

### 3.2 Problem 2: Size and scope

Beal (1999) was able to transliterate the whole of Spence's (1775) dictionary because it is relatively short, consisting of approximately 17,000 entries. A pilot study carried out in 2010<sup>i</sup> established that a highly competent research assistant was able to transliterate 3,378 entries from Walker (1791) over 40 hours. This only covered the entries from *abacus* to *borage*, indicating that it would take a great deal of time and therefore expense to include every word from every eighteenth-century source in the proposed database. Whilst, as Beal (1999) has demonstrated, access to a complete lexicon does provide valuable evidence of lexical diffusion, many of the words recorded in eighteenth-century dictionaries are obscure and/ or now obsolete. Examples from the pilot project include *arundinacious*, *atrabilariousness*, *belswagger*, and *bezoardick*, all magnificent words but unlikely to be amongst those included in studies of English historical phonology (unless, of course, the object of the research was an investigation of stress patterns in polysyllabic words). In order to keep the database to a manageable size, we propose to restrict the entries to the words used by Wells (1982) to illustrate his standard lexical sets. In Wells's system, as illustrated in figure 5, each keyword 'stands for a large number of words which behave the same way in respect of the incidence of vowels in different accents' (1982: 120). Since, as shown in figure 6, this system also includes subsets which differentiate between historical lexical sets, it is as useful for diachronic as for diatopic comparisons.

INSERT FIGURES 5 and 6 HERE.

Including all the words provided by Wells in subsets of lexical sets would give 1,739 items, about one tenth of the size of Spence's (1775) dictionary, but we can see from the example of the FLEECE set in figure 6 here that not all of these would be included in eighteenth-century pronouncing dictionaries. Although some of these dictionaries did include proper names, *Keith*, and *Sheila* are unlikely to appear (though *Peter* may be in some); likewise *casino*, and *ski* are first cited in 1789 and 1755 respectively in the *Oxford English Dictionary*.<sup>1</sup> On the other hand, since Wells's lexical sets are designed for the comparison of

---

<sup>1</sup> The 1755 quotation for *ski* is highlighted in the OED as an 'isolated early use'. The next earliest quotation dates from 1885.

vowel phonemes and their distribution, further sets will need to be provided if users of the database are to have access to information concerning consonantal variants, such as /hw ~w/ discussed below. Subtracting from Wells's list such words as do not appear in the eighteenth-century sources and augmenting it with a small number of consonantal sets would yield a database of manageable size which would nevertheless provide a rich amount of information on the diachronic, diatopic and lexical distribution of phonological variants in eighteenth-century English. The next section consists of a case study in which a supplementary lexical set for /hw ~w/ was compared across a subset of nine eighteenth-century pronouncing dictionaries to reveal patterns of variation and change.

#### **4 A pilot study: The eighteenth-century pronunciation of 'wh'**

We describe plans above for a searchable database of eighteenth-century English phonology. In this section, we test whether such a resource might usefully answer questions about phonological variation and change.

Our test case involves the representation of 'wh' in nine eighteenth-century pronouncing dictionaries. In present-day standard southern British English, words such as *whale*, *what*, and *where* begin with /w/, whilst *who* and *whole* have initial /h/. Eighteenth-century sources present evidence, through their orthographic systems, of variation across authors between /hw/ and /w/ for the first set, hence a preserved versus unpreserved /hw ~w/ contrast in *where* ~ *wear*. The nine dictionaries were selected to ensure that variation in both pronunciation as well as in geography and chronology were amply represented. We recorded in a spreadsheet the pronunciations of 50 words which occur in as many as possible of the nine dictionaries, consisting of (1) 39 words beginning with the spelling 'wh' which are pronounced with /w/ in present-day southern British English, (2) 6 words with initial 'wh' which are now pronounced with initial /h/, and (3) 5 words with 'wh' word internally, which are now all pronounced with internal /w/. The nine authors were arranged as columns in chronological order, with an additional column displaying the total of how many times each word appears in quotations used by the OED dating from the eighteenth century (1701-1800), to give an indication of their frequency. The words under consideration were listed as rows. Figure 7 presents the evidence as described.

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 presents the loss of the /hw ~w/ contrast as a special case of 'h-dropping', which was just beginning to attract social stigma at this time in lower-class London English. The proposed database would include such information.

As our study aims to ascertain and explain the variation in pronunciation of ‘wh’ in the eighteenth century, it is first useful to present background research on two aspects: firstly, a reconstruction of the nature of the /hw ~ w/ contrast going into the eighteenth century, and secondly, the phonetic nature of the sound or cluster we are treating as /hw/.

#### **4.1 The starting point: /hw/ before the eighteenth century**

Words containing /hw/ in English (< Old English hw < Common Germanic \*xw < Proto-Indo-European \*kw) had already begun to be pronounced with simple /w/ in the twelfth century in many southern dialects, notably in London (Dobson 1957: 974). However, /hw/ was clearly not unknown in southern speech for many centuries, as shown by the fact that spellings with simple <w> are much sparser than would be expected in the fifteenth to seventeenth centuries if this was the regular pronunciation (Wyld 1936: 312). Johnston (1764: 9) comments that the ‘h’ element in these words was at the time ‘very little heard’, which appears to indicate, from the context, that these forms had weak aspiration in normal speech, and not that *few people* pronounced them as /hw/. Contrary to the southern position, most northern English and Scottish dialects robustly preserved /hw/, which persists to this day in Scottish dialects.

The development of \*/hw/ to /h/ before back, rounded vowels such as /u/ (e.g. *who*) seems to date from the thirteenth or fourteenth centuries, but only entered conservative registers in the seventeenth century (Dobson 1957: 980-81). The /h/ pronunciation was reasonably settled in southern England by the eighteenth century, but data from north-east England (Spence 1775) suggests that /hw/ persisted in these dialects for longer (see §4.3).

