

# Spelling in the Mind: phonemic-graphemic correspondence hypotheses

Paul Tench

Rhiannon Vaughn Griffiths

Ben Clarke

*Centre for Language and Communication Research, Cardiff University*

[tenchp@cf.ac.uk](mailto:tenchp@cf.ac.uk)

This study is an attempt to investigate the orthographic systems that operate in the mind of native speakers of English – not so much their spelling competence in respect of mental lexicons, but rather the orthographical systems themselves as a resource for spelling novel words or dialect words or even nonsense items that conform to the rules of English word phonology. For instance, how might an English speaker spell a ‘word’ they heard as /kli/? That is, what choices would they make from possible options; what are the systems in the mind that they could choose from; what is the total resource they hold in their minds for such a task? And, what strategies do they adopt when called upon to engage in such a task?

People have to engage in such a task if, for instance, they need to write a list of surnames of people who they are not familiar with, or if they need to write down what is to them an unfamiliar word when taking notes in a lecture. The reverse task is also often required: how to pronounce a name from its written form if they have never heard it before (how **do** you pronounce *Loughton* as the name of an underground destination in London?), or a pharmaceutical name, let’s say, like *Ibuprofen*, when you need to get some in a chemist’s shop. This particular task is a matter of matching graphemes with phonemic correspondences, but the former task – which we are attempting to investigate in this study – is to match phonemes and phonemic sequences with corresponding graphemes that are stored in the mind. So, for example, if a native speaker of English hears the name /'laʊtən/, how do they write it down if they have never seen it spelled? And what choices do they make? And what difference does it make if, instead, that ‘word’ is presented, say, as a verb?

If we follow Sebba’s recommended use of the terms *spelling*, *orthography*, *script* and *writing system* (Sebba 2007:10-11), we might conclude that we are investigating a native speaker’s deployment of the Roman script (= writing system) according to English orthographical rules that are stored in the mind to represent phonemic-graphemic correspondences in the spelling of novel words or actual coinages.

The study draws explicitly from four distinct ‘practices’ in applied linguistics, or ‘ways of doing things’ (Fairclough 2007), namely psycholinguistics, studies of social practice, description (for contrastive analysis) and field studies in orthography for unwritten languages. The main inspiration from psycholinguistics is Aitchison (2003), in which the mental lexicon is described in terms of a series of networks including the phonological, but, perhaps surprisingly, omitting the orthographical. She devotes considerable space to sound patterns and structure, but spelling receives scant attention, despite the observation that people who are literate know (or think they know!) how to spell all the words they can use. Derwing & Dow (1987) supplies examples of orthography in psycholinguistic research, and Derwing, Priestly & Rochet (1987: 45) actually suggests the kind of investigation that we have invested attention in.

An examination of social practice in orthography is amply provided by Kress (2000) in the case of children’s attempts at expressing meaning and creating an impact through their spelling, and by Sebba (2007) in the case of graffiti from around the world. We would be investigating how a group of adults engage their orthographic competence as a resource for the spelling of novel or unfamiliar words, and discovering the degree of agreement among them.

Thirdly, descriptions of the correspondences between graphemes and phonemes in English abound: for example, Wijk (1966), Venezky (1970), Scragg (1974), Carney (1994), Rollings (2004), Brown (2005), Ott (2006). Cook (2004) is a more light-hearted approach to the subject, but very illuminating. Contrastive studies in orthography, however, do **not** abound although Lado (1957) devoted a separate chapter to ‘writing systems’. Van Berkel (2005) is an example of a recent study, in this case contrasting Dutch and English spelling from the angle of a Dutch person learning English.

The fourth kind of practice in applied linguistics that is relevant to this particular investigation is the work of creating orthographies for unwritten languages. Practical measures, of an applied linguistic nature, for such an enterprise have been in place since Pike (1947) and Smalley (1963). Recent work includes Winter (1983), Coulmas (1989), Hinton & Hale (2001), Grenoble & Whaley (2006), Gippert, Himmelmann & Mosel (2006), and is exemplified in Tench (2007). The approach invariably involves a description of the phonology of a language as one basis for the creation of an orthography, with phonemes being matched to graphemes.

## **Bases for the hypotheses**

The focus of enquiry was to ascertain whether adult native speakers of English have a developed orthographical system that in some senses at least is parallel to a mature phonological competence, given the extensive discrepancy between pronunciation and spelling in current English. It seemed best to investigate this possibility by asking a group of adults to try and spell unfamiliar words from dictation, and see what correspondences between phonemes and letters each person displayed, and to what degree those adults held such correspondences in common. Thus, we would be able to see whether there was a system in the use of the orthographic resources in the mind, and offer an explanation to supplement Aitchison (2003).

Detailed hypotheses were prepared to cover British adults' competence in all the phonemic-graphemic correspondences that are required for English, based on observations of the most common and regular ways of representing each phoneme. It soon emerged that this was not to be a simple task. It became quickly apparent that we three collaborators did not share a simple equivalence of phonemes and letters, and that differences between us led us to realize that not only were there variations for many correspondences, but also that position in a word would be critical, and word class also: there does seem to be a paradigm shift in the orthographic conventions for 'common' and technical words. The designation 'technical/scientific' was unproblematic, but we took some time to agree on a term to describe 'non-technical' words; the term 'common' was chosen as the least problematic.

Another paradigm shift appears to happen with the spelling of words deemed to be of foreign origin; we did not have the resources to investigate this notion, but evidence for it did eventually appear in the analysis of the experimental data.

Eventually, the set of hypotheses comprised 10 general hypotheses and a series of hypotheses for each consonant and vowel phoneme for both common and technical words. They were all tested in pilot and full experiments that involved the dictation of 50 'novel' words; the hypothesis that participants would distinguish one system for 'common' words from a system for technical words was also tested. The data was analyzed for both the degree of uniformity among the subjects and the degree of conformity to the set of hypotheses.

The hypotheses that eventually emerged are based on the following assumptions:

- that we all store all our words in our minds; each person has their own personal store, which might be called their 'mental lexicon' – whatever their language or languages, and whatever their age or cultural background
- that all literate people store for each word a pronunciation and a spelling, in addition to meanings, sense relationships to other words, connotations, collocations, syntactic behaviour, etc
- that all literate people are aware of the pronunciation and spelling of each word in their mental lexicon and can call them to mind (even if a spelling happens not to conform with standard forms as found in published dictionaries – see Cook 2004)
- that most linguistically untrained people are not aware of the phonological and orthographical systems that operate in their minds that regulate the pronunciation and spelling of each word
- that there is a small minority of words in each person's lexicon which they feel uncertain about in either pronunciation or spelling, or both (see, for example, Kress 2000: 25); but these uncertainties do not undermine the claim for the existence and operation of either the phonological or the orthographical system
- that the orthographical system derives from the phonological system in all cultures that use a non-semantic writing system, ie those that use alphabets, syllabaries or some form of rebus writing (in which visual images stand for sound, not meaning - in contrast to pictographs)

- that just as it is possible to investigate the nature and operation of a phonological system, it is also possible to investigate the nature and operation of a corresponding orthographical system.

That final assumption is, basically, what we set out to do.

