Hickey, T. M. & Stenson, N. (2011): Irish orthography: what do teachers and learners need to know about it, and why? Language, Culture and Curriculum, 24, 1, 23-46
http://dx.doi.org/10.1080/07908318.2010.527347

Irish orthography: what do teachers and learners need to know about it and why?

Tina Hickey[a*] and Nancy Stenson[b]

[a]School of Psychology, University College Dublin, Belfield, Dublin 4, Republic of Ireland;
[b]Institute of Linguistics, University of Minnesota, Minneapolis, MN, USA

Irish has significant State support, but lacks a research base to support the teaching of Irish reading. Current approaches to teaching Irish reading are presented, and outcomes summarised. Issues of consistency and complexity in Irish orthography are discussed in light of an analysis of a corpus of early reader texts, and the formulation of rules for discriminating between words which are regular by letter-sound and grapheme-sound rules is outlined. While the most frequent words show a high level of regularity, underlying rules are very complex. The need to target decoding skills early is discussed. Recommendations regarding the teaching of aspects of Irish orthography are presented.

Keywords: Irish; orthography; heritage language teaching; teacher training; teaching methods; beginning readers

1. Introduction

Despite a recent burgeoning of research interest in the literacy problems of children for whom English is an additional language (c.f. Harrison & Krol, 2007), the needs of learners (both children and adults) learning their L2 literacy in a minority language have received relatively little attention. Lesser-used languages suffer from poor resourcing, as well as practical difficulties such as limited materials production. This paper aims to consider some of these gaps in the case of one such language, Irish.[1] Irish is designated by the EU both as a 'regional, minority language' and, more recently, as an official language of the EU. It is the 'first official language' of the Republic of Ireland, although most citizens are English-speaking. While Irish has significant State support in terms of an established place in the primary school curriculum in the Republic of Ireland, it lacks a body of research on the teaching and learning of literacy in the language to support it. The fundamental research question to be addressed here regards the difficulties this system poses for learners. Following necessary background information in the remainder of this section, we briefly review the current teaching of Irish literacy. We then examine the Irish spelling system in its own right, following the corpus study methodology previously used for English. This study will show a high degree of systematicity to Irish spelling, but this systematicity is based on very complex rules, leading to a discussion of the need to consider complexity alongside regularity. Finally, the practical challenges posed by this orthography for beginning readers will be discussed, concluding with recommendations for the explicit teaching of Irish orthographic patterns.

*Corresponding author. Email: tina.hickey@ucd.ie

1.1. Who reads in Irish?

Most children in the Republic of Ireland are native English speakers, and begin to learn Irish as a second language at school entry, aged 4–5 years. The Department of Education and Science (DES) (Revised Curriculum, 1999) recommends that children attending English-medium schools (about 95%) should learn initial literacy in English, and that after three years of schooling (by which time they are expected to have a good foundation in English reading and oral Irish) they should be introduced to reading in Irish as L2. Conversely, Irish native speakers and those English-speakers who attend Irish-medium schools (about 5% of the primary sector) learn their initial literacy in Irish and begin English reading later (schools vary regarding when).

Other children learning Irish come from varied language backgrounds, a result of immigration during the economic boom since the 1990s. Many of these children learn Irish as their L3. Only children who have already received most of their primary education outside of the state or those with a diagnosed reading or intellectual disability are exempt from learning Irish. Overall, then, most children in the Republic of Ireland are reading or learning to read in both English and Irish by age 9, and have been taught Irish as either L2 or L1 from school entry at age 4 (Harris, Forde, Archer, Nic Fhearaile, & O'Gorman, 2006; Harris, 2007). Outside Ireland, some children of the Irish diaspora also attend community language classes, particularly in the UK (http://www.colaiste-na-ngael.com/classes.html), and to a lesser extent in the USA and Australia.

A growing number of adults also learn Irish, both in Ireland and internationally. Many courses are offered within Ireland to adult learners wishing to improve their school Irish as well as to visitors from abroad. Recent decades have also seen a worldwide flourishing of Irish courses for adults, both credit-bearing university-level courses and informal study-groups. Although Irish is mostly studied in anglophone countries, Irish language learners are also found in Germany, Russia, Finland, Norway, Sweden, France, Poland, the Czech Republic, and Japan, among others. While they tend to have high levels of L1 reading fluency and often good knowledge of other languages as well, these learners bring different levels of proficiency in Irish and in language learning to the task of Irish reading. That task is briefly presented in the outline of Irish orthography below.

1.2. Irish orthography

The current Irish alphabet comprises 5 vowels and 13 consonants, shown in (1):

(1) *a b c d e f g h i l m n o p r s t u*

This alphabet represents approximately 50 basic Irish sounds. The letters <j, k, q, v, w, x, y, z> are also used in loan words. The five spoken vowels of Irish can be either short or long, giving 10 sounds, and in the simplest cases a length mark (síneadh fada) is placed above the vowel (á, é, í, ó, ú) to indicate its lengthening. Compare, for example, the short vowel in < bá> /ba/ 'cows' vs. the lengthened vowel in < b**á** > /ba:/ 'understanding'. Digraphs like ao, ae, eo also signal length, and vowel lengthening (or diphthongisation) occurs in closed syllables before nn and ll (and, in some dialects, rr and m as well). Short vowels on which no stress is placed are pronounced as schwa /ə/ but this is not represented orthographically.

While the script now used for written Irish is familiar to readers of English, the orthographic conventions differ significantly. One feature of particular relevance to a discussion of Irish orthography is the contrast between slender (palatalised) and broad (non-palatalised or velarised) consonant forms. Orthographically, the quality of the spoken consonant is indicated by adjacent vowels, i.e. by preceding/succeeding the consonant symbol with a slender (*i or e*) or broad (*a, o, or u*) vowel. Thus, some vowels in Irish orthography are used solely to signal consonant quality, and are not themselves pronounced. Examples (2)–(5) show minimal pairs of broad and slender consonants, with the palatalised consonant marked with a following /ʹ/ according to the norms of Irish phonetic transcription:

(2)	*buí*	/bi:/	'yellow'	*bí*	/bʹi:/	'be'
(3)	*bó*	/bo:/	'cow'	*beo*	/bʹo:/	'alive'
(4)	*bád*	/ba:d/	'boat'	*báid*	/ba:dʹ/	'boats'
(5)	*teas*	/tʹas/	'heat'	*tais*	/tasʹ/	'damp'

Note that the 'i' in báid and tais in examples (4) and (5) should not be syllabified as /id/ or /is/ but signals palatalisation of the following /dʹ/ and /ʹ/ ʃ(=/sʹ/).