Therefore, entering the eighteenth century, the /hw ~ w/ contrast was only weakly realised in southern English, and the /hw ~ h/ contrast before back, rounded vowels no longer realised, whereas /hw ~ w/ was robustly preserved in northern English and Scottish dialects, and there is evidence to indicate that /hw ~ h/ also remained in some northern English varieties.

| Figure 7    | 1757     | 1772     | 1773    | 1775   | 1775  | 1780     | 1786 | 1791   | 1797  | 1701-1800 |
|-------------|----------|----------|---------|--------|-------|----------|------|--------|-------|-----------|
|             | BUCHANAN | JOHNSTON | KENRICK | SPENCE | PERRY | SHERIDAN | BURN | WALKER | JONES | OED COUNT |
| WHALE       | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 84        |
| WHARF       | hw       | w        | w       | w      | w     | hw       | w    | hw     | hw    | 15        |
| WHAT        | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 2611      |
| WHEAT       | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 157       |
| WHEEDLE     | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 2         |
| WHEEL       | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 190       |
| WHEEZE      | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 5         |
| WHELM       | hw       | hw       | hw      | hw     | hw    | hw       | w    | hw     | hw    | 0         |
| WHELP       | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 5         |
| WHEN        | NA       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 3742      |
| WHENCE      | NA       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 264       |
| WHERE       | NA       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 1900      |
| WHERRY      | hw       | hw       | hw      | hw     | w     | hw       | w    | hw     | hw    | 6         |
| WHET        | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 5         |
| WHETHER     | hw       | hw       | hw      | hw     | w     | hw       | w    | hw     | hw    | 629       |
| WHEY        | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 15        |
| WHICH       | NA       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 9201      |
| WHIFF       | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 8         |
| WHIFFLE     | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 0         |
| WHIG        | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 69        |
| WHILE       | NA       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 852       |
| WHIM        | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 18        |
| WHIMPER     | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 2         |
| WHIN        | hw       | NA       | w       | hw     | w     | hw       | w    | hw     | hw    | 13        |
| WHINE       | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 11        |
| WHIP        | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 57        |
| WHIRL       | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 29        |
| WHISK       | hw       | hw       | hw      | hw     | hw    | hw       | w    | hw     | hw    | 14        |
| WHISKERS    | hw       | hw       | hw      | hw     | hw    | hw       | w    | hw     | hw    | 20        |
| WHISPER     | hw       | hw       | hw      | hw     | hw    | hw       | w    | hw     | hw    | 29        |
| WHIST       | hw       | hw       | hw/w    | hw     | hw/w  | hw       | w    | hw     | hw    | 31        |
| WHISTLE     | hw       | hw       | hw      | hw     | w     | hw       | w    | hw     | hw    | 60        |
| WHIT        | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 16        |
| WHITE       | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 1621      |
| WHITHER     | hw       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 12        |
| WHITLOW     | hw       | hw       | hw      | hw     | hw    | hw       | hw   | hw     | hw    | 1         |
| WHITSUNTIDE | hw       | hw       | hw      | hw     | w     | hw       | hw   | hw     | hw    | 2         |
| WHIZ        | hw       | hw       | hw      | hw     | hw    | hw       | hw   | hw     | hw    | 4         |
| WHO         | hw       | h        | h       | hw     | h     | h        | h    | h      | h     | 4333      |
| WHOLE       | h        | h        | h       | hw     | h     | h        | h    | h      | h     | 1221      |
| WHOM        | NA       | NA       | h       | hw     | h     | h        | h    | h      | h     | 475       |
| WHOOOP      | hw/w     | h        | h       | hw     | h     | h        | h    | h      | h     | 11        |
| WHORE       | h        | h        | h       | h      | h     | h        | h    | h      | h     | 9         |
| WHOSE       | NA       | h        | h       | hw     | h     | h        | h    | h      | h     | 986       |
| WHY         | NA       | hw       | w       | hw     | w     | hw       | w    | hw     | hw    | 392       |
| SOMEWHERE   | NA       | hw       | NA      | hw     | w     | hw       |      | hw     | hw    | 49        |
| SOMEWHAT    | NA       | NA       | NA      | hw     | w     | hw       |      | hw     | hw    | 484       |
| OVERWHELM   | hw       | hw       | NA      | hw     | hw    | hw       |      | hw     | hw    | 5         |
| NOWHERE     | NA       | NA       | NA      | hw     | w     | hw       |      | hw     | hw    | 39        |
| ELSEWHERE   | NA       | NA       | NA      | hw     | hw    | hw       |      | hw     | hw    | 140       |

## 4.2 The phonetics of /hw/

The phonological nature of ‘wh’ (when not simply /w/) in eighteenth-century English is unclear: it could either be analysed as a consonant cluster /hw/ or a single voiceless labial-velar fricative or approximant /ɸ/. The question is discussed further in §4.3 and §4.6, but generally the notation /hw/ is employed throughout this pilot study. In either case, there are labial and velar place-of-articulation elements, and a breathy, aspirated element which may or may not produce audible frication, hence the phonetic realisation of the phonologically different representations need not be different. Ladefoged & Maddieson (1996: 326) report that, in the world’s languages, /ɸ/ is usually non-fricative, but go on to acknowledge that a fricative realisation is a possibility, in which case it is ‘better described as a voiceless labialized velar fricative’, namely [x<sup>w</sup>], as ‘the voiceless counterpart of w cannot have friction at both the labial and velar places of articulation’. Reconstructing such a phonetic realisation of the phoneme or cluster would be consistent with one historical observation and one piece of contemporary commentary. Historically, greater restrictions seem to have developed between /hw/ and the labiality of the *following* vowel rather than the preceding, an asymmetry which would be predicted by a labializing secondary articulation, as Ladefoged & Maddieson (1996: 357) note that a labial articulatory gesture is anchored in terms of its timing to the release rather than the formation of a primary articulation. Hence, we have \*/hw/ > /h/ in words such as *who*, where a vowel with lip-rounding follows. Secondly, Douglas ([1779] 1991: 141) comments that ‘The Scottish pronounce the *wh* like their guttural *ch*’, and ‘When they endeavour to correct this fault they are apt to omit the *h*, so as to pronounce *whit* and *wit*.. in the very same manner’. Given that the ‘fault’ appears to be the production of velar frication, it would be predicted that the omission of such an element would result in precisely a labial-velar approximant, thus /hw/*it* > /w/*it*. However, Douglas’ comment does suggest that the difference between *whit* and *wit* ought then to be preserved in a different way from the production of strong velar frication, suggesting two alternatives: weak velar frication or labial frication.