The following hypotheses relate to the English orthographic system for British adults. They have been set out as a basis for investigating the nature and operation of orthographic competence as a resource for spelling novel words which conform to the phonological structure of British English words, ie the productive orthographic system in the mind. The hypotheses are based on high frequency, regular, conventional correspondences between phonemes, phonemic sequences and syllable structure on the one hand and orthographic correspondences on the other; we believe that, as a general rule, this productive orthographic system will ignore low frequency, irregular, idiosyncratic spellings, such as

- for /k/ and /kw/ *liquor, liqueur, lacquer, choir*
- ‘silent’ letters as in *wrap, white, two, who*

but it will not ignore a perception of foreign loan words and technical terms, for which different systems exist.

It should be noted that the hypotheses do not correspond to a phonemic-style re-spelling transcription system as exists, for instance, in the *Oxford BBC Guide to Pronunciation* (Olausson & Sangster, 2006). Nor do they correspond to the type of ‘regularized spelling’ that is advocated for language teaching purposes – see, for example, Derwing, Priestly & Rochet (1987). These provide written forms as guides to pronunciation, ie grapheme-to-phoneme correspondences.

Rather, the hypotheses in our study seek to represent the **Sound-to-Spelling** (phoneme-to-grapheme) correspondences that exist as a system in the mind, which enables a native speaker/writer to **do** something, namely to spell novel, native-like words. We are well aware that issues of age, gender and accent have to be taken into account. For example, on hearing a ‘word’ as /vɑ:θ/, an adult with a southern British accent may well ‘spell’ it as <vath>, whereas a Northerner may well be more likely to ‘spell’ it as <varth>. However, for the sake of ease of reference, the phonemic inventory that the hypotheses are based on is that of Southern England Standard Pronunciation (formerly known as ‘RP’):

#### *consonants*

plosives:	p, b, t, d, k, g
nasals:	m, n, ŋ
fricatives:	f, v, θ, ð, s, z, ʃ, ʒ, h
affricates:	tʃ, dʒ
approximants:	l, r, j, w

#### *vowels: strong*

short:	ɪ, ɛ, æ, ɒ, ʊ, ʌ
long monophthongs:	i:, ɑ:, ɔ:, u:, ɜ:
long diphthongs:	eɪ, aɪ, ɔɪ, əʊ, aʊ, ɪə, eə, ʊə
<i>weak:</i>	ɪ, u, ə

#### *stress*

primary and secondary stress, and no (‘weak’) stress

The orthographic system is presented as sets of ‘graphemes’, with terminology deriving from Sgall (1987):

‘protographemes’ (single letters):	a, b, c, d, e, etc
which may be ‘doubled’:	bb, cc, dd, ee, etc
‘bigraphemes’ (regular pairs for single phonemes):	ch, sh, th, ai, ay, ea, er, ge, etc
‘trigraphemes’ (regular trios for single phonemes):	air, are, ear, eer, oor, our, etc
‘quadrigrapheme’ (regular quad for single phoneme):	aire

Thus a ‘grapheme’ may consist of one or more letters to represent a single sound. One ‘quasi-protographeme’ (a single letter that corresponds to a phoneme sequence) is recognized: <x> for /ks/ or /gz/; and one ‘quasi-bigrapheme’ (a regular pair for a phoneme sequence) is recognized: <qu> for /kw/. Phonemes are represented as: /b/; graphemes as <b>.

The hypotheses are presented as

- predictions (“that they will...”; 67% - 100%)
- options (“that they will ... either ...or” ...; 33% - 66%)
- possibilities (“that they may ...”; 1% - 32%)

## General hypotheses

We postulate 10 hypotheses of a very general nature that relate to the use of the alphabet, stress, grammatical information and lexis:

**G1** that adult native speakers of English will utilize all the letters of the Roman alphabet as used for Standard English for attempting to spell novel words that conform to word phonological systems of English

**G2** that they will use the consonant letters <b, c, d, f, g, h, j, k, l, m, n, p, r, s, t, v, w, y, z> as single graphemes (‘protographemes’), in all positions of a word, except <h, j, k, v> in final position.

**G2a** that they will also use the consonant letters <b, c, d, f, g, l, m, n, p, r, s, t, z> doubled in certain contexts (‘double graphemes’)

**G2b** that they will not use the consonant letters <h, j, k, v, w, x, y> doubled

**G2c** that they will use the following consonant letter pairs (‘bigraphemes’) to represent single phonemes: <ch, ph, sh, th; sc; ck, dg>

**G2d** that they will use <tch> as a ‘trigrapheme’

**G3** that they will use the vowel letters <a, e, i, o, u> as single graphemes

**G3a** that they will use the vowel letters <e, o> doubled to represent single phonemes

**G3b** that they will not use the vowel letters <a, i, u> doubled

**G3c** that they will use the following vowel letter pairs (‘bigraphemes’) to represent single phonemes: <ai, au, ea, ia, ie, oa, oe, ou, ua, ue>

**G4** that they will also use combinations of vowels and consonants to represent single phonemes

**G4a** that they will use each vowel letter with <r> as pairs (‘bigraphemes’) to represent single phonemes: <ar, er, ir, or, ur>

**G4b** that they will use the following pairs (‘bigraphemes’) in word-final position: <aw, ay, ew, ow, oy; se, ce, ge, ve; ci, si, ti> and also the following sequences (‘trigraphemes’ and one ‘quadrigrapheme’) in word-final position: <gue, the, dge; air, are, ear, eer, ere, ire, igh, oar, oor, ore, our, ure; aire>

**G5** that they will use <x> as a ‘quasi-protographeme’ to represent /ks, gz/ and <qu> as a ‘quasi-bigrapheme’ to represent /kw/

**G6** that they will use a minimum of three letters for any word that is perceived to be a ‘common’ lexical item, even though grammatical items may consist of only one or two letters – the ‘three letter rule’. (Compare *I/eye; an/Ann; at/add; so/sow; be/bee*, etc; but note *do, go* as lexical as well as grammatical items. There is one notable exception: *ox*.)

**G6a** that they may use a minimum of two letters for a word perceived to be a foreign loan word or a technical term: *id, pi*

**G7** that they will not indicate stress



**G8** that they may use a hyphen <-> as a link between a morpheme with a perceived final vowel letter and a morpheme with a perceived initial vowel letter, eg *co-operate*, *de-ice*, *psycho-analysis*

**G9** that they may take account of grammatical and lexical information in specific cases, eg

- the interpretation of word-final /-ks/ as a component of a base form (eg *tax*, *fix*, *lax*), or the *s*-form of a noun or verb (eg *lacks*), or as a non-count noun denoting a branch of study (eg *physics*); or word-final /d/ as a component of a base form (eg *pride*), or the past tense form of a verb (eg *pried*), etc;
- parts of words that conform to perceived morphemes, ie inflections, and derivations like *-tion*, *-al*, *-ous*, etc;
- perceived word classes, eg word-final /-əs/ as an adjective (eg *calamitous*), or a noun denoting an object (eg *apparatus*) or, for some speakers, a noun denoting a medical condition (eg *arthritis*); or word-final /-l/, -əl/ as a noun (eg *title*) or an adjective (eg *vital*), etc

**G10** that they will distinguish a spelling system for words that are perceived to conform to common English words from those that are perceived to conform to technical words or loan words from another language.