The use of vowel symbols to signal consonant quality is expressed by a spelling rule of vowel harmony in Irish orthography: *leathan le leathan, caol le c*aol 'broad [vowel] with broad [consonant], slender with slender', a rule which is taught explicitly in schools. This rule explains a sub-group of unpronounced vowels which are added to achieve a balance in vowel quality on either side of consonants, thereby signalling consonant quality unambigu- ously. Example (6) contrasts two words whose final syllables sound and mean the same, but are spelled differently, with an additional /e/ in the second case, simply because the preced- ing consonant and pronounced vowel are slender:

(6)	*Glanann*	/glanən/	and	*Briseann*	/bʹrʹisʹən/

The Irish system of consonant mutation is also uniquely reflected in its orthography. An initial consonant may undergo pronunciation changes that signal various morphological processes such as gender/number/case, possession, adjective agreement, and tense. The orthographic effects of the two commonest mutations, lenition and eclipsis, are illustrated in (7):

	Citation form	Lenition	Eclipsis
(7)	*bord*	*bhord*	*mbord*
	cat	*chat*	*gcat*
	fraoch	*fhraoch*	*bhfraoch*

In all, nine consonants are affected by lenition and seven by eclipsis. The result of these mutations is that word forms are highly unstable, with frequent word-onset changes producing complex word-initial clusters, as in *bhruach* /vru@x/, *ndroichead* /nrihəd/, *d'fhreastal* /dəʳʹastəl/, *tsráid* /trʹa:dʹ/.

To summarise this brief overview, modern Irish uses 18 letters of the Latin alphabet, and a length diacritic on vowels, to represent at least 50 basic sounds, varying among three major dialects. Its orthography distinguishes between pairs of slender (palatalised) and broad (non-palatalised or velarised) consonants,[2] using unpronounced vowels before/ after consonants to indicate consonant quality. Initial mutations result in considerable variability in word forms of nouns, verbs and adjectives, and present complex word-initial clus-

ters of consonants. Standardisation and spelling reform have been criticised as incomplete (Ó Murchú, 1977; Ó Sé, 1990), and also as imposing a variety that does not represent any dialect consistently (Mac Cárthaigh, 2006). Indeed, the government in 2010 initiated a review of the written standard for Irish, with the goals of simplification and increasing internal consistency (http://www.pobail.ie/ie/Lar-AonadAistriuchain/Scoipanphroisis).

2. Current teaching of Irish reading

The current curriculum of the Department of Education in the Republic of Ireland (revised in 1999) includes specific guidelines on the sequencing of and approach to the teaching of reading in Irish and English. Recognising that most children begin reading with English, it acknowledges that 'a certain degree of skill transference may occur' from English to Irish reading (Department of Education and Science [DES], 1999, pp. 5–6), and advises teachers to engage in an analysis of similarities and differences between Irish and English sound-symbol correspondences. However, there appear to be gaps between the aspiration and implementation of the Revised Curriculum with regard to the teaching of Irish reading. An evaluation of the teaching of Irish (DES, 2008) identified best practice in only 8% of observed classes. The DES evaluation notes further that word analysis skills were not formally taught in 15% of classes observed, resulting in pupils who had difficulty reading common words, and who could not identify words already learned orally. Hickey (2007) and Ó Faoileáin (2006) point to an even greater reluctance to teach graphophonic correspondences than these figures suggest, and a widespread tendency to rely on a 'look-and-say' approach. Where sound-symbol correspondences are taught, these tend to be only the most straightforward ones, e.g. *í* = /i:/. Other recent research (Harris et al., 2006) has also noted that teachers in English-medium schools tend to rely on a class textbook for Irish reading, with relatively little exploration of the other Irish books available. Harris et al. found that immersion schools do expose pupils to a wider variety of texts and genres in Irish, but their reading in Irish tends to be restricted to school, and teachers report difficulties engaging even immersion school pupils in leisure reading in Irish. Thus, it appears that many pupils receive limited instruction in Irish word analysis and recognition, as well as limited opportunities to practise reading in Irish. This may be because relatively few materials are available for teachers in English-medium schools to present the grapheme–phoneme rules of Irish to beginners, and there is low awareness even of those available.

Another concern is the time-limited training in the teaching of Irish reading given at pre- and in-service level to enable teachers to engage in the recommended comparisons between English and Irish spelling. Given the infrequency of explicit discussion of Irish orthography with pupils in the majority of classrooms or lack of detailed methods of teaching Irish sound-symbol correspondences as noted by the DES (2008) evaluation, it is unsurprising to find that the whole-word approach figures largely in teaching young children to read in Irish. The DES evaluation and observations from teacher trainers indicate that Irish reading is apparently viewed as a secondary part of the Irish class, rather than being approached as part of the biliteracy development in English and Irish. Lack of orthographic analysis during pre-service education results in a reliance on sight-word training in both English and Irish reading and minimal use of phonics for graphemes representing similar sounds in both languages, but with little exploration of the differences between the orthographies or of the regularities unique to Irish.

2.1. Outcomes

The outcome with regard to reading is evident in evaluations of Irish language achievement. According to the DES (2008) report, 'in approximately one-third of classes, pupils had significant gaps in their skills of word recognition and reading comprehension' (p. 60), while 11% had significant difficulties with word recognition in Irish. Harris et al. (2006) found that 21% of parents said that their children in the final year of primary school (aged 11–12) had problems reading Irish, whereas only 8% noted a problem with English reading at this age. Four-fifths (80%) of teachers surveyed also claimed that standards in Irish reading have declined significantly since the 1988 assessment. Interestingly, schools were found to differ more on Irish reading scores than on English reading and mathematics scores. For younger children, a study of Irish reading among Grade 2 children in English-medium schools (Hickey, 2005) showed that many had difficulty decoding even high-frequency Irish words, and words with consistent grapheme–phoneme correspondences. The children's miscues showed only partial analysis, depending on first letters and overlooking even regular grapheme–phoneme relationships, and also showed intrusion of English grapheme–phoneme relationships into sight-reading.

Parsons and Lyddy (2009) compared the English and Irish reading of 8-year-old children in Irish-medium and English-medium schools, as well as a group of L1 Irish speakers in a Gaeltacht (Irish-speaking community). They found that the least proficient Irish readers gave a large number of 'failure to read' responses, i.e. refused to attempt a word in Irish, and made as many real-word substitution errors (of Irish or English words) as non-word errors. The most proficient readers in this sample, on the other hand, rarely refused to attempt a word, and their errors fell mainly in the non-word category, while English word substitution and Irish word substitution were less frequent. Parsons and Lyddy interpret this as showing that successful Irish readers have worked out the grapheme–phoneme correspondences in Irish and are using a phonological reading strategy, which supports their attempts at unknown words, resulting more often in a non-word than a whole word substitution. The unsuccessful readers, on the other hand, appear to have difficulty with grapheme–phoneme correspondences, resulting in a high rate of refusals to attempt an unknown word, and a higher production of substitutions of real words than the more successful readers, indicating only partial analysis at best or a whole-word strategy rather than graphophonic analysis.

In contrast to child learners of Irish as a second language, who are still honing their native language reading skills as they begin their L2, adults who come to the study of Irish do so with fully developed L1 literacy skills, usually in English, but sometimes in another native language. Perhaps for this reason, and because of biases in the Euro-American educational system towards written learning, more courses for adults than for children focus on written language from the outset. It is therefore inevitable that learners bring their knowledge of English orthography to the learning of Irish, and tend to transfer the orthographic conventions of the language in which they are already literate. A certain amount of re-education is therefore necessary to help learners link the pronunciations of Irish vocabulary to their spellings. However, as in the teaching of children described above, adult courses and texts typically lack all but the most cursory attention to explicit teaching of grapheme–phoneme correspondences (c.f. Ó Siadhail, 1995; Sheils & Ó Sé, 2004). The assumption underlying most teaching materials for adults seems to be that literacy in L1 will transfer without explicit teaching to the L2, thereby missing an opportunity for promoting more effective learning of reading in the second language.