A logical alternative realisation of /hw/ or /ɸ/ would be a voiceless velarized labial fricative [ɸ<sup>v</sup>], i.e. frication produced at the lips, not the velum, and this is the pronunciation concluded by M. Jones (2008) for the present-day Scottish variant, based on acoustic and articulatory evidence. Jones identifies weak frication produced at the lips, as suggested by the generally flat acoustic spectrum, and the fact that lip aperture is different between /w/ and /ɸ/, indicating a different modification of the airflow at the lips between the two sounds.<sup>2</sup> The acoustic formats are generally those for a labial-velar, indicating that there is an approximant-degree constriction at the velum in addition to the labial frication. Labial frication historically in Scotland appears to be corroborated by the development of \*/hw/ to

---

<sup>2</sup> M. Jones (2008) finds that /ɸ/ has a *wider* lip aperture than /w/, which would be opposite relationship expected between approximant /w/ and fricative /ɸ/. A possible explanation is that the lip aperture used to execute Scottish /w/ is too small to produce audible labial frication.

/f/ in north-east Scottish dialects, a change that is only plausible if we posit that the velar element of the sound was lost due to its minimal acoustic effect.

We should therefore bear in mind when analysing the eighteenth-century data that there may be geographical variation in the phonetic realisation of the sound represented by ‘wh’.

### 4.3 Analysing the data 1: Geography

Three main patterns emerge from the data based on geographical distribution: (1) The London authors prefer /hw/ to /w/, with the exception of Kenrick, (2) Two out of the three Scottish authors prefer /w/ (Perry, Burn), whereas the earliest, Buchanan (1775) prefers /hw/, and (3) Spence, from Newcastle, has /hw/ even in words containing a following back, rounded vowel, e.g. *who*; all the other authors have /h/ in this position with exceptions discussed in §4.6.1 below. Sheridan, from Ireland, has /hw/ consistently, and /h/ where expected, but he patterns in many ways with the London-based authors, as discussed below.

The London authors’ preference for /hw/ appears to be inconsistent with the conclusion in §4.1 above that /hw/ was only weakly realised in southern English at the start of the eighteenth century. We might have expected the /hw ~w/ contrast to be further eroded and not identified by the dictionary authors in the second half of the century, and indeed most indirect sources of evidence suggest that /w/ was the norm in London at this time. However, the near-consistent transcription using /hw/ appears to reflect a prescriptive attempt to revitalise this contrast under the influence of spelling, on the basis that the simplification of /hw/ to /w/ could be considered a form of ‘h-dropping’, that is, the common omission of the initial fricative in words such as *happy* in the lower-class London English pronunciation of the time (Beal 1999: 176-78). This phenomenon was beginning to attract social stigma in the middle of the eighteenth century, and was one precisely proscribed by London-based authors such as Walker and Jones, who both explicitly classed the /w/ pronunciation as ‘h-dropping’, and made overt comments labelling this practice a vulgarity. The Irishman Sheridan, who spent a number of years in London, also specifically proscribed the /w/ pronunciation, aligning it with the stigmatized ‘h-dropping’. This account of the London pattern is supported by the fact that the only London-based exception to /hw/ is Kenrick (1773), one of the earliest of the group; presumably the stigmatization of /w/ had not yet fully taken effect by this time. The otherwise consistent /hw/ pronunciation in the dictionaries instead of the apparently regular London /w/ can therefore be considered ‘collateral damage’ from the stigma of ‘h-dropping’.

The preference for /w/ among the Scottish authors is again curious in the light of the conclusions in §4.1 and §4.2 above that /hw/ remained the regular Scottish pronunciation throughout this period and indeed to the present day. Recall, however, Douglas’ ([1779] 1991: 141)) comments that ‘The Scottish pronounce the *wh* like their guttural *ch*’, and ‘When they endeavour to correct this fault they are apt to omit the *h*’. This observation is consistent with Perry’s and Burn’s near-consistent /w/ (the exceptions are discussed below in §4.5.2, §4.6.3, and §4.6.4), if we posit that these authors were advising a more London-

like pronunciation to avoid the Scottish /hw/, stigmatized due to its clear regional connotations. 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. Finally, and similarly to the London pattern, the exception to the Scottish pattern is the earliest from that geographical area, Buchanan (1757), who has near-consistent /hw/, presumably reflecting the standard Scottish pronunciation of the time and up to the present day.

Spence (1775) from Newcastle in north-east England has extremely consistent /hw/, even in words containing a following back, rounded vowel (the only exceptions are *wharf* and *whore*, discussed below in §4.6). One can reasonably assume that Spence’s transcription accurately reflects the regional pronunciation of the time (Beal 1999: 179-80). Furthermore, Spence is the only author to use a special symbol for the /hw/ sound: an upper-case WH ligature. This could indicate that in Spence’s dialect, the sound was not a consonant cluster /hw/, but a single phoneme /ʍ/, and this account is further supported by the preservation of the labial element even in words such as *who*, *whole*, *whom*, *whoop* and *whose*, all of which Spence transcribes using his ligature. The preservation of the labial element in a single phoneme would then be entirely in parallel with its preservation in the voiced labial-velar phoneme /w/ before a back, rounded vowel in almost all English dialects, including those of north-east England, as in *wound*, *womb*, *wool*, *wood*, etc. Undoubtedly a single phoneme, /w/ did not delete or develop to a velar or glottal sound in this environment. Also analysing /ʍ/ as a single phoneme in Spence’s Newcastle dialect explains this identical behaviour, and implies that in most other English dialects – which have /h/ before back, rounded vowels – the correct phonological analysis of the ‘wh’ sound is a cluster /hw/. Delabialisation could then be analysed as cluster simplification, with loss of the /w/ element before a back, rounded vowel, in parallel with the widespread loss of /w/ in other clusters in this exact environment, e.g. *sword*, where \*/sw/ > /s/ several centuries earlier.<sup>3</sup> Recall from §4.1 that \*/hw/ > /w/ in this environment began in the thirteenth to fourteenth centuries.