Evidence for **G10** includes the relative lack of double graphemes in technical words. In common words, a doubled consonant letter usually indicates a single consonant phoneme with a preceding short vowel in a non-final syllable, eg *supper* in contrast to *super* and *sup*. If, however, the short vowel is spelled with a vowel bigrapheme, eg as in *couple*, the doubled form does not occur; compare *supple*. Furthermore, the doubled form occurs in common monosyllabic lexical items that have no onset consonant, as **G6** (the 'three letter rule'); other examples are *ass*, *ebb*, *egg*, *ill*, *inn*, *odd* (also *err*). The doubled form often also occurs as a single consonant phoneme after an initial weak but before a following strong syllable, in words of Latin origin, eg *supply*, *appear* (compared with *apart*), *effect* (cf *elect*), *immense* (cf *imagine*), *oppose* (cf *opaque*).

On the other hand, doubled consonants rarely appear in technical words. This no doubt reflects the classical origins of technical words, since the distinction between long and short vowels is not a feature of Classical Greek and Latin phonology; and doubled consonants are generally confined to morpheme boundaries, as in the examples immediately above. Indeed, words of classical origin that entered the English language as a result of the Renaissance often offended the historical Anglo-Saxon 'rule', which was designed to compensate for the relative lack of vowel letters in the Roman alphabet for the larger vowel systems of Germanic languages. A few examples must suffice: *chapel* (compared with *apple*), *model* (cf *muddle*), *metal* (cf *mettle*), *stomach* (cf *tummy*). Nevertheless, the observation remains that generally speaking, it is the stock of common words that employs the doubled consonants, rather than the technical words.

It is also the case that the stock of common words employs doubled vowel graphemes, but that technical words do not; again, this reflects the simpler vowel inventories of the Classical languages.

Technical words also display a variety of alternative bigraphemes for single consonants, eg <ph> for /f/, <ch> for /k/, <rh> for /r/. Furthermore, there is no /ð/; /j/ is confined to occurring with /u:/; /w/ is rare apart from occurring with /k/; /ʃ, tʃ/ do not, as a rule, occur in word-initial position. <j> rarely is used for /dʒ/. Technical words also use <y> as an alternative to <i>; all the vowel letters, as protographemes, are used for both long and short vowels, eg <i/y> for both /i, ai/ (eg *glycerine*; *glycogen*); <e> for /ɛ, i:/; <a> for /æ, eɪ/; <o> for /ɒ, əʊ/; <u> for /ʌ, u:/ . The vowels /ʊ, au/ do not seem to occur in technical words. The centring diphthongs - /ɪə, eə, ʊə/ - are more usually spelled with final <a> rather than with <r>.

It is thus necessary to distinguish between hypotheses of phonemic-graphemic correspondences in the registers of common and technical words. It would also be necessary to establish another set of hypotheses for words that sound like foreign loan words, eg with word-final /ʒ/ and with word-final primary stress after an earlier secondary stress, eg /bu:'keɪ/ (*bouquet*), /meɪzə'net/ (*maisonette*); but that is beyond the scope of this present study.

## Hypotheses for phoneme-grapheme correspondences

### Common words

The hypotheses are set out simply phoneme by phoneme. The consonants have a basic form, a doubled form and, in some cases, a special form. The **basic form** is the grapheme which represents the consonant phoneme

in *word-initial* position;

in *medial* position after a long vowel, or a short vowel perceived as a vowel bigrapheme, and after a weak vowel;

in *final* position, including before another consonant, but excluding the case of a monosyllabic lexical item that has no onset consonant.

Example: *pat, paper, staple, couple, callipers, multiple, tap, apt, tulip*.

The **doubled form** is the grapheme which represents the consonant

in *medial* position after a short vowel perceived as a single grapheme (protographeme), and optionally after a word-initial weak vowel;

in final position in a monosyllabic lexical item that has no onset consonant (the ‘three letter rule’).

Example: *supper* (~*super*), *apart/appear*, *ebb*.

Any **special form** is the grapheme which represents the consonant in specific environments, as indicated.

Table 1 should be read as: for the given phoneme *x* adult native speakers of English will use grapheme *y* in its basic or doubled form as required, except in special cases when they will use either one grapheme or another (ie “optionally”), or may (ie “possibly”) use an alternative grapheme *z*. The simplest hypotheses are listed first, roughly following the traditional order of phonological studies:

	Basic	Doubled	Special
p	<b>p</b>	<b>pp</b>	
b	<b>b</b>	<b>bb</b>	
t	<b>t</b>	<b>tt</b>	
d	<b>d</b>	<b>dd</b>	
m	<b>m</b>	<b>mm</b>	
n	<b>n</b>	<b>nn</b>	
f	<b>f</b>	<b>ff</b>	<b>ff</b> also in final position after a short vowel, in lexical items (eg <i>cuff, mastiff</i> )
r	<b>r</b>	<b>rr</b>	
h	<b>h</b>		
k	<b>c</b>		before <a, o, u> and consonants; in final position after weak /ɪ/ (eg <i>music</i> )
	<b>k</b>		before <e, i, y>; in final position after long vowels and short vowels perceived as vowel bigraphemes (eg <i>beak, book</i> )
		<b>cc</b>	optionally for <c> after word-initial weak vowel (eg <i>account/acute</i> )
		<b>ck</b>	as the doubled form after a short vowel perceived as a single grapheme; in final position after short and weak vowels (other than weak /ɪ/) (eg <i>pack, paddock</i> )
			<b>qu</b> for the sequence /kw/ (eg <i>frequent</i> )
			<b>x</b> for the sequence /ks/ (eg <i>axle</i> ), except in a final unstressed syllable in a noun denoting study (eg <i>electronics</i> )
g	<b>g</b>	<b>gg</b>	possibly <b>gu</b> before <e, i, y> (eg <i>guess, guide, guy</i> )
			<b>gue</b> in final position after long vowels (eg <i>vague</i> )
			<b>x</b> for the sequence /gz/ before a vowel (eg <i>exist</i> )
ŋ	<b>n, ng</b>		<b>n</b> before consonants; <b>ng</b> in final position (eg <i>wink, wing</i> )
v	<b>v</b>		<b>ve</b> in final position (eg <i>have</i> )
θ	<b>th</b>		
ð	<b>th</b>		<b>the</b> in final position (eg <i>breathe</i> )
s	<b>s</b>		optionally <b>c</b> before <e, i, y> (eg <i>cent, circle, cycle</i> )
			either <b>se</b> or <b>ce</b> for word-final /s/ after a long vowel or consonant (eg <i>loose, lace; sparse, scarce; dense, pence</i> )