2.2. Alphabetic writing systems

Ziegler and Goswami (2005) defined reading as the process of matching visual symbols to units of sound, but noted that languages vary significantly in the consistency with which their orthography represents phonology. Thus, children learning to read alphabetic writing systems must discern the nature of phonetic representation in a particular orthography (Juel, 2006). As a result of this variation in how different orthographies represent the sounds of their languages, developmental differences have been documented in children's reading strategies and rates of reading difficulties in different languages. A number of studies (Ellis et al., 2004; Goswami, Gombert, & de Barrera, 1998; Seymour, Aro, & Erskine, 2003) have compared the difficulty of learning to read in languages such as English, French, German, Albanian, Turkish, Spanish, Italian, and Welsh, all of which use the same script, but whose orthographies adopt different conventions to represent the sounds of the language. The results show that the level of consistency in the orthography significantly impacts on the rate of children's progress in early reading. Seymour et al. attributed differences in children's progress to the orthographic depth of the language they were learning (the level of consistency with which each letter/group of letters represents sounds in the language) and to syllabic complexity.

Thus, orthographic consistency predicts the ease and speed of reading development, with more consistent languages such as German, Finnish, and Spanish showing earlier accurate reading than languages that have inconsistencies in either reading (Danish) or spelling (French), which in turn showed better early performance than languages with inconsistencies in both reading and spelling (English). Ziegler and Goswami (2005) argue that the orthography must be taken into account in the teaching of reading in a particular language, since learners of consistent languages benefit from a focus on 'the small psycholinguistic grain size of the phoneme' (p. 19), whereas learners of an inconsistent orthography need to learn correspondences for larger units such as syllables or whole words. Thus, the level of consistency in the grapheme–phoneme correspondences in a language should influence the early teaching of reading in that language, in orienting the beginning reader to the types of units which will offer the greatest gain. They note that children learning to read English show stronger influences from whole-word phonology, while children learning to read orthographically consistent languages such as German or Spanish rely heavily on grapheme–phoneme recoding strategies.

The issue of adapting the teaching of reading to the orthography is also relevant when two orthographies are learned consecutively. Researchers in L2 reading (Bernhardt, 2003; Grabe & Stoller, 2002; Koda, 2004) have begun to systematise conclusions from L1 reading research as the basis for recommending improvements in L2 reading instruction. It is notable that word recognition fluency and efficient word recognition (Grabe, 2004; Pikulski & Chard, 2005; National Institute of Child Health and Human Development [NICHD], 2000) are highlighted as critical components of reading both in L1 and L2. Here it is argued that the teaching of Irish reading needs to be informed by research leading to a greater understanding of its orthography. This would enable teachers to evaluate the benefits of explicit teaching of symbol-sound correspondences to improve decoding by early readers. A considerable body of research has indicated that sight word reading alone is less helpful than explicit teaching of grapheme – phoneme relationships (Ehri et al., 2001; NICHD, 2000; O'Sullivan & Goosney,2007; Pikulski & Chard, 2005). Other research has found that early flawed strategies based on partial word analyses are widespread in beginning readers and can be hard to repair once established (Harrison, 1998). The tendency among Irish teachers to rely on the 'look and say' approach in Irish reading (Ó Faoileáin, 2008), with a reluctance to teach graphophonic correspondences outside of those which overlap with English, may promote the partial analysis evident in learners' miscues even in the most regular grapheme – phoneme relation- ships (Hickey, 2005). This points to a need to focus explicitly on teaching methods that promote analysis of grapheme – phoneme correspondences, particularly of the most frequent words, in order to improve word recognition. To achieve this

end, teachers need guidance and instruction in order to understand the features of the Irish orthographic system and teach sound-symbol correspondences explicitly.

The first step is to determine the nature and degree of regularity in Irish orthography. Such analysis is assisted by the development of corpora in the language. Goswami (2006) has noted the contribution of corpus analysis to the fuller understanding of English orthography in recent years. She points to the need for the teaching of reading in a particular language to be informed by the script and orthography of that language, rather than assuming that the approach which works for one language will be equally effective for all. The issue of a'fit' between the teaching approach used and the orthography of the language being learned is also discussed in Goswami et al. (1998), who noted that sight-word training tends to play a greater role in teaching inconsistent orthographies than in more regular ortho- graphies, which can rely more on teaching sound-symbol correspondences. We argue here that although the complexities of Irish orthography are numerous and can cause considerable difficulty to learners (especially adults already literate in English), an over-reliance on sight- word training for children learning Irish as a result of insufficient understanding of how the orthography works fails to capitalise on the regularities that do exist in the system, and leaves unexplained the features of the orthography that differ significantly from English.

We attempt below to assess objectively the challenge involved in Irish reading and to base this on the kind of language confronting the language learner, using a corpus of children's books in Irish. The first question to be asked concerns just how regular the Irish spelling system is. We address this question with a corpus study of early Irish readers of the sort most likely to be encountered by beginning readers. This analysis addresses a gap in the literature on Irish in analysing the regularity and complexity of Irish orthography. The approach taken is similar to the analysis of English by Stuart, Dixon, Masterson, and Gray (2003) who developed a corpus of children's books in English aimed at early readers. The database is first described briefly, with a discussion of the characteristics of early vocabulary, particularly with regard to regular and irregular spelling patterns, following Stuart et al. This is followed by a brief consideration of the implications of prior English literacy for the teaching and learning of Irish reading.

3. The Early Reader Corpus in Irish: methodology

The assembly of an Irish corpus of books aimed at adults began in recent years (Institiúid Teangeolaíochta Éireann, 2003). A corpus of Irish children's books ('Corpas na Leabhar Gaeilge do Pháistí', henceforth referred to as CLGP) assembled by the first author com- prises books aimed at preschoolers and children in primary school (up to about age 14); this includes fiction and non-fiction, but excludes textbooks. As far as possible, texts

were corrected to bring them into line with the final book (electronic texts were frequently not the final version of the text published). Where a glossary of terms was provided in books (a recent, but fairly rare occurrence), the Irish words were retained in the text but the English gloss was excluded. The text aimed at adults (e.g. the cover 'blurb') and publishing details (printer, page numbers, etc.) were excluded. Series from Northern Ireland (publisher Áisaonad) were included and the more recent reading scheme, Séideán Sí, developed by An Gúm (various years) for Irish-medium schools has also been added, up to Rang III (third grade). Some of these books are translations of English or other language originals while others are published only in Irish. The total CLGP includes books to be read aloud to young children, early readers and stories, and longer books aimed at mid-teenage readers. In all, 386 books made up the corpus, which comprises 761,779 word tokens, and 27,816 word types.