To summarise, the collection of evidence from several eighteenth-century dictionaries, both from pronunciations and commentaries, allows us to reconstruct and explain the geographical distribution of the variant realisation of the ‘wh’ sound, from London to Newcastle in England, through to Scotland.

#### 4.4 Analysing the data 2: Chronology

The chronological patterns for the distribution of ‘wh’ have been identified above: Kenrick, a London author from an earlier period of the time-frame, has a /w/ pronunciation because

---

<sup>3</sup> The cluster analysis is in line with that of Vachek (1954), who presents other arguments regarding sonorant-cluster simplification in its favour, and Hickey (1984, 2007: 319-320), who presents arguments from sonority.

the stigma of ‘h-dropping’ and the identification of /w/ with this phenomenon was only recently taking hold in educated London speech. Buchanan, the earliest Scottish author in the pilot study, has a /hw/ pronunciation, unlike the later Scottish authors, who are presumably recommending a more London-like, and less ‘guttural’, pronunciation, unaware that trends in London were simultaneously changing. In future studies using a more complete database, the chronological dimension of the problem under investigation will be easily visualised and notable patterns more readily extracted.

## 4.5 Analysing the data 3: Lexical factors

### 4.5.1 Homophones

Certain authors make use of the /hw ~ w/ contrast to differentiate the two meanings of the words *whist* and *whoop*, constructing minimal pairs from the contrast. Kenrick and Perry both have /hw/ for *whist* ‘be quiet’, but /w/ for *whist* ‘card game’. It is perhaps not coincidental that these are the two authors (in the pilot study) who are most variable in their selection of /hw/ or /w/: both generally prefer /w/, but a number of unexpected /hw/ words appear, discussed further in §4.5.2 and §4.6.3 below. Both were arguably sensitive to the contrast, and presumably believed there to be differences between lexical items as to which was the correct sound (i.e. not an ‘across-the-board’ /hw/ or /w/). Similarly, Buchanan has /hw/ for *whoop* ‘a cry’, but /w/ for *whoop* ‘a bird’. Buchanan elsewhere consistently has /hw/, suggesting that the contrast is being used solely to differentiate between words that would otherwise be homophones.

A near-minimal pair can also be identified in Burn’s and Perry’s pronunciations for *Whitsuntide* with /hw/ and *whit* with /w/. Note that both authors were Scottish, but both generally preferred /w/, presumably in conflict with the regular pronunciations of their regions. The construction of a minimal pair therefore suggests sensitivity to the contrast.

### 4.5.2 Onomatopoeia and sound symbolism

The /hw/ pronunciation for *whist* ‘be quiet’ and *whoop* ‘a cry’ reported above could also be analysed as examples of onomatopoeic pronunciation. Further plausible examples of onomatopoeia or sound symbolism again come from Kenrick and Perry:

- (1) Kenrick’s /hw/ pronunciations analysable as onomatopoeia/symbolism

*whisk, whisper, whistle, whiz*

- (2) Perry’s /hw/ pronunciations analysable as onomatopoeia/symbolism

*whisk, whisper*

It is perhaps also not a coincidence that these words all have a relatively high, front vowel following the /hw/ sound, as such a tongue position might plausibly enhance velar frication (see §4.6.3).

### 4.5.3 Word frequency

In this pilot study, we attempted to gauge the frequency of the words under consideration to ascertain whether any pattern emerged. Frequency has been shown to play an important role in the spread of sound changes, in the lexical diffusion model (e.g. Philips 1984). If our data enables us to catch a change mid-stream, whether that change is /hw/ > /w/ by regular sound change, or /w/ > /hw/ due to prescriptive or social factors, word frequency might show us how far the change has progressed, if the change is lexically diffused. If the change is not of this type, we might expect frequency to play a minimal role. Incorporating word frequency information in any database would therefore provide a valuable tool to test theories of sound change.

We compiled a word count of all the forms under consideration in quotations used by the *Oxford English Dictionary* during the time-period 1701-1800, to give a rough indication of the frequencies of the words at the time. Few patterns emerge from the word count, other than that Kenrick's and Perry's unexpected /hw/ words which are not onomatopoeic have very low counts: *whelm* has zero and *wherry* only six. Compare these figures with words containing /w/ in these dictionaries: *whale* has a count of 84, and *whig* 69. However, the pattern is not wholly corroborated by the data, in that other apparently low-frequency items have the /w/ pronunciation, such as *wheedle* (two) and *wheeze* (five). It is possible that a more complete word count will provide clearer patterns or a clearer indication of their absence, to enable us to ascertain whether frequency played a role. We therefore intend to use ARCHER to extend the word-count information for the final database.

## 4.6 Analysing the data 4: Phonology

### 4.6.1 Before high, rounded vowels

The clearest phonological environment conditioning the distribution of 'wh' variants is of course before a high, rounded vowel. From the forms in the data, it appears that any vowel that is higher and more round than /ɔ/ on the vowel quadrilateral results in a realisation of the sounds written 'wh' as /h/, and not /hw ~ w/, and this is presumably due to a labial co-occurrence constraint. In §4.3, we concluded that this might best be analysed as deletion of /w/ in a /hw/ cluster, as dialects where the sound was arguably a phonological singleton (Spence's Newcastle dialect) resisted the deletion. The conclusion that the vowel had to be higher than /ɔ/ arises from the absence of /h/ realisations in *wharf* in any of the dictionaries.

*Whore* consistently has a pronunciation with /h/ in every one of the nine dictionaries in the pilot study, despite (1) possibly having a vowel /ɔ/, although OED reports that the eighteenth-century pronunciation was (the now dialectal) /huə/, and (2) Spence having no /h/ realisations, except in this instance. The solution appears to be historical: *whore* etymologically has initial /h/ (Old English *hore*), not /hw/, hence the pronunciation variant with /hw/ arguably never arose.

A second instance where etymology provides a solution is Buchanan's choice of /hw/ for *who* (like Spence), but /h/ for *whole* (unlike Spence). Not only does *whole*

etymologically have /h/ (Old English *hāl*), furthermore the standard spellings of *whole* in Scots were *hail* and *hale*, hence presumably the Scottish Buchanan considered the /h/ pronunciation the robust standard. Buchanan's other interesting choice in this phonological environment is in the *whoop* minimal pair with /hw/ and /w/ discussed in §4.5 above.