	<b>ss</b>	possibly <b>ss</b> or <b>ce</b> for word-final /s/ after a weak vowel (eg <i>harness, palace</i> ) possibly also in final position after a short vowel (eg <i>lass</i> ; cf <i>gas</i> ), and the long vowel perceived as <a> in Southern England Standard Pronunciation, eg <i>pass</i> after a word-initial weak vowel, optionally <b>sc</b> before <e, i, y> (eg <i>assent, ascent/asylum</i> )
<b>z</b>	<b>z</b>	optionally <b>s</b> in medial position after a long vowel (eg <i>measles</i> ) either <b>ze</b> or <b>se</b> in final position after a long vowel (eg <i>prise, prize</i> ) <b>s</b> in final <-ism> (eg <i>prism</i> )
	<b>zz</b>	possibly <b>z</b> or <b>s</b> (eg <i>buzzard, wizard, visit</i> ) <b>zz</b> also in final position after a short vowel, in lexical items (eg <i>buzz, jazz</i> )
<b>ʃ</b>	<b>sh</b>	<b>ti</b> after long vowels and consonants, and <b>ssi</b> after short vowels for /-ʃən/ as a noun for a process or object (orthographic <i>-tion, -ssion</i> ) <b>ci</b> for /-ʃən/ as a noun for a profession (orthographic <i>-cian</i> ) <b>ti</b> or <b>ci</b> before /-əl/ as an adjective (orthographic <i>-tial, -cial</i> ) <b>ss</b> for /-ʃə/ as a noun for a process or object (orthographic <i>-ssure</i> )
<b>ʒ</b>		<b>si</b> for /-ʒən/ (orthographic <i>-sion</i> ) <b>s</b> or <b>z</b> for /-ʒə/ as a noun (orthographic <i>-sure, -zure</i> ) <b>s</b> before /-uəl/ as an adjective (orthographic <i>-sual</i> ) <b>ge</b> in final position (eg <i>beige</i> )
<b>tʃ</b>	<b>ch</b>	<b>tch</b> after a short vowel in lexical items (eg <i>satchel, match</i> ; but <i>much</i> ) <b>t</b> after a long vowel or a consonant for /-tʃə/ (orthographic <i>-ture</i> ); also before /-uəl/ as an adjective (orthographic <i>-tual</i> )
<b>dʒ</b>	<b>j</b>	optionally <b>g</b> before <e, i, y> in initial and medial positions (eg <i>gem, gin, gym; dungeon, rigid</i> ) <b>ge</b> in final position after a long or weak vowel, or a consonant (eg <i>rage, marriage, range</i> ) <b>dg</b> in medial position after a short vowel (eg <i>fidget</i> ) <b>dge</b> in final position after a short vowel (eg <i>hedge</i> )
<b>l</b>	<b>l</b> <b>ll</b>	either <b>le, el</b> or <b>al</b> for syllabic /l/ and final /-əl/ (eg <i>peddle, model, pedal; rabble, rebel (n), oval</i> )
<b>j</b>	<b>y</b>	optionally unspelled before /u:, ʊə, u/ (eg <i>usual, Europe</i> )
<b>w</b>	<b>w</b>	except after /k/ (= <qu>)

Table1: Phonemic-graphemic correspondence hypotheses for consonants in common words

The vowels have a **basic form** for syllables in non-final position in a word, a **final form** for open syllables in word-final position where permissible, and, in some cases, a **special form** which presents the vowel phoneme in specific environments. Short vowels are given first, followed by long monophthongal and diphthongal vowels, the weak vowels, and special forms with preceding /w/. Table 2 should be read as: for the given phoneme *x* adult native speakers of English will use grapheme *y* (options *y/z*) in its basic or final form as required, except in specific environments when they will use either one grapheme or another (ie “optionally”), or may (ie “possibly”) use an alternative.

	Basic	Final	Special
<b>i</b>	<b>i</b>		
<b>ε</b>	<b>e</b>		
<b>æ</b>	<b>a</b>		
<b>ɒ</b>	<b>o</b>		
<b>ʌ</b>	<b>u</b>		possibly <b>o</b> before <v> (eg <i>cover, dove</i> )
<b>ʊ</b>	<b>oo</b>		<b>u</b> before /ʃ, l/ (eg <i>push, pull</i> )
<b>i:</b>	<b>ee/ea</b>	<b>ee/ea</b>	<b>i</b> before final /ʒ/, and possibly before final /n/ (orthographic <i>-ige, -ine</i> ) possibly <b>e</b> before a single consonant (eg <i>completion</i> )
<b>ɑ:</b>	<b>ar</b>	<b>ar</b>	<b>a</b> before final /ʒ/, and possibly before final /s/ (orthographic <i>-age, -ass</i> ) NB <b>ar</b> in non-final position must be followed by a consonant (eg <i>barter</i> )
<b>ɜ:</b>	<b>er/ir/</b>	<b>er/ir/</b>	possibly also <b>ear</b> and <b>our</b> in non-final position (eg <i>earth, journey</i> )

	<b>ur</b>	<b>ur</b>	possibly <b>eur</b> as primary stress in final position (eg <i>chauffeur</i> )
ɔ:	<b>or/oar/</b> <b>au/aw</b>	<b>or/oar/</b> <b>ore/aw</b>	NB <b>au</b> before a consonant only
u:	<b>oo/u</b>	<b>oo/ue</b>	<b>u.e</b> in a final closed syllable (eg <i>rule</i> )
(j)u:	<b>ew/eu/</b> <b>u/you</b>	<b>ew/ue</b>	NB <b>eu/u</b> before a consonant only; <b>u.e</b> in a final closed syllable (eg <i>tune</i> )
eɪ	<b>ai/a</b>	<b>ay</b>	<b>ay</b> before a vowel in medial position; <b>ai</b> or <b>a.e</b> in a final closed syllable (eg <i>gait, gate</i> )
aɪ	<b>i</b>	<b>y/ie</b>	<b>igh</b> before /t/ (eg <i>light</i> ); <b>i.e</b> in other final closed syllables (eg <i>line</i> )
ɔɪ	<b>oi</b>	<b>oy</b>	<b>oy</b> before a vowel in medial position
əʊ	<b>o/oa</b>	<b>o/oe/</b> <b>ow</b>	<b>o, oa</b> or <b>o.e</b> in a final closed syllable (eg <i>post, moat, mote</i> )
aʊ	<b>ou/ow</b>	<b>ow</b>	<b>ow</b> before a vowel in medial position
ɪə	<b>ear/eer</b>	<b>ear/eer/</b> <b>ere</b>	
ɛə	<b>air/ar</b>	<b>air/are</b>	possibly <b>aire</b> as primary stress in final position (eg <i>solitaire</i> ) NB <b>air/ar</b> in non-final position must be followed by a vowel (eg <i>wary</i> )
ʊə	<b>oor/our</b>	<b>oor/our/</b> <b>ure</b>	
ɪ	<b>i/e</b>	<b>y/ey</b>	<b>ie</b> before perceived inflection (eg <i>berries; buries</i> ) possibly <b>ie</b> after a <k> in final position (eg <i>bookie</i> )
ɪ (weak)	<b>i/e</b>		(NB word-initial <ex->)
(j)u	<b>u</b>		
ə	<b>a/e/o/</b> <b>u/ou</b>	<b>a/er/re/</b> <b>or/ure</b>	
wʌ	<b>wo</b>		(eg, <i>won, wonder, worry</i> , but <i>swum, swung</i> )
wɜ:	<b>wor</b>	<b>were</b>	(eg <i>work, world</i> ; when preceded by /k/: <b>quir</b> (eg <i>squirt</i> ))
wɒ	<b>wa</b>		(eg <i>want, wander</i> ; when preceded by /k/: <b>qua</b> (eg <i>quality</i> ))
wɔ:	<b>war/</b> <b>wor</b>	<b>war/</b> <b>wore</b>	(eg <i>warn/worn; war/wore</i> ; when preceded by /k/: <b>quar/quor</b> (eg <i>quart, quorum</i> ))

Table 2: Phonemic-graphemic correspondence hypotheses for vowels in common words

### Technical words

Consonants have a much simpler set of phonemic-graphemic correspondences in technical words. Since doubled forms rarely occur, it is hypothesized that as soon as they perceive a word to be technical, they will disregard that ‘option’. Consonants have a **basic form**; in some cases, there is an **alternative form** (labelled below as ‘Alt.’); and also a **special form** which represents the consonant in specific environments. Table 3 should be read in a similar way as above.