This analysis focuses on the most frequent words as word types, rather than grouping them under word families or lemmas. Gardner (2004) cites a body of research showing that young readers' ability to make word family connections varies significantly depending on whether children are reading in their L1 or L2, their general reading skills, the size of their existing vocabulary, and instructional methods. We make the conservative assumption that the majority of pupils learning to read Irish as an L2 in a classroom setting are likely to treat words as separate entities initially, rather than linking them to word families, especially as some forms of words are fairly unstable due to initial mutations, case and number inflections of nouns, inflection of prepositions for person and inflection of verbs for tense and person. These result in a range of word forms which may require high exposure before they are linked by readers with limited competence in the language. Since Stuart et al.'s (2003) analysis also does not link items under word families or lemmas, the decision to treat separate forms of words as individual types here has the added advantage of comparability with their analysis. This is particularly important in exploring regularity, as a base form of a word may be regular, but its more frequent variant form may be irregular by the letter-symbol or grapheme–phoneme rules, and it would therefore underestimate the challenge of the word to count it as regular in all cases.

The decision to focus on a sub-set of the corpus consisting of the books aimed at early readers also follows Stuart et al. (2003). Their Early Reader Corpus aimed at children in Key Stage 1 of the British educational system included 19 reading schemes and series used in the surveyed schools, showing the choice available in the teaching of English, an extremely well-resourced language. Their corpus contained 9748 words (types) and 268,028 individual occurrences of these words (tokens). The comparable array of books aimed at early independent readers in Irish (a subset of the CLGP) comprises in total 186 books in series from various publishers which were judged to lie within the 'early reader' category, as aimed at young independent readers of Irish. This sub-corpus of books is small by comparison with that assembled by Stuart et al., with 3445 words (types) and 43,052 occurrences of these (tokens), but Meara (2005, p. 32) has noted that, while there has been concern about which lists provide the best standard for a particular language, corpus size is really only an issue at low levels of frequency. He argues that different lists in a language tend to agree broadly about the most frequent words, despite large differences in size. Thus, while the corpus of early readers in Irish is small by corpus standards, it is considered sufficiently large to allow comparison with Stuart et al.'s Early Reader Corpus in English, and as Table 1 indicates, it bears some very interesting similarities to it in terms of distribution of the most frequent types and tokens.

Stuart et al. (2003, p. 588) noted that the most striking characteristic of the vocabulary in the English early reader database was the extent to which it is skewed towards lowerfrequencies, with over half (51%) of the word types appearing only once or twice. Similarly, in the Irish Early Reader database 53.7% of words occur just once or twice in total. Thus, even in these texts aimed at young learners where there is an attempt at vocabulary control, about half of the word types encountered will be seen only once or twice. As Stuart et al. (2003) point out, the large number of low frequency words in these corpora aimed at children starting to read presents a learning problem, given the evidence from Stuart, Masterson, and Dixon (2000) that 5-year-old children need more than 36 exposures to a word in a text in order to store and facilitate recognition of that word. In fact, 97.7% of the word types in the English Early Reader Corpus and 94.6% of the Irish Early Reader Corpus appeared less than 36 times across this category of books.

At the other end of the spectrum, Table 1 shows that the 100 most frequent word types account for over half of the tokens in each corpus; Table 2 gives further details. In the Irish Early Reader Corpus the top 500 words account for 78.8% of all tokens in the entire corpus, similar to the proportion covered by the top 500 in the English Early Reader Corpus (76%). The top 1000 words account for 87.96% (Stuart et al. [2003] do not provide the 1000 word coverage for their corpus).

Table 1. Comparison of coverage in English and Irish Early Reader corpora.

	English		Irish	
	% of types	% of tokens	% of types	% of tokens
100 most frequent words	1	54.1	2.9	54.9
500 most frequent words	5.1	76.3	14.5	78.8
Types appearing once/twice only	51	2.4	53.7	5.2
Types appearing <36 times	97.7	n/a	94.6	n/a

Hu and Nation (2000) estimate that readers need to know 98% of the word tokens in a text to read it independently for meaning, without consulting an adult or dictionary. Even at this level, they estimate that the reader will meet one unknown word in about every 50 words. Nation (2006) estimates that 92% coverage would result in one unknown word in about every 15 words. Since most of the young L2 readers in the early years will not have an Irish (L2) vocabulary even approaching 1000 types, the challenge is considerable, as they require significant support in terms of preparation to deal with unknown vocabulary, and pre-reading and during-reading strategy deployment to maximise comprehension. Nevertheless, we claim, Table 2 also shows emphatically the value of a more systematic approach to establishing a firm foundation in even just 300 word types, which would acquaint children with 71% of the words they encounter, offering teachers a more attainable goal for the earliest classes, and a basis on which to anchor decoding in the new orthography. This, in conjunction with the teaching of grapheme – phoneme correspondences and strategies like extracting word meaning from context and pictorial clues, would offer children in the earlier grades a good return on investment in learning the most frequent words.

Table 2. Coverage of tokens by the most frequent words in the Irish Early Reader Corpus.

Irish Early Readers	Tokens	% total tokens
Top 100 word types	23,678	54.99
101–200	4444	10.31
201–300	2623	6.11
301–400	1814	4.2
401–500	1390	3.23
Top 500 word types	33,949	78.85
Top 1000 word types	37,867	87.96

In order to provide the background analysis (hitherto lacking for Irish) which would underpin such a teaching approach, the most frequent words in the Irish Early Reader Corpus are examined below following the analyses by Stuart et al., with some consideration of the incidence of content vs. function words, before a fuller discussion of regularity.

4. Results

4.1. Content and function words

The corpus was analysed to ascertain the distribution of content words (e.g. nouns, verbs, adjectives) and function words (e.g. prepositions, pronouns, auxiliaries, conjunctions). Some forms can be either content or function words depending on the context, and were classified according to their dominant use in concordance analysis. For example, trí might represent either the numeral 'three' or the preposition 'through'. Only 4% of the occurrences of trí in the database were prepositional, so it is treated as a content word here. Stuart et al.'s definition of function words included the usual auxiliaries, pronouns, articles, conjunctions, prepositions and some adverbs, but also common 'light' verbs like be, give, put, go, come, said, say, etc. Applying this definition, we find that the top 100 words of the Irish Early Reader Corpus include 73 function words, compared with 85 in Stuart et al.'s English corpus. Thus, in both Irish and English Early Readers, three-quarters to over four-fifths of the most frequent word types are function words. Applying a stricter definition of function words than Stuart et al. (one that eliminates the non-auxiliary uses of light verbs), it was found that there were still 64 function words in the Irish top 100 and, indeed, the 21 most common Irish Early Reader words are function words by these criteria. Some function words are shown in (8):

(8)	*an*	*a*	*ach*	*ag*	*agam*	*agat*	*agus*	*ann*	*anois*	*anseo*	*ansin*	*ar*	*as*
	the	Poss part	but	at	at-me	at-you	and	there/in-3masc	now	here	then	on	out-of

The most frequent word, *an*, represents several distinct morphemes, which have the same spelling and pronunciation but are unrelated. Most frequently, *an* functions as the singular definite article. *An* is also the particle which precedes verbs (or replaces the copula is) to indicate that the sentence is a question. Finally, stressed *an* can be prefixed to nouns or adjectives, to give intensive force. An occurs 1387 times in the Early Reader Corpus, and 88% of its occurrences are as the definite article (*an cailín* 'the girl', *an leaba* 'the bed'). Only 9% of its uses are as a pre-verbal particle (*an bhfuil* 'is?' *an dtiocfaidh?* 'will come?') or as interrogative copula before a noun or pronoun (*an tusa* 'is-it you?', *an libhse*, 'is [X] yours?'). The remaining 3% of an tokens are intensive, preceding either an adjective (*an-sásta* 'very pleased', *an-te* 'very hot') or a noun (*an-lá* 'a great day', *an-sásamh* 'much satisfaction').