#### 4.6.2 *Wharf*

The initial sound in *wharf* follows whichever practice is generally adopted by seven of the nine authors (e.g. /hw/ in Walker, /w/ in Burn), but both Spence and Johnston rather curiously have /w/ instead of the usual /hw/ given by these two authors. The occurrence here of /w/ in Spence is particularly striking, given that there are no other such pronunciations of 'wh' in his dictionary, and he has /hw/ even in words with a following high, rounded vowel (*whore* being the only exception, as above). How can this peculiarity be explained?

Clearly, the rounding of the vowel is not a relevant factor, as a labial dissimilation would presumably have yielded /h/, not /w/, and it is clear that such phenomenon does not take place in Spence's dialect (§4.3). We must discover an explanation that accounts for the loss of the *fricative* element. We identified in §4.2 that the alternatives for the phonetic realisation of /hw/ or fricated /w/ were (1) a labialized velar fricative [x<sup>w</sup>], or (2) a velarized labial fricative [ϕ<sup>Y</sup>]. In account (1), frication is produced by a high tongue back, so presumably a following vowel also requiring a high tongue back (a high, back vowel) might compromise the perceptibility of that frication as a consequence of vowel anticipation (i.e. the listener factors out the acoustic effect of the high tongue back as simply the result of the following vowel and not intended in the consonant). This might result in the loss of the velar fricative element, yielding simply a labial-velar approximant /w/. An insurmountable problem arises from this account: it does not explain the occurrence of /hw/ in all the other forms in Spence where there is a following high, back vowel (*who*, *whose*, *whom*, *whole*, *whoop*), and also does not explain why these forms have the /h/ pronunciation (with labial dissimilation) in Johnston. The alternative, account (2) with a velarized labial fricative, performs much better. If /hw/ had labial and not velar frication in the varieties of Spence and Johnston, the loss of the initial fricative element in this word, but not the others, can easily be accounted for by positing a non-adjacent dissimilation in labial frication due to the final labiodental fricative /f/: /hw...f/ > /w...f/. Note that it is frication, not labiality that is dissimilating on this account, thus yielding a correct prediction. Of course, this analysis opens up a further area of phonetic/phonological investigation as to the variation in the phonetic realisation of /hw/. Why, for example, should Spence in Newcastle and Johnston in London pattern together?

#### 4.6.3 Unexpected /hw/ words in Kenrick and Perry

As noted in §4.5 above, both Kenrick and Perry unexpectedly have words with /hw/, in contrast to the usual /w/ in these authors. Identifying these words allow us to contemplate a phonological pattern.

(3) Kenrick's unexpected /hw/ words

- a. *whelm, wherry, whether*
- b. *whisk, whiskers, whisper, whistle*
- c. *whitlow, Whitsuntide, whiz*

(4) Perry's unexpected /hw/ words

- a. *whelm*
- b. *whisk, whiskers, whisper* (but NOT in *whistle*)

The clearest patterns are firstly that the following vowel is always front /*ɛ*/ or /*ɪ*/ (possibly enhancing velar frication with a raised tongue front), and secondly that there is a group of forms with /*s*/ following the vowel, possibly with a tighter restriction for Perry that the /*s*/ has to be in the same syllable (thus /*hwisk*/ *whisk*, but /*wɪ.səl*/ *whistle*). No clear phonological explanations are forthcoming (e.g. /*hw*/ does not occur before all front vowels, or in all forms with /*ɛ*, *ɪ*/), but these patterns could form the starting-point for a more detailed investigation. The non-phonological explanation of onomatopoeia/symbolism for a number of these words is posited in §4.5.2, but this cannot account for all of the forms.

#### 4.6.4 Stress and compounds

The final five words in the data list comprise forms in which 'wh' occurs word-internally, not initially: *somewhere*, *somewhat*, *overwhelm*, *nowhere*, and *elsewhere*. Not all of the authors list all of the words, but those that have initial /*hw*/ (Johnston, Spence, Sheridan, Walker, and Jones) also consistently have internal /*hw*/. However, Perry, the only author listing these words who usually has /*w*/, has pronunciations with /*w*/ for *somewhere*, *somewhat* and *nowhere*, /*hw*/ for *overwhelm* and *elsewhere*. Recall that Perry has some unexpected /*hw*/ words, which show some indistinct patterns (§4.6.3). However, in this instance, the pattern is unambiguous: /*hw*/ is licensed internally in the onset of a primarily stressed syllable, whereas /*w*/ occurs as the unstressed-syllable onset variant. The stress markers are clearly indicated by Perry himself to corroborate this pattern.

(5) Stress (acute accent) and the distribution of word-internal /*hw* ~ *w*/ in Perry

- a. Onset of internal stressed syllable: /*hw*  
*overwhélm, elsewhére*
- b. Onset of internal unstressed syllable: /*w*/  
*sómewhere, sómewhat, nówhere*

This distribution is exactly paralleled by the distribution of internal /*h*/ in present-day English: /*h*/ is absent in the onset of an unstressed syllable (*véhicle* with no /*h*/ pronounced), but present in the onset of a stressed one (*vehícular*).

## 4.7 Conclusions from the pilot study

We have 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 incorporating other types of evidence into our analyses (phonetics, phonology, etymology, typology, frequency), we have been able to posit accounts for many of the patterns which such an orderly approach to the data has allowed us to ascertain. Some of the analyses are more robust than others, but all of them at the very least present a starting-point for further investigation. This could progress in two ways: either by further examining the phonetics, phonology, etc. of the problem, or by incorporating much more contemporary data in order to identify much more robust patterns and potential sources of explanation. It is hoped that a database of sources of eighteenth-century phonology will allow us to develop this latter approach so as to place the former on a much firmer footing.