	Basic	Alt.	Special
p	<b>p</b>		
b	<b>b</b>		
t	<b>t</b>		
d	<b>d</b>		
g	<b>g</b>		
m	<b>m</b>		
n	<b>n</b>		
ŋ	<b>n</b>		before /k, g/ only
f	<b>f</b>	<b>ph</b>	

v	<b>v</b>	
θ	<b>th</b>	
h	<b>h</b>	
s	<b>s</b>	optionally <b>c</b> before <e, i, y> possibly <b>tz</b> for the sequence /ts/
z	<b>z</b>	optionally <b>x</b> in initial position optionally <b>s</b> in medial position
l	<b>l</b>	possibly <b>ll</b> after a short vowel (eg <i>callisthenics</i> )
r	<b>r</b>	<b>rh</b> possibly <b>rr</b> or <b>rrh</b> after a short or weak vowel (eg <i>diarrhoea</i> )
k	<b>c</b>	<b>k</b> before <a, o, u> and <l, r>; <b>k</b> in initial position (eg <i>kaolin</i> )
	<b>ch</b>	<b>k</b> before <e, i, y> and optionally <l, r>; <b>k</b> in initial position (eg <i>kerosene</i> ) possibly <b>cc</b> or <b>cch</b> after a short vowel in medial position (eg <i>saccharine</i> ) <b>qu</b> for the sequence /kw/ <b>x</b> for the sequence /ks/; possibly <b>cc</b> before <e, i, y> in medial position <b>cs</b> for the sequence /ks/ in a final unstressed syllable in a noun denoting study
ʃ		<b>ti</b> after long vowels and consonants, and <b>ssi</b> after short vowels for /-ʃən/ as a noun for a process or object (orthographic <i>-tion, -ssion</i> ) <b>ci</b> for /-ʃən/ as a noun for a profession (orthographic <i>-cian</i> ) <b>ti</b> or <b>ci</b> before /-əl/ as an adjective (orthographic <i>-tial, -cial</i> ) <b>ss</b> for /-ʃə/ as a noun for a process or object (orthographic <i>-ure</i> ) possibly <b>x</b> for /kʃ/ before /u/
ʒ		<b>si</b> for /-ʒən/ (orthographic <i>-sion</i> ) <b>s</b> or <b>z</b> for /-ʒə/ as a noun (orthographic <i>-sure, -zure</i> ) <b>s</b> before /-uəl/ as an adjective (orthographic <i>-sual</i> ) <b>ge</b> in final position after a long vowel (eg <i>trriage</i> )
tʃ		<b>t</b> after a long vowel or a consonant for /-tʃə/ (orthographic <i>-ture</i> ) also as /-uəl/ as an adjective (orthographic <i>-tual</i> )
dʒ		<b>g</b> before <e, i, y> in initial and medial positions <b>j</b> before <a, o, u> (eg <i>conjunctivitis</i> ) <b>d</b> as /-uəl/ as an adjective (orthographic <i>-dual</i> )
j		unspelled before /u:, uə, u/ (eg <i>urology</i> )

Table 3: Phonemic-graphemic correspondence hypotheses for consonants in technical words

The vowels have a **basic form** without distinctions for open final syllables; in some cases, they also have an **alternative form**, and also a **special form** which represents the vowel in specific environments. Table 4 follows the pattern for vowels above, except that no special provision is needed for vowels following /w/.

	Basic	Alt.	Special
ɪ	<b>i</b>	<b>y</b>	possibly <b>e</b> in 'ex'
ɛ	<b>e</b>		
æ	<b>a</b>		
ɒ	<b>o</b>		<b>a</b> following /kw/ (orthographic <i>qua-</i> )
ʌ	<b>u</b>		
i:	<b>e</b>		<b>e.e</b> in a final closed syllable <b>i</b> possibly before final /n/ (orthographic <i>-ine</i> )
ɑ:	<b>ar</b>		
ɜ:	<b>ur</b>	<b>er/ir</b>	
ɔ:	<b>au</b>	<b>or</b>	
u:	<b>u</b>		<b>u.e</b> in a final closed syllable
(j)u:	<b>u</b>	<b>eu</b>	<b>u.e</b> in a final closed syllable

eɪ	<b>a</b>	<b>a.e</b> in a final closed syllable
aɪ	<b>i</b> <b>y</b>	<b>i.e/y.e</b> in a final closed syllable
ɔɪ	<b>oi</b>	<b>oy</b> before a vowel in medial position
əʊ	<b>o</b>	<b>o.e</b> in a final closed syllable
ɪə	<b>er</b>	the <r> must be followed by a vowel; possibly <b>ia</b> or <b>ea</b> in final position
ɛə	<b>ar/er</b> <b>aer/ear</b>	the <r> must be followed by a vowel
ʊə	<b>ur</b>	the <r> must be followed by a vowel; possibly <b>ua</b> in final position
(jʊə)	<b>ur</b> <b>eur</b>	the <r> must be followed by a vowel; possibly <b>ua</b> in final position
i	<b>i</b> <b>e</b>	(before a vowel) <b>y</b> in final position
ɪ (weak)	<b>i</b> <b>e</b>	(before a consonant)
(j)u	<b>u</b>	
ə	<b>a/e/o/</b>	<b>a/er/re</b> in final position

Table 4: Phonemic-graphemic correspondence hypotheses for vowels in technical words

The hypotheses for technical words are a good deal simpler than for common words. Common words require 71 predictions, options and possibilities for consonants, and 67 for vowels; technical words, on the other hand, require just 48 predictions, options and possibilities for consonants, and 47 for vowels. This comprehensive set of hypotheses highlights the magnitude of the complexity of phonemic-graphemic correspondences that, we claim, are contained in the mind of adult native speakers of English. Now, to test the claim.