The remaining content words in the top 100 words in the Irish Early Readers included:

(9)	*am*	*beag*	*béar*	*breá*	*buí*	*cat*	*chonaic*	*daidí*	*deas*	*fear*	*lá*	*leaba*	*maith*
	time	little	bear	fine	yellow	cat	saw	daddy	nice	man	day	bed	good

The most frequent content word, *maith* 'good', appears 170 times across 60 texts in the early reader database, placing it 22nd in the frequency ranking. Features of these words are discussed further below.

4.2. Regularity

A central aim of Stuart et al. was to consider the degree of regularity in the most frequent words in their corpus. They found that 52 of the 100 most frequent words in their English corpus were regular according to rules of grapheme–phoneme correspondence, noting that this is, in fact, a higher proportion than is popularly believed with regard to the words encountered by learners of English literacy. Of these, 27 English words were regular by letter-sound rules (e.g. and, but, can, help, frog), while 25 were regular by grapheme–phoneme rules (e.g. with, out, down, then, like, see, very, house, children, saw). Here we examine the level of regularity in the 100 most frequent words in the Irish Early Readers' Corpus, summarised in the appendix. Prior to this analysis, however, a fundamental issue needs to be addressed. In the absence of an educational research literature on the Irish orthographic system, we cannot assume an established set of rules, as did Stuart et al. for English, and it is important to clarify the grounds on which we based our decisions of regularity. While we do not expect unanimous agreement with the rules developed, we hope that initiating discussion of these issues will help those teaching Irish reading. We determined regularity according to the criteria detailed below.

4.3. Determining regularity in Irish: rules proposed

(1) Vowels and diphthongs

We assume a standard vowel system of five short and five long vowels and the diphthongs /ai/ and /au/, with approximate IPA values allowing for minor variations from dialect to dialect. Short /o/, say, may be pronounced differently in Munster and Connacht but as long as it is consistent across words it is treated as regular. Only when pronunciation differs within a dialect from the usual pronunciation for that dialect is it treated as irregular.

(2) Unstressed vowels

In unstressed syllables, reduction to schwa is the norm outside of Ulster, and is treated as regular non-initially. However, as a rule, monosyllables in Irish are assumed to bear stress. Therefore, clitics such as prepositions and verbal proclitic particles, which are inherently unstressed, are treated as irregular on the grounds that beginning learners of Irish cannot be expected to know which monosyllabic words are stressed and which are not.

(3) Palatalisation vs. velarisation

Recognising regularity in Irish requires acknowledgement of the role played by vowels in signalling consonant quality. Regular consonants include digraph sequences of consonant symbols plus vowels signalling palatalisation or velarisation (distinguishing slender from broad consonants) adjacent to a (phonetic) vowel of different quality (i.e. a palatalised (slender) consonant next to a back vowel or a velarised (broad) consonant next to a front vowel). That is, sequences such as Ce before a phonetic back vowel or iC following one, Ca or aC, oC adjacent to a pronounced front vowel, etc., are treated as regular according

to phoneme–grapheme rules. By this assumption, words in (a) of Table 3 (below) are counted as regular.

(4) Initial mutations
Consonant mutations apply both historically within words and synchronically in the initial position under specific lexical/grammatical conditions. They are marked consistently by digraphs, treated here as regular graphemes. The lenition mutation is marked orthographically by an h following the affected consonant, and eclipsis is marked by a symbol representing the mutated pronunciation before the consonant of the citation form. Examples are found in (b) of Table 3 below.

(5) Vowel digraphs
Common vowel digraphs are likewise treated as regular: e.g. ao (= /e:/ or /i:/ depending on dialect), ei (= /e/), eo (= /o:/) and ae (= /e:/). Even though in some cases, most notably ao, the pronunciation varies from dialect to dialect, it is consistent within a given area. There is an internal consistency in treating these digraphs representing (mostly) long vowels as the norm, with the rarer case of a short vowel in words such as seo being treated as the exception.[3]

(6) Vowel rules
Vowels that undergo certain broadly applicable rules are also considered regular. For example, words like *ann* 'there', *am* 'time', *poll* 'hole', *carr* 'car' are counted as regular, although not pronounced with the short vowels indicated by their spelling but rather with long vowels or diphthongs (varies by dialect), consistently triggered by the following nn, ll, rr, or m in a closed syllable.

(7) Epenthesis
Similarly, the epenthetic, but unwritten vowel of words such as gorm /gorə̯m/ 'blue' and *banbh* /banə̯v/ 'piglet' is not treated as an irregularity, because it occurs quite predictably between a sonorant and voiced consonant.

(8) Double consonants
Both single and double sonorants (nn, ll, rr) are treated as regular, although the distinct pronunciations for single and double graphemes found in some regions are neutralised in others. This process is regular, however, in the regions where it takes place, and is therefore a consistent phoneme–grapheme pattern (although details vary regionally). It will not, however, be treated as a letter-sound regularity. See Rule 6 above for the effect of these consonants on vowels.

(9) Treatment of dialect variation
When dialects vary in the pronunciation of words in ways that lead to irregularities in some regions but not others, we have treated a form as regular if in any dialect the phoneme–grapheme correspondences are regular or are altered from the usual value of the spelling in general and consistent ways (as by the lengthening rule mentioned above).

Table 3 presents application of these principles in individual words. In the word bean 'woman', the first two letters <be> are treated as a digraph representing the palatalised /bʲ/ and in buí 'yellow' the first two letters <bu> are treated as representing the velarised /b/ (cf. Rule 3).

Table 3. Identifying digraphs according to the rules developed.

Word	Phonetic transcription	Grapheme–phoneme correspondence	Gloss
(a) Quality			
Bean	/bʲan/	be - a - n	'Woman'
Áit	/a:tʲ/	á - it	'Place'
rúnaí	/ru:ni:/	r - ú - na - í	'Secretary'
Caitríona	/kat'r'inə/	c - a - it - r - í - on - a	Woman's name
faitíos	/fatʲi:s/	f - a - it - í - os	'Fear'
buí	/bi:/	bu - í	'Yellow'
(b) Mutation			
Thadhg (<Tadhg)	/haig/	Th - adh - g	Man's name
Mhamó (<Mamó)[a]	/wamo:/	Mh - a - m - ó	'Granny'
gcat (<cat)	/gat/	gc - a - t	'Cat'
dtír (<tír)	/dʲi:rʲ/	dt - í - r	'Country'
bhfaitíos (<faitíos)	/watʲi:s/	bhf - a - it - í - os	'Fear'
ndoras (<doras)	/norəs/	nd - o - r - a - s	'Door'
(c) Vowel digraphs			
saor	/si:r/ or /se:r/	s - ao - r	'Free'
deis	/dʲeʲsʲ/	d - ei - s	'Opportunity'
ceol	/kʲo:l/	c - eo - l	'Music'
tae	/te:/	t - ae	'Tea'

[a]We recognise that the pronounced value of the a in Mamó is not identical to that of the a in bean. The difference is not phonemic, however, but is predictable from the phonological context in which the vowel is found (here, the preceding palatalised vs. velarised consonant). We are concerned here with phonemic values of the spellings and gloss over allophonic differences such as these.