## References

- Abercrombie, David (1981). 'Extending the Roman alphabet: some orthographic experiments of the past four centuries' in Asher, R. E. and Henderson, E. (eds.) *Towards a History of Phonetics*. Edinburgh: Edinburgh University Press.
- Anderson, Wendy and Corbett, John (2009). *Exploring English with Online Corpora*. Basingstoke: Palgrave Macmillan.
- Beal, Joan C. (1999). *English Pronunciation in the Eighteenth Century: Thomas Spence's 'Grand Repository of the English Language' (1775)*. Oxford: Clarendon Press.
- Beal, Joan C. (2004). *English in Modern Times 1700-1945*. London: Arnold.
- Beal, Joan C. (2007). "To Explain the present: 18<sup>th</sup> and 19<sup>th</sup>-century antecedents of 21<sup>st</sup>-century Levelling and Diffusion", in Bueno Alonso, J. L., González Álvarez, D., Pérez-Guerra, J and Rama Martínez, E. (eds), *'Of Varying Language and Opposing Creed': New Insights into Late Modern English*. Bern: Peter Lang, 25-46.
- Beal, Joan C. (2012a). '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 Multidimensional Approach*. Amsterdam, Benjamins, 13-29.
- Beal, Joan C. (2012b). "Evidence from Sources after 1500", in Nevalainen, T. and E.C. Traugott (eds.), *Handbook on the History of English: Rethinking Approaches to the History of English*, Oxford: Oxford University Press, 63-77.
- Benzie, W. (1972). *The Dublin Orator*. Menston: Scolar Press.
- Buchanan, James (1757). *Linguae Britannicae Vera Pronuntiatio*. London: A. Millar.
- Burn, John (1786). *A Pronouncing Dictionary of the English Language*. 2nd edn. Glasgow: Alex Adam for the Author and James Duncan.

- Cooper, Christopher (1687). *The English Teacher*. London
- Denison, David (1998). 'Syntax', in Romaine, S. (ed.) *The Cambridge History of the English Language*, vol. IV, 1776-1997, 92-329.
- De Smet, Hendrik (2005). 'A corpus of Late Modern English', *ICAME Journal* 29: 69-82.
- Dobson, Eric J. (1957). *English Pronunciation 1500-1700. Vol. 2, Phonology*. London: Oxford University Press.
- Douglas, Sylvester [1779] (1991). *A Treatise on the Provincial Dialect of Scotland*. ed. by Charles Jones. Edinburgh: Edinburgh University Press.
- Fitzmaurice, Susan (2007). 'Questions of standardization and representativeness in the development of social network-based corpora: the story of the Network of Eighteenth-century English Texts', in Beal, Joan C., Corrigan, Karen P. and Moisl, Hermann L. (eds.) *Creating and Digitizing Language Corpora volume 2: Diachronic Databases*. Basingstoke: Palgrave Macmillan, 49-81.
- Fritz, Clemens (2007). *From Early English in Australia to Australian English 1788-1900*. Frankfurt: Peter Lang.
- Görlach, Manfred (2001). *Eighteenth-Century English*. Heidelberg: Winter.
- Hickey, Raymond (1984). 'Syllable onsets in Irish English', *Word* 35: 67-74.
- Hickey, Raymond (2007). *Irish English: History and Present-day Forms*. Cambridge: Cambridge University Press.
- Hickey, Raymond (ed.) (2010). *Eighteenth-Century English*. Cambridge: Cambridge University Press.
- Hodges, Richard (1644). *The English Primrose*.
- Johnston, William (1764). *Pronouncing and Spelling Dictionary*. London: the Author.
- Johnston, William (1772). *Pronouncing and Spelling Dictionary*. 2nd edn. London: the Author.
- Jones, Charles (1989). *A History of English Phonology*. London: Longman.
- Jones, Charles (2006). *English Pronunciation in the Eighteenth and Nineteenth Centuries*. Basingstoke: Palgrave Macmillan.
- Jones, Mark (2008). 'What? Whence? And whither? A phonetic analysis of Scottish English "wh"'. Paper presented at the BAAP Colloquium, University of Sheffield, Tuesday 1st April 2008.
- Jones, Stephen (1797). *Sheridan Improved: A General Pronouncing and Explanatory Dictionary of the English Language*. London: Verner & Hood, J. Cushell, Ogilvie and Son and Lackington, Allen & Co.
- Kenrick, William (1773). *A New Dictionary of the English Language*. London: John and Francis Rivington, William Johnston et al.
- Kökeritz, Helge (1953). *Shakespeare's Pronunciation*.

- Kytö, Merja, Rydén, Mats and Smitterberg, Erik (2006). *Nineteenth-century English: Stability and Change*. Cambridge: Cambridge University Press.
- Ladefoged, Peter & Ian Maddieson (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.
- Local, John (1983). Making a transcription: The evolution of A. J. Ellis's Palaeotype. *Journal of the International Phonetic Association* 13(1). 2–12.
- MacMahon, Michael (1998). 'Phonology' in Romaine, S. (ed.) *The Cambridge History of the English Language*, vol. IV, 1776-1997.
- Maguire, Warren (2012). 'Mapping *The Existing Phonology of English Dialects*', *Dialectologica et Geolinguistica* 20: 84-107.
- OED = *The Oxford English Dictionary*. 2nd edn. OED Online. Oxford. (1989). Available from <http://dictionary.oed.com/>.
- Perry, William (1775). *The Royal Standard English Dictionary*. Edinburgh: David Willison for the Author.
- Phillips, Betty S. (1984). 'Word frequency and the actuation of sound change', *Language* 60: 320-342.
- Rissanen, Matti, Merja Kytö, Leena Kahlas-Tarkka, Matti Kilpiö, Saara Nevanlinna, Irma Taavitsainen, Terttu Nevalainen and Helena Raumolin-Brunberg (1991). *The Helsinki Corpus of English Texts*. Department of English, University of Helsinki.
- Robinson, Robert (1617). *The Art of Pronuntiation*. London: Nicolas Okes.
- Sheridan, Thomas (1761). *A Dissertation on the Causes of the Difficulties which Occur in Learning the English Tongue*. London: R. and J. Dodsley.
- Sheridan, Thomas (1780). *A General Dictionary of the English Language*. London: W. Straham.
- Smitterberg, E. (2005). *The Progressive in 19th-century English: a Process of Integration*. Amsterdam: Rodopi.
- Spence, Thomas (1775). *The Grand Repository of the English Language*. Newcastle: Thomas Saint.
- Tieken-Boon van Ostade, Ingrid (2009). *An Introduction to Late Modern English*. Edinburgh: Edinburgh University Press.
- Vachek, Josef (1954). 'On the phonetic and phonemic problems of the southern English WH-sounds'. *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft* 8: 165-94.
- Walker, John (1791). *A Critical Pronouncing Dictionary*. London: G. G. J. and J. Robinson and T. Cadell.
- Wallis, John (1653). *Grammatica Linguae Anglicanae*.
- Wells, John C. (1982). *Accents of English*. Cambridge: Cambridge University Press.
- Wyld, Henry C. (1936). *A History of Modern Colloquial English*. 3rd edn. Oxford: Blackwell.