## Experiment

An experimental mechanism to test all 233 hypotheses would clearly become unwieldy, even though a number of hypotheses could be tested within a single novel ‘word’. We decided that 50 ‘words’ would suffice to contain enough representative hypotheses and not over-tax participants. A pilot study proved to be an invaluable resource for the design of the experiment proper. It helped

- to define more precisely the procedure for conducting the experiment, including the request for ‘qualitative data’;
- to construct items that could realistically be regarded as ‘English’, eg to eliminate sequences that were not typical of English, like /mu:sk/ with a long vowel before final /-sk/ (apart from the distinctive Southern England pronunciation of *ask*, etc);
- to eliminate items that were too similar to taboo words for comfort;
- to eliminate items that could be spelt as existing words;
- to expect ‘legitimate’ alternative spellings;
- to exclude items as unnecessary for further investigation because of total agreement, eg /lʌs/ as <luss>, /kədək/ with <dd>, etc;
- to give experience for the timing of the list of items; it gave us confidence that a list of 50 items would not in fact be too onerous – it took less than 20 minutes to complete;
- to confirm a decision to use two speakers, one for the ‘common’ words (PT) and one for the technical words (BC), but without indicating to the subjects why; it helped us to talk about “PT’s words” and “BC’s words” without reference to there being any difference in the nature of the words themselves;
- to confirm our suspicions that some grammatical information is occasionally necessary, eg final /d/ as a past participle marker, or not;
- to require information on any recognized linguistic disorder in any participant, eg dyslexia;
- and to decide to audio record the dictation of items as a checking mechanism for consistency.

A new set of 50 items was presented to 45 participants, native English-speaking students with a narrow age range (19-23 years), similar cultural background and level of education, who remained anonymous but were required to indicate their gender, regional background in terms of accent, and any linguistic disorder. The nature and purpose of the experiment was explained: research into the links between the sounds of words and their spelling and in the processes that take place in the mind in deciding how to spell a novel word. “All we want you to do is to write down on the sheet of paper in front of you the

most likely spelling for each of the 50 words which you will hear. This is **not** testing you in any way, and there is no right or wrong answer here, but your responses will help us to test these hypotheses”. They were also encouraged to write alternative spellings. They were also encouraged to use their best print form of handwriting. They were also invited to provide comments on the way they made their decisions. They were provided with a sheet with numbers 1 to 50, and with space at the bottom for their comments; the design of the sheet was to expedite our own analysis.

The 50 words appear in Table 5. The first 35 were dictated by PT, and the final 15 by BC, as in the pilot. The very first 15 were said to be verbs, and as a reminder of this, occasionally an item was presented as *to* ..., but they were designed to especially test the spelling of vowels in final open syllables. The following 10 were said to be nouns, and occasionally were presented as *a* ..., but they were designed to especially test vowels in closed syllables. These first 25 items, naturally, also provided opportunities to test consonant hypotheses. Numbers 25 to 35 were said to be adjectives, and were occasionally presented in the form *they were really/very/pretty...*, but were designed to test a variety of consonant and vowel hypotheses. The final 15, dictated by BC, were designed to sound technical. All were delivered as closely as possible to a Southern England Standard Pronunciation.

In order to save space, we also present our expectations from the hypotheses, and an analysis of the participants’ responses. Only significant percentages are given; hence, not all responses amount to 100%. (RVG took responsibility for 1-35; BC for 36-50.)

	Expectation	Responses
1. kli:	<i>clee/clea</i>	<i>c</i> 70.4; <i>k</i> 29.6   <i>ee</i> 78; <i>ea</i> 19.2
2. rɪ'pɑ:	<i>repar</i>	<i>e</i> 97.8   <i>ar</i> 87.2; <i>are</i> 10.5
3. bə'lɔ:	<i>belor/-oar/-ore/-aw</i>	<i>e</i> 98.9   <i>ore</i> 65.6; <i>or</i> 13.3; <i>aw</i> 6.7; <i>our</i> 5.6
4. kən'vju:	<i>convue/-view</i>	<i>o</i> 100   ( <i>i</i> ) <i>ew</i> 63.3; <i>ue</i> 33.3
5. draʊ	<i>drow</i>	<i>ow</i> 65.5; <i>ough</i> 17.2
6. vei	<i>vay</i>	<i>ay</i> 52.3; <i>ey</i> 23.8; <i>igh</i> 8.3
7. pə'trɜ:	<i>pertrer/-ir/-ur</i>	<i>e</i> 59; <i>a</i> 23.1; <i>or</i> 7.7   <i>er</i> 28.2; <i>ur</i> 16.7; <i>eur</i> 16.7; <i>ure</i> 7.7
8. kaɪ	<i>kie</i> (or <i>kye</i> – three letter rule)	<i>k</i> 87.3   <i>ai</i> 53.6; <i>ie</i> 10; <i>ye</i> 10; <i>ay</i> 9.2; <i>igh</i> 4.6
9. fɔɪ	<i>foy</i>	<i>oy</i> 51.1; <i>oi</i> 30.7; <i>oye</i> 13.6
10. raɪə	<i>rire</i>	<i>yer</i> 23.2; <i>ier</i> 12.2; <i>ya</i> 8.5; <i>ior</i> 7.3; <i>ia</i> 4.9
11. trəʊ	<i>tro/-oe/-ow</i>	<i>ow</i> 62.5; <i>o</i> 11.4; <i>ough</i> 11.4; <i>ou</i> 8; <i>owe</i> 4.5; <i>oe</i> 3.4
12. ə'nɪə	<i>anear/ann/-eer/-ere</i>	<i>a</i> 45.2; <i>e</i> 42.9   <i>n</i> 88.9; <i>nn</i> 8.9   <i>ear</i> 63.1; <i>eer</i> 29.8
13. bə'dʒʊə	<i>bedure</i>	<i>ure</i> 34.5; <i>uer</i> 22.6; <i>ewer</i> 15.5; <i>ua</i> 8.3; <i>uar</i> 6
14. raʊə	<i>rower</i>	<i>ow</i> 56.8; <i>aw</i> 19.3
15. ə'meə	<i>amare/amm/-air</i>	<i>a</i> 72; <i>e</i> 24.4   <i>m</i> 95.6; <i>mm</i> 4.4   <i>are</i> 31.7; <i>air</i> 17.1; <i>ayer</i> 15.9; <i>ere</i> 12.2; <i>aire</i> 6.1
16. klaɪt	<i>clight/clite</i>	<i>k</i> 52.6; <i>c</i> 42.9   <i>igh</i> 38.4; <i>i.e</i> 30.2; <i>ai</i> 9.3; <i>igh</i> 7
17. gleɪt	<i>glait/glate</i>	<i>a.e</i> 69.3; <i>ai</i> 9.1; <i>ai.e</i> 8; <i>igh</i> 6.8
18. 'kwɒtɪə	<i>quattier/-ia</i>	<i>qu</i> 90.5   <i>qua</i> 66.7; <i>quo</i> 21.4; <i>qu/coi</i> 7.1   <i>t</i> 84.4; <i>tt</i> 15.6   <i>ier</i> 82.1; <i>ia</i> 6
19. 'si:nɪŋ	<i>seening/sea/-sce-</i>	<i>s</i> 64.8; <i>sc</i> 22.7; <i>c</i> 12.5   <i>ee</i> 64.8; <i>e</i> 15.5; <i>ei</i> 11.4; <i>ea</i> 8   <i>ng</i> 97.8
20. 'ju:flɪks	<i>youflicx/uflicx/-cks/-cs</i>	<i>u</i> 43.9; <i>eu</i> 29.2; <i>yu/ou</i> 24.4   <i>x</i> 89; <i>cks</i> 11
21. tʃʌv	<i>chove</i>	<i>chuv</i> 84.4; <i>chove</i> 6.7; <i>chuve</i> 4.4
22. ʌd	<i>udd</i>	<i>udd</i> 73.3; <i>ud</i> 22.2
23. 'hæzədʒ	<i>hazage/-s-</i>	<i>z</i> 51.3; <i>s</i> 42.5   <i>age</i> 41.3; <i>udge</i> 18.8; <i>uge</i> 10; <i>arge</i> 7.5
24. ɪg'zɜ:stʃə	<i>exersture/-ir/-ur-</i>	<i>ex</i> 85   <i>er</i> 59; <i>ur</i> 40   <i>ture</i> 52.5; <i>ti</i> 31.25
25. ə'peɪzə'neə	<i>apasionaire/app-</i>	<i>a</i> 94.9   <i>p</i> 76.7; <i>pp</i> 23.3   <i>a</i> 55.1; <i>ai</i> 38.5; <i>ay</i> 5.1   <i>ge/i</i> 42.3; <i>si</i> 25.6; <i>ga/o</i> 9; <i>j</i> 7.7; <i>ti</i> 7.7   <i>aire</i> 73.1; <i>air</i> 14.1; <i>are</i> 6.4
26. 'fræki	<i>fracky/-ey</i>	<i>ck</i> 76.5; <i>k</i> 9.1; <i>kk</i> 8   <i>y</i> 64.8; <i>ie</i> 20.5; <i>ey</i> 11.4
27. 'lɔ:kɪʃ	<i>laukish/law/-lor/-loar-</i>	<i>or</i> 60.2; <i>aw</i> 24; <i>au</i> 8   <i>k</i> 95.4   <i>ish</i> 97.7
28. 'sʌktʃʊəl	<i>suctual</i>	<i>c</i> 68.3; <i>ck</i> 20.2   <i>tual</i> 95.5