Based on the rules above and their elaboration for each of the most frequent words in this corpus (see the appendix), Table 4 summarises the level of regularity for the 101 most frequent words in the Irish Early Readers Corpus, in line with the analysis carried out by Stuart et al. for the comparable English corpus.

Table 4. Regularity in the 101 most frequent words of the Irish Early Reader Corpus.

Regular by letter-sound rules	Regular by grapheme–phoneme rules	Irregular or exceptions
agam (C, U), agat (C, U), agus (C), as, cá, cad, cat (M, U), cé, dul, é, i, í, in, ina, lá, le, léi, Liam, mé, mise, mór, ná, ní, níl, ó, ocras, orm, sásta, sé, sí, siad, sin, sise, suas, tá, trí, tú	*ach (C, U), am, ann, aon, béar, bhfuil, bhí, breá, buí, chonaic, chuir, Daidí, dearg, deas, deir, dtí, duit, féach, fear, féin, leaba, leat, léim, leis, Mamaí, mhamaí, múinteoir, nach, nuair, Róisín, sibh (M, U) teach, tháinig, thug*	*a, ag, [agam (M), agat (M)], air, amach, an, anois, anseo, ansin, ar, arsa, atá, beag, caithfidh, [cat(C)], chuaigh, do, go, is, isteach, liom, maith, mhaith, mo, na, raibh, rith, sa, Sciob, seo, [sibh(C)], siopa, thosaigh*
37	34	30 irregular in all dialects
= 71 regular in some dialect		34 irregular in some dialect

Note: When dialects vary in the pronunciation of words in ways that lead to irregularities in some regions but not others, we have treated a form as regular if in any dialect the phoneme–grapheme correspondences are regular or are altered from the usual value of the spelling in general and consistent ways, with C representing Connamara dialect, U Ulster dialect, and M Munster dialect.

Using these rules, we found an even higher incidence of regularity for the most frequent hundred words in the Irish corpus than was noted by Stuart et al. in English. At least 71 of the first 101 words in this corpus can be clearly counted as regular in at least one dialect according to the grapheme – phoneme rules, and 37 of those are also regular according to letter-sound rules. All but eight of the latter (or 89%) are function words (pronouns,prepositions, determiners, particles, auxiliaries). The regular content words according to letter-sound correspondences are *cat* 'cat', *dul* 'going', *lá* 'day', *Liam* 'man's name', *mór* 'big', *ocras* 'hunger', *sásta* 'pleased, satisfied', *trí* 'three'. Content words that are regular by grapheme – phoneme patterns but not single letter-sound patterns include *am* 'time', *béar* 'bear', *breá* 'fine', *buí* 'yellow', *chonaic* 'saw', *chuir* 'put', *Daidí* 'Daddy', *dearg* 'red', deas 'nice', deir 'says', *féach* 'look', *fear* 'man', *leaba* 'bed', *léim* 'jump', *mamaí* 'Mommy', *múinteoir* 'teacher', *Róisín* 'woman's name', *teach* 'house', *tháinig* 'came', *thug* 'gave', raising the percentage of content words among the regular spellings from 11% to 39%. The difference between forms that are regular only according to grapheme – phoneme patterns and those with a one-to-one letter-sound regularity is due primarily to the presence in the former group of digraphs regularly representing lenition and eclipsis (thug, mhamaí, nach) and silent vowels signalling (according to a regular pattern) consonant quality in words like *buí, leaba, deir, chuir.*

Of the remaining words, six (eight in Munster) would be regular except for an irregular stress pattern (*amach* 'outward', *anois* 'now', *anseo* 'here', ansin 'there, then', *atá* 'which is', *isteach* 'into' and in Munster also agam 'by me', *agat* 'by you'), and another seven are irregular only by virtue of being inherently unstressed, with their sole vowel accordingly reduced to schwa (a 'verbal particle', *an* 'the', *do* 'your',, *mo* 'my', *na* 'the-pl.', *sa* 'in-the'). This leaves only 15 words (16 in Connacht) of the top 101 that are unambiguously irregular in spelling for reasons other than stress. Of these, 11 are irregular by virtue of an unexpected vowel for the spelling: *ag* 'by, at', *air* 'on it/him', *anseo* 'here', *ar* 'on', *arsa* 'said', *beag* 'little', *liom* 'with me', *raibh* 'was (dependent form)', *siopa* 'shop', *Sciob* 'Name' and *seo* 'this'. The remainder involve consonant irregularities: is 'is' has a broad consonant that would be expected from the spelling to be slender, and *chuaigh* 'went', *thosaigh* 'began' contain silent final consonants in Connacht, whereas in Munster they contain a stop rather than the fricative normally indicated by spelling. Similarly, *sibh* in parts of Connacht is pronounced /sʹibʹ/ rather than with the final /v/ found elsewhere and indicated in spelling. The final th of m(h)aith 'good' and rith 'run' is also silent when not followed by a vowel. With this degree of regularity in the early Reader Corpus, not teaching grapheme – phoneme correspondences is an opportunity missed.

5. Discussion: consistency and complexity in Irish orthography

These results support Lyddy's (2005) claim that Irish is less inconsistent than English, but consistency is not the full story. The regularities established for Irish depend crucially on a complex set of grapheme–phoneme rules, of which those outlined above are only the beginning. One cannot assume that the consistency found will be apparent to beginning Irish readers. The complexity of the rules used to determine consistency provides an additional argument for explicit teaching of grapheme–phoneme correspondences, in a systematic, age-appropriate fashion.

An essential backdrop to this picture of a consistent but complex set of orthographic rules is the unavoidable sociolinguistic context, whereby the powerful effect of transfer from what is already known of English spelling cannot be ignored. Many Irish phonological patterns have no analogue whatever in English. Most notably, consonant-vowel combinations which signal the broad/slender distinction are by no means self-evident, and are rarely explained to learners, which can contribute to spelling pronunciations of these vowels as extra syllables, rather than the appropriate consonant change. Thus, even expert L2 learners of Irish, including teachers, have been found to be unaware of such salient features of Irish orthography. Such explanation requires materials which have not yet been developed.

Given the minority status of Irish and the fact that the vast majority of beginning Irish readers are English-dominant, it is imperative to recognise that learners who already have literacy skills in English will find challenges in both the learning of new patterns and the unlearning of old ones. The problem may be even greater for adult learners, who have already automatised their response to English orthography. Even highly educated adults with advanced metalinguistic skills have difficulty with these issues, as illustrated by the following quote from linguist Andrew Carnie regarding orthographic representation of the palatalised/velarised distinction:

> I . . . had more problems . . . to know which vowels were pronounced, which vowels were there to indicate broad/slender, and which vowels were both. This was compounded by the fact that the broad/slender pronunciation differences are changed by which vowel is actually pronounced. (http://www.linguistlist.org, 14 April 2003)

Lengthy experience with teaching adult learners (second author) indicates that students continue to pronounce even early-learned words according to English phoneme–grapheme rules despite many months of study. Explicit contrastive analysis of Irish and English spelling conventions (as advocated by the DES) could be a pivotal strategy for promoting more efficient acquisition of Irish reading skills.