Wyld, Henry C. (1923). *A History of Modern Colloquial English*. Oxford: Basil Blackwell.

## Websites

A Representative Corpus of Historical English Registers (ARCHER)

A Corpus of Late Eighteenth-Century Prose

<http://personalpages.manchester.ac.uk/staff/david.denison/late18c>

Corpus of Early English Correspondence Extension (CEECE)

<http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/ceece.html>

Corpus of Historical American English (COHA) <http://corpus.byu.edu/coha/>

Corpus of Late Modern English Texts (extended version) (CLMETEV)

<http://perswww.kuleuven.be/~u0044428/>

Diachronic Electronic Corpus of Tyneside English (DECTE) <http://research.ncl.ac.uk/decte/>

Eighteenth-Century Collections Online (ECCO) [http://gale.cengage.co.uk/product-](http://gale.cengage.co.uk/product-highlights/history/eighteenth-century-collections-online.aspx)

[highlights/history/eighteenth-century-collections-online.aspx](http://gale.cengage.co.uk/product-highlights/history/eighteenth-century-collections-online.aspx)

Eighteenth-Century English Grammars (ECEG)

<http://www.llc.manchester.ac.uk/research/projects/eceg/ecegdatabase/>

IvIE (Intonation in the British Isles) [www.phon.ox.ac.uk/IViE/](http://www.phon.ox.ac.uk/IViE/)

Literature On Line (LION) <http://lion.chadwyck.co.uk>

Network of Eighteenth-Century English Texts (NEET)

SAMPA <http://www.phon.ucl.ac.uk/home/sampa/>

Old Bailey Corpus <http://www.uni-giessen.de/oldbaileycorpus/>

Old Bailey Online <http://www.oldbaileyonline.org/>

Phonologie d'Anglais Contemporain (PAC) <http://w3.pac.univ-tlse2.fr/>

Penn-Helsinki Parsed Corpus of Modern British English (PPCMBE)

<http://www.ling.upenn.edu/hist-corpora/>

---

<sup>i</sup> This project was funded by the Rapid Response Fund of the Faculty of Arts, University of Sheffield, and the research assistant was Christine Wallis.

### Figures

Figure 1: Spence's 'New Alphabet'

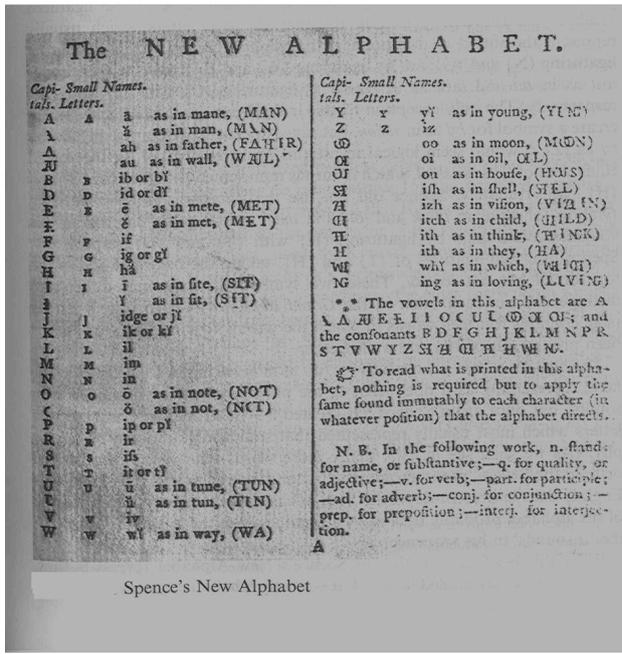


Figure 2: Walker's 'Table of Simple and Diphthongal Vowels'

[ xvi ]

*A Table of the Simple and Diphthongal Vowels referred to by the Figures over the Letters in this Dictionary.*

| ENGLISH SOUNDS.   | FRENCH SOUNDS.                   |
|---|----------------------------------|
| 1. à. The long slender English a, as in fâte, pâ-per, &c. (73)          | - - - é in fêc, épée.            |
| 2. â. The long Italian a, as in fâr, fâ-ther, pa-pâ, mam-mâ, (77)       | - - - a in fable, rable.         |
| 3. ä. The broad German a, as in fâll, wâll, wâ-ter, (83)                | - - - ä in âge, Cédäons.         |
| 4. ä. The short found of this Italian a, as in fât, mât, mâr-ry, (81)   | - - - a in fat, matin.           |
| 1. ê. The long e, as in mê, hère, mê-tre, mê-dium, (93)                 | - - - e in mitre, épître.        |
| 2. ê. The short e, as in mêt, lét, gêt, (95)                            | - - - e in mette, nette.         |
| 1. î. The long diphthongal i, as in plne, tl-tle, (105)                 | - - - ai in laïque, naïf.        |
| 2. î. The short simple i, as in pin, tlt-tle, (107)                     | - - - i in inné, titré.          |
| 1. ô. The long open o, as in nô, nôte, nô-tice, (162)                   | - - - o in globe, lobe.          |
| 2. ô. The long close o, as in nôve, prôve, (164)                        | - - - ou in mousoir, pouvoir.    |
| 3. ô. The long broad o, as in nôr, fôr, ôr; like the broad ä, (167)     | - - - o in or, for, encor.       |
| 4. ô. The short broad o, as in nôt, hôt, gôt, (163)                     | - - - o in botte, cote.          |
| 1. û. The long diphthongal u, as in tûbe, cû-pid, (171)                 | - - - iou in Cîoutat, chiourme.  |
| 2. û. The short simple u, as in tûb, câp, sâp, (172)                    | - - - eu in neuf, veuf.          |
| 3. û. The middle or obtuse u, as in bûll, fûll, pâll, (173)             | - - - ou in boule, foule, poule. |
| ôî. The long broad ô, and the short î, as in ôil, (299)                 | - - - ôi in cycloïde, herôïque.  |
| ôâ. The long broad ô, and the middle obtuse â, as in thôâ, pôând, (313) | asû in Aouîté.                   |