29. 'gærən <i>garren</i>	rr 78.9; r 18.9   e 52.3; a 20.9; o 20.9
30. 'praʊnə <i>prouner/prow-</i>	ow 72.7; ou 19.3   er 86.4; a 13.6
31. 'kɒʃənət <i>coessionate</i>	c 72.8; k 27.2   sh 86.4; ssi 3.7   -e- 35.8; -i- 25.9; -io- 17.3; -o- 8.6; -u- 6.2   ate 61.7; et 17.3; ut 7.4; at 6.2
32. brəʊð <i>brothe/-oa-</i>	o..e 46.7; ow..(e) 26.5; oa..(e) 11.9; ou..e 5.7   the 58.4; th 32.6
33. dʊp <i>doop</i>	dup 68.5; duvp 11.2; doop 9; doup 5.6
34. 'dʒɪpəl <i>jipple/g/-el/-al</i>	j 69.3; g(i/y) 30.7   pp 90; p 10   le 92; el 3.3
35. wəʊθ <i>woth/-oa-</i>	o..(e) 42; oa..(e) 39.8; ow..(e) 12.5; ou 5.7   th 75; the 25
36. flɪdʒ (verb) <i>flidge/ph-</i>	f 100   dge 100
37. 'haɪdrəm (adjective) <i>hidrum/hy/-am</i>	hy 74.4; high 14.4; hie 5.6; hi 4.4   u 90; a 8.9
38. 'eɪzəprɪn (noun) <i>asaprine/az/-prene</i>	a 82.2; ai 10; ay 6.7   z 50; s 47.8   -a- 55.6; -u- 12.2; -e- 11.1; -ia- 6.7   een 39.5; ine 36.8; ene 23.7
39. sɔ:'keɪʃən (noun, a process) <i>saucation/sor-</i>	or 34.3; aw 30; au 26.7; ore 6.7   c 97.8   a 97.8   tion 86.7; sion 8.9
40. daɪ'lʌkʃuəs (adjective) <i>diluxuous/dy-</i>	di 84.4; dy 11.1   u 97.8   x 48.6; ct 48.6   u 71.1; i 17.8   ous 100
41. 'voɪsəʊklaɪd (noun) <i>voisoclide/-yde</i>	oi 95.6   voiceo- 46.7; voico- 33.3; voiso- 26.7   yde 40; ide/ite 37.8
42. nu:'mɛtʃə (adjective) <i>numeture/neu-</i>	u 68.9; eu 15.6; ew 15.6   ture 52.2; tua 8.9; tuer 5.6
43. 'ziərɒks (noun) <i>zerox/x-</i>	z- 63.3; x- 36.7   e 85.6   -x 97.8
44. ru:'mætsɔɪd (noun) <i>rumatoid/rh/-eu/-tz-</i>	r 68.9; rh 31.1   u 62.2; eu 28.9   tz 58.9; ts 32.2   oid 93.3
45. zaɪ'lɒdʒɪnəs (adjective) <i>zilogenous/x-/zy-/xy/-ge-</i>	z 57.8; x 42.2   y 80; i 15.6   ge/i 80; dge/i 11.1   e 60; i 22.2; u 8.9; io 4.4   ous 84.4; us 4.4; ace 2.2
46. glɪθɪ'rɔɪtɪks (noun) <i>glithiroitics/gly/-ix</i>	gli 56.7; gly 43.3   thi 42.2; thy 26.7; the 22.2   oi 65.6; oy 18.9   ix 58.8; ics 28.9; ex 6.7; icks 5.5
47. krɔɪtɪ'laitəs (noun) <i>croitilitis/chr/-ly/-tus</i>	c 72.7; k 21.2; ch 6.1   oi 47.7; oy 41.1   ti 80; te 11.1   li 73.3; ly 20   us 44.4; is 33.3; ous 22.2
48. θaɪlə'dʒɪlɪks (noun) <i>thilogilics/thy/-la/-le/-gy/-ix</i>	thy 74.4; thi 14.4; thigh 8.9   lo 46.7; li 22.2; la 17.8; le 13.3   gi 73.3; gy 11.1; ji 6.7   ics 44.4; ix 42.2; icks 8.9
49. glə'mætsə'rɪə (noun) <i>glamatsarea/glo-/gle/-tz/-or/-er/-ia</i>	gla 60; glo 13.3; gly 11.1; gle 6.7; glu 6.7   ts 57.8; tz 28.9   -o- 45.6; -e- 27.8; -a- 20   r 93.3; rh 6.7   ia 56.7; ea 22.2; eia 6.7
50. 'dʒɛərɔɪ'setəli:n (noun) <i>geroisetaline/gear-/oice/-te/-to/-ene</i>	ger 76.7; jer 10   roi 36.7; roy 33.3; ro 21.1   s 61.1; c 31.1   ta 62.2; to 15.6; te 15.6   ine 80; ene 13.3

Table 5: Test items, expectations and responses

The experiment was conducted in a single session; thus each participant heard exactly what the others heard. Six participants wrote the first item as *clean*; an independent judge inspected the audio recording and declared that no /n/ was discernible. We assume that these six participants had not yet adjusted themselves to ‘transcribing’ items that they were told were not real. These six responses were excluded from the analysis of the first item.