Dialect variation is also a significant issue, especially for L2 learners of Irish. ESL is typically taught assuming a standard ('received') pronunciation, but the Irish official standard is explicitly oriented towards the written language, and was designed to allow for dialect variation. For native English speakers the variation children might encounter from teachers is offset by community support for their L1. In contrast, the teacher is often the only source of Irish input for L2 learners. This poses problems, especially for primary schools, insofar as a child may, for one grade, have a teacher who speaks one dialect with a particular sound-symbol correspondence for some of the combinations above, and the next year have a teacher with a different dialect and different correspondence for the same spelling (e.g. bord, corresponding in some regions to /o:/ in the list above, corresponds to /au/ in parts of Connemara). It is also common among teachers who have learned Irish as an L2 to have a mixture of dialects, with a variety of item-specific sound-symbol correspondences. Dialect is also relevant to children who are native speakers: Mac Cárthaigh (2006) argued that the Standard Language in which texts are written seems remote to native speakers in the *Gaeltacht* (Irish-speaking communities),

as an artificial variety imposed by Dublin, and poses particular difficulties for native speakers of dialect when they are learning to read.

Therefore, while Irish spelling may indeed be more regular than that of English at least in early reader texts,[4] it is our contention that the complexity and unfamiliarity to L2 readers of the rules underlying that regularity make Irish orthography as difficult as that of English, and particularly difficult for native English speakers already literate in that language; one cannot assume that L2 readers will simply pick up Irish with the ease with which more nearly phonemic orthographies like Welsh and Spanish can be learned. The problems are exacerbated by the existence of major dialect differences within a small geographic area, and dialect mixing in the speech of many L2 Irish speakers (which includes most teachers). For this reason, we argue, the promotion among teachers of greater awareness of spelling rules and dialect variation in pronunciation, as well as more explicit teaching of the patterns and rules of Irish orthography, would be of enormous benefit to beginning readers, by providing them with points of comparison through formal study. Some sample recommendations are outlined below.

6. Teaching Irish orthography: some recommendations

Our recommendations must be viewed as preliminary, pending more detailed analysis of Irish grapheme–phoneme rules. Thus, the development of resources for teachers is a first priority. The current under-exploitation of the teaching of the regularity that does exist in Irish is in part due to the dearth of academic analysis of the issue and to the resulting lack of materials and resources. Such materials as do exist and are currently used in Gaeltacht and immersion schools could be beneficial in all schools. In addition, certain spelling regularities are readily accessible and could be presented systematically even in the absence of more developed resources. We recommend that the teaching of Irish orthographic patterns be treated as an ongoing process, revisited periodically as learners acquire larger vocabularies and increased exposure to both written and spoken forms. A significant aim would be to prevent the development of English-spelling pronunciation patterns becoming fossilised, particularly with regard to the sounding out of vowels which are meant only to indicate consonant quality. Individual rules could be presented one at a time, reinforcing previously taught ones with known examples. Among the things worth covering early on, we propose the following.

6.1. Vowels

(1) The basic values of single (orthographic) vowels, and the length difference signalled by the síneadh fada (acute accent) – *mo* 'my' vs. *mó* 'more' or *na* 'the (pl.)' vs. *ná* 'nor'.
(2) The fact that, if a fada is present in a vowel sequence, the vowel with the fada is the one to pronounce, e.g. the á in *Seán*, the í in *Caitríona*.
(3) The phonetic values of consistent digraphs, such as ei, ea(i), eo(i), ae(i), and ao(i).
(4) The fact that word-final vowels are never silent as they may be in English: i.e. the final e is always pronounced in words like *míle* 'mile/thousand', *mise* 'me'.
(5) The use of specific vowels to mark consonant quality. The long digraphs listed above in (1) could serve as an entry point to this discussion, as the presence or absence of i is determined by the quality of the following consonant, so that (near) minimal pairs can easily be found to illustrate, e.g. *saor* 'free', saoirse 'freedom'. Once the consistent use of i to mark a following palatal is acquired, more variable patterns can be introduced, such as the use of u between a non-palatal consonant and subsequent front vowel (e.g. the use of u between a non-palatal consonant and following front vowel in *cuid* 'part', duit 'to you', etc., but a in *chonaic* 'saw'. Hickey (2007) has discussed the option of learner materials where the vowel

which serves only to indicate consonant quality is presented in a light colour, and the functioning vowel in a dominant colour in order to orient learners to the latter.

(6) Pan-dialectal stress patterns: initial stress and reduction of unstressed short vowels, as well as systematic deviations from these patterns, such as unstressed particles (a, an, mo, etc.) and the set of adverbs with unstressed first syllables (anois, isteach, etc.). The more complex Munster patterns for non-initial stress might be saved for later lessons, with early words showing the shift being taught as exceptions, until enough have been learned to establish the rules.

6.2. Consonants

Features of the consonant system that warrant early discussion and practice include:

(1) The fact that in Irish c is always pronounced as /k/ and g as /g/, in contrast to English, where their pronunciations vary (e.g. cat vs. city, gift vs. gist).
(2) The fact that in Irish the letter s represents two sounds spelled differently in English (as s and sh), the choice depending on adjacent vowel symbols.
(3) Following on (2), the general contrast between palatal and non-palatal (slender and broad) consonants, and the vowels that identify minimal pairs: *bí* 'be' vs. *buí* 'yellow', *bád* 'boat' vs. *báid* 'boats'.
(4) The use of h to mark lenition of a consonant and the change in pronunciation associated with this process. Hickey (2005) found that L2 learners frequently disregard lenition, and this makes the learning of the grammatical functions associated with it more difficult. Some of these alternations can be explicitly taught as a relaxing of the articulation, as is the case with c vs. ch, contrasting, e.g. *caith* 'throw (command)' with *chaith* 'threw' and *teach* 'house'. Others, where historical changes have further altered the pronunciations, may be better taught without reference to specific articulatory relations between the lenited and unlenited forms. Thus, th and sh might simply be taught as alternative representations for the phoneme /h/, equivalent in pronunciation to the single letter h. We recommend focusing at first on initial position when teaching the values of digraphs with h, especially *th*, *sh*, *dh*, *and gh*, *a*s their values medially and finally are more complex and vary across dialects.
(5) The sequences that mark eclipsis, *bp, mb, bhf, dt, nd, gc, and ng,* would benefit from explicit discussion. For all but the last of these, a general rule that only the first letter is sounded can be offered (assuming that lenited bh has already been introduced as representing a single sound), but ng is not pronounced as /n/ and must be dealt with separately. Learners who already know the orthographic value of ng in English can easily transfer its pronunciation to Irish; what must be learned in this case is that in Irish the sound occurs word-initially.