Figure 3: Recoding of Spence's 'New Alphabet'

| New Alphabet | Recoding | IPA | New Alphabet | Recoding | IPA    |
|--------------|----------|-----|--------------|----------|--------|
| {A}          | A        | e:  | {P}          | P        | p      |
| {A}          | a        | æ   | {R}          | R        | r      |
| {A}          | l        | ai  | {S}          | S        | s      |
| {AU}         | 2        | ɑ:  | {T}          | T        | t      |
| {B}          | B        | b   | {U}          | U        | ju:    |
| {D}          | D        | d   | {U}          | u        | u (aʔ) |
| {E}          | E        | i:  | {V}          | V        | v      |
| {E}          | e        | ɛ   | {W}          | W        | w      |
| {F}          | F        | f   | {Y}          | Y        | j      |
| {G}          | G        | g   | {Z}          | Z        | z      |
| {H}          | H        | h   | {Ø}          | w        | u:     |
| {I}          | I        | ai  | {Q}          | 3        | ɑ:     |
| {H}          | i        | i   | {CU}         | 4        | au     |
| {J}          | J        | dʒ  | {SI}         | s        | ʃ      |
| {K}          | K        | k   | {Z}          | z        | ʒ      |
| {L}          | L        | l   | {CH}         | C        | tʃ     |
| {M}          | M        | m   | {H}          | 5        | θ      |
| {N}          | N        | n   | {H}          | 6        | ð      |
| {O}          | O        | o:  | {WH}         | 7        | ʍ      |
| {C}          | o        | ɒ   | {NG}         | 8        | ŋ      |

Figure 4: Walker's (1791) Notation Transliterated into SAMPA

| Walker | Transcription (SAMPA-based) | IPA            |
|--------|-----------------------------|----------------|
| a1     | e                           | e:             |
| a2     | A                           | ɑ:             |
| a3     | Q:                          | ɔ:/ ɒ:         |
| a4     | a                           | a              |
| e1     | i                           | i:             |
| e2     | E                           | ɛ              |
| i1     | I                           | ɑɪ             |
| i2     | l                           | ɪ              |
| o1     | o                           | o:             |
| o2     | u                           | u:             |
| o3     | O                           | ɔ:             |
| o4     | Q                           | ɒ              |
| u1     | ju                          | ju:            |
| u2     | V                           | ʌ              |
| u3     | U                           | ʊ              |
| o3i2   | 3                           | ɔɪ             |
| o3u3   | 4                           | au             |
| sh     | S                           | ʃ              |
| zh     | Z                           | ʒ              |
| tch    | tS                          | tʃ             |
| th     | T                           | θ              |
| th     | D                           | ð              |
| ng     | N                           | ŋ              |
| j      | dZ                          | dʒ             |
| y      | j                           | j              |
| e      | ø                           |                |
| '      | '                           | primary stress |
| -      | -                           |                |

Figure 5 Wells's Keywords

(50) The standard lexical sets

|     | RP              | GenAm | keyword |     | RP              | GenAm | keyword |
|-----|-----------------|-------|---------|-----|-----------------|-------|---------|
| 1.  | ɪ               | ɪ     | KIT     | 13. | ɔː              | ɔ     | THOUGHT |
| 2.  | e               | ɛ     | DRESS   | 14. | əʊ              | o     | GOAT    |
| 3.  | æ               | æ     | TRAP    | 15. | uː              | u     | GOOSE   |
| 4.  | ɒ               | ɑ     | LOT     | 16. | aɪ              | aɪ    | PRICE   |
| 5.  | ʌ               | ʌ     | STRUT   | 17. | ɔɪ              | ɔɪ    | CHOICE  |
| 6.  | ʊ               | ʊ     | FOOT    | 18. | aʊ              | aʊ    | MOUTH   |
| 7.  | ɑː              | æ     | BATH    | 19. | ɪə <sup>1</sup> | ɪr    | NEAR    |
| 8.  | ɒ               | ɔ     | CLOTH   | 20. | ɛə <sup>1</sup> | ɛr    | SQUARE  |
| 9.  | ɜː <sup>1</sup> | ɜr    | NURSE   | 21. | ɑː <sup>1</sup> | ɑr    | START   |
| 10. | iː              | i     | FLEECE  | 22. | ɔː <sup>1</sup> | ɔr    | NORTH   |
| 11. | eɪ              | eɪ    | FACE    | 23. | ɔː <sup>1</sup> | ɔr    | FORCE   |
| 12. | ɑː              | ɑ     | PALM    | 24. | ʊə <sup>1</sup> | ʊr    | CURE    |

<sup>1</sup> with /r/ following before a vowel only.

Figure 6: Wells's FLEECE subset

- (62) FLEECE (a) *creep, meet, seek, beech, reef, teeth, seed, sleeve, seethe, cheese, seem, green, feel, see, tree, agree, needle, feeder, sweeten, ... ; grebe, these, Peter, even, ... ; shriek, brief, piece, believe, field, ... ; ceiling, Keith, Sheila, ... ; be, me, ... ; key, people;*
- (b) *reap, meat, speak, teach, leaf, beneath, peace, leash, bead, league, leave, breathe, please, team, mean, deal, sea, tea, feast, reason, weasel, easy, Easter, ... ; metre, equal, decent, legal, penal, complete, scene, ... ; deceive, receive, seize, ... ; Caesar, an(a)emic, Aesop, ... ; phoenix, subpoena, f(o)etus, ... ; quay;*
- (c) *police, unique, machine, prestige, elite, mosquito, casino, visa, trio, ski, chic, ... .*