Participants heard each item twice at least, and, in some cases, three or four times, upon request. We had explained that the experiment was not a test situation; we wished for maximum attention. It was clear from the participants' behaviour that they rehearsed each item silently to themselves in order to reinforce the aural impression before they committed themselves to the task of spelling. In doing so, it is inevitable that they would interpret the item in terms of their own phonological system or accent; for example, it is clear that many – if not all – interpreted the first syllable of 42 as /nju:/ rather than /nu:/. They then, again inevitably, spell out their own mental version of the item. From the qualitative data, it is clear that two main strategies emerged. First, they matched their mental versions with lexical and grammatical elements that were already stored in their minds, eg prefixes like *re-* (2), *be-* (3, 13), *con-* (4), suffixes like *-ing* (19), *-ish* (27), *-ous* (40, 45), *-oid* (44), lexical elements like *view/-vue* (4), and for a few *high* (37). Second, they matched phonemes and phoneme sequences to their store of graphemes. After spelling an item, it is also clear that they sometimes reviewed their attempt if they perceived alternatives and/or improvements or wished to avoid a spelling that already existed in the language. For instance, one participant revised their attempt at 16, from <cliet> to <clite>; another recognized their attempt at 14 as *rougher*, which already exists, and revised it to <rower>, probably not realizing that they had opted for another potential homograph!

## Concluding remarks

The responses in the experiment clearly demonstrate that these young adults do operate a series of systems of orthography for managing the spelling of novel items that conform to regular patterns of word phonology in English. There was not a single nil response in the 2,250 items of data; only very occasionally was there a partial or a confused attempt. Adults certainly do appear to have orthographic potential in their minds which they can readily apply.

Evidence for the general hypotheses has been furnished, except that **G8** was not tested. **G6** was tested by item 22; only 23.3% produced a two letter spelling for this word presented as a noun: 76.7% reproduced the 'three letter rule', principally as <udd>, but also as <udh, ughd>. The vast majority of the 233 phonemic-graphemic hypotheses were proved, but not all of them, and these exceptions require a little thought.

21 /tʃʌv/. Final /v/ is almost invariably spelled <ve> in Standard English; parallel words include *shove*, *dove*, *love*, *glove*, *above*, *cover*, *covet*, *govern*, *oven*, *slovenly* – assumed to be enough guidance. But not to these students, who overwhelmingly chose <chuv>. The basis for their choice appears to be informal spellings like *spiv*, *guv*, *rev(s)*, and probably most importantly *chav*. Hypotheses **G2** and **v** in common words will need reconsideration, certainly for the younger generation.

18 /'kwɒtɪə/. /t/ following a short vowel is almost invariably spelled <tt>. Perhaps this item sounded rather like a word of classical origin, or indeed a foreign loan word, in which the doubled consonant is not expected. There were three responses that reflected French spelling, including <coitieur>.

8 /kaɪ/. *Kai* was offered 53.6%, no doubt with certain celebrities in mind to the younger generation! <kih> seems as reasonable as the expected <kie, kye>; the /aɪ/ hypothesis needs reconsideration.

Initial /k/, other than for 8 /kaɪ/. The **c** hypotheses need reconsideration, since <k> appeared before <l> (1, 16), before <o> (31) and before <r> (47) quite extensively. Perhaps again, a difference between the generations, with the younger more disposed to <k> as a result of exposure to popular brand names.

Space does not permit consideration of other exceptions and the details of alternatives; they will have to await another opportunity. But what we have established is that just as it is possible to investigate and describe the phonological systems that a person operates, it is also possible to investigate and describe the orthographical systems, including the paradigm shift in spelling what are perceived to belong to the technical register. The study therefore completes Aitchison's *Words in the Mind*, contributes a much more comprehensive set of phonemic-graphemic correspondences than van Berkel (2005) for second language research, extends Kress's study into young adulthood, matching Halliday's *Learning how to Mean*, and provides evidence of literate competence in any culture.

## References

- Jean Aitchison.** 2003. *Words in the Mind*. 3<sup>rd</sup> ed. Blackwell: Oxford.
- Ans van Berkel.** 2005. The role of phonological strategy in learning to spell English as a second language. In Cook, V. & Bassetti, B. *Second Language Writing Systems*. Multilingual Matters: Clevedon.
- Adam Brown.** 2005. *Sounds, symbols and spellings*. McGraw-Hill: Singapore.
- Edward Carney.** 1994. *A Survey of English spelling*. Routledge: London.
- Vivian J. Cook.** 2004. *Accommodating broccoli in the cemetery; or, Why can't anyone spell?* Profile: London.
- Florian Colmas.** 1989. *The Writing Systems of the world*. Blackwell: Oxford.
- Bruce Derwing & Maureen L. Dow.** 1987. Orthography as a Variable in Psycholinguistic Experiments. In Philip A. Luelsdorff. 1987.
- Bruce Derwing, Tom M.S. Priestly & Bernard L. Rochet.** 1987. In Philip A. Luelsdorff. 1987.
- Norman Fairclough.** 2007. Global capitalism and change in Higher Education: dialectics of language and practice, technology, ideology. BAAL conference: Edinburgh
- Jost Gippert, Nikolaus P. Himmelmann & Ulrike Mosel (eds).** 2006. *Essentials of Language Documentation*. Mouton de Gruyter: Berlin.
- Lenore A. Grenoble & Lindsay J. Whaley.** 2006. *Saving Languages*. Cambridge University Press: Cambridge.
- Leanne Hinton & Ken Hale (eds).** 2001. Academic Press: San Diego.
- Gunter Kress.** 2000. *Early Spelling: between convention and creativity*. Routledge: London.
- Robert Lado.** 1957. *Linguistics across cultures; Applied Linguistics for Language Teachers*. University of Michigan Press: Ann Arbor.
- Philip A. Luelsdorff (ed).** 1987. *Orthography and Phonology*. John Benjamins Publishing Co: Amsterdam.
- Lena Olausson & Catherine Sangster.** 2006. *Oxford BBC Guide to Pronunciation*. Oxford University Press: Oxford.
- Philomena Ott.** 2006. *How to manage spelling successfully*. Routledge: London.
- Kenneth L. Pike.** 1947. *Phonemics: a technique for reducing languages to writing*. University of Michigan Press: Ann Arbor.
- Andrew G. Rollings.** 2004. *The spelling patterns of English*. Lincom Europa: München.
- Mark Sebba.** 2007. *Spelling and society: the culture and politics of orthography around the world*. Cambridge University Press: Cambridge.
- Donald G. Scragg.** 1974. *A history of English spelling*. Manchester University Press: Manchester.
- Petr Sgall.** 1987. Towards a Theory of Phonemic Orthography' in Luelsdorff, P. A. 1987.
- William A. Smalley (ed).** 1963. *Orthography Studies*. United Bible Societies: London.
- Paul Tench.** 2007. *Illustrations of the IPA: Tera*. In *Journal of the International phonetic Association*, 37/2, 2007, pp227-234.
- Richard L. Venezky.** 1970. *The Structure of English orthography*. Mouton: The Hague.
- Axel Wijk.** 1966. *Rules of pronunciation for the English language: an account of the relationship between English spelling and pronunciation*. Oxford University Press: London.
- W. Winter.** 1983. Tradition and innovation in alphabet making. In Florian Coulmas & Konrad Ehlich. *Writing in focus*. Mouton: Berlin.
- Writing in focus*. Mouton: Berlin.