As learners master the basics of orthography covered above, more complicated spelling-pronunciation patterns can be considered, such as the diphthongs created by merger of dh, gh with preceding vowels in medial position in words like praghas /prais/ 'price', or the various dialects' treatments of unstressed final syllables in -(a)igh, -(a)idh, -amh, -adh. Exceptions and sub-patterns that vary from the regularities we have discussed can be taken up as they begin to emerge in the vocabulary being learned.

7. Conclusions

Irish orthography has a very complex system of rules which operate fairly consistently, but learners tend to have significant difficulty in discerning or understanding these rules, partly because they signal phenomena not familiar to L2 learners of Irish and not discussed explicitly outside of linguistic circles. This difficulty has knock-on effects for learners' grammatical and phonological accuracy, and contributes to the strengthening of 'spelling pronunciations', when learners have little exposure to native speakers. Thus, it is vital that teachers are enabled to understand the rules and patterns as well as irregularities in this system, so that they can assist learners systematically. To this end, there needs to be more discussion of this topic among Irish scholars, in order to lay the foundations for the kind of targeted training teachers need, and to develop supporting materials. Our claim here is that more attention to teaching orthographic rules in teacher preparation courses would greatly facilitate the teaching of Irish reading, and help avoid later, enduring problems for learners. We do not argue for a wholesale change towards the teaching of rules only, in isolation from a whole language approach that supports reading for meaning and developing children's interest in Irish reading. Cummins (2003) noted that a focus on:

> . . .explicit systematic phonics teaching is not in any way incompatible with a concurrent or later focus on encouraging extensive reading for meaning. It is unfortunate that the central message of whole language teaching regarding the importance of focusing on meaningful engagement with text and encouraging extensive reading of a wide variety of linguistic genres has gotten lost in the ideological conflicts over reading. (p. 6)

Such systematic teaching is particularly justified in view of both the consistency and the complexity of Irish orthography. Huge challenges face the beginning reader of Irish, particularly those coming from an English background, but even native speakers must grapple with the issue of variation in dialect. Teachers could be enabled to capitalise more on the consistencies that do exist in Irish spelling to help learners, by teaching (simplified versions of) the rules that have been identified. They cannot do this, however, without being helped to become aware of those rules and consistencies, as well as of the real irregularities that all readers must eventually come to grips with. This paper is a call for greater attention to the egregiously overlooked area of Irish orthography.

Acknowledgements

This research was supported by a Government of Ireland senior fellowship to the first author from the Irish Research Council for the Humanities and Social Sciences, and by support from the Dublin Institute for Advanced Studies and the University of Minnesota to the second author. The help of An Gúm and ITÉ is also gratefully acknowledged. The authors would like to thank the anonymous reviewers for their insightful comments.

Notes

1. In Irish, the language is referred to as Gaeilge, and the term 'Gaelic' is derived from this. However, the term 'Irish' is preferred in Ireland when referring to the language in English, in order to distinguish it from (Scots) Gaelic.
2. In dialects, where nn, ll, rr, represent distinct phonemes from n, l, r, this number may be higher.
3. A digraph like eo functions like the single vowel in cód /ko:d/, identifying both vowel pronunciation and surrounding consonant quality. It regularly signals a long /o:/ following a slender consonant and preceding a broad one, functioning identically to ó, used when both preceding and following consonants are broad. The only fixed sequence that does not represent a long vowel is ei, which always represents short /e/ before a consonant. It is simply a quirk of Irish orthography that the grapheme e appears alone only in final position (e.g. te 'hot'). Long /e:/ is also always written as a digraph before consonants, éi before a slender consonant or éa when it is broad. In what follows, we will treat ei as a regular digraph, since the i is essentially redundant with respect to consonant quality, but will consider the a which follows é and precedes a broad consonant to form a digraph with that consonant, according to the rule presented in (3). The a of the short digraph ea is usually the pronounced vowel, with the e signalling the quality of the preceding consonant and it is treated as a digraph with that consonant.
4. No studies are available for written Irish targeting adults (and non-learners), but we suspect that the vocabulary found in such works, apart from the commonest function words listed above, will most likely vary more than in children's literature, leaving open the possibility of greater irregularity across texts.

References

An Gúm. (various years). Séideán sí. Dublin: Government Publications.

Bernhardt, E. (2003). Challenges to reading research from a multilingual world. Reading Research Quarterly, 38(1), 112–117.

Cummins, J. (2003). Reading and the bilingual student: Fact and friction. In G. Garcia (Ed.), English learners: Reaching the highest level of English literacy (pp. 2–33). Newark, DE: International Reading Association.

Department of Education and Science (DES). (1999). Gaeilge. Retrieved October 22, 2010, from http://www.ncca.ie/uploadedfiles/Curriculum/Primary_Gaeilge_Curriculum.pdf

Department of Education and Science (DES). (2008). Irish in the primary school: Promoting the quality of learning. Dublin: Stationery Office. Retrieved from http://www.education.ie/servlet/blobservlet/des_irish_primary_school_cont.htm

Ehri, L., Nunes, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. Reading Research Quarterly, 36(3), 250–287.

Ellis, N., Natsume, M., Stavropoulou, K., Hoxhallari, L., van Daal, V., Polyzoe, N., & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic and logographic scripts. Reading Research Quarterly, 39(4), 438–468.

Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. Applied Linguistics, 25(1), 1–37.

Goswami, U. (2006, September 20–21). A psycholinguistic grain size view of reading acquisition across languages. Paper presented at EPS Research Workshop: The Role of Orthographies in Reading and Spelling, Middlesex University.

Goswami, U., Gombert, J.E., & de Barrera, L.F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. Applied Psycholinguistics, 19(1), 19–52.

Grabe, W. (2004). Research on teaching reading. Annual Review of Applied Linguistics, 24, 44–69.

Grabe, W., & Stoller, F. (2002). Teaching and researching reading. New York, NY: Longman.

Harris, J. (2007). Bilingual education and bilingualism in Ireland North and South. International Journal of Bilingualism and Bilingual Education, 10(4), 359–368.

Harris, J., Forde, P., Archer, P., Nic Fhearaile, S., & O'Gorman, M. (2006). Irish in primary schools: Long-term national trends in achievement. Dublin: Department of Education and Science.

Harrison, C. (1998). Methods of teaching reading: Key issues in research and implications for practice. EducatiON-Line, 39. Retrieved October 22, 2010, from http://www.leeds.ac.uk/educol/documents/000000488.htm

Harrison, G., & Krol, L. (2007). Relationship between L1 and L2 word-level reading and phonological processing in adults learning English as a second language. Journal of Research in Reading, 30(4), 379–393.

Hickey, T. (2005). Second language writing systems: Minority languages and reluctant readers. In V. Cook & B. Bassetti (Eds.), Second language writing systems (pp. 398–423). Clevedon: Multilingual Matters.

Hickey, T.M. (2007). Fluency in reading Irish as L1 or L2: Promoting high frequency word recognition in emergent readers. International Journal of Bilingual Education and Bilingualism, 10(4), 471–493.

Hu, M., & Nation, I.S.P. (2000). Vocabulary density and comprehension. Reading in a Foreign Language, 13(1), 403–430.

Institiúid Teangeolaíochta Éireann. (2003). Corpas Náisiúnta na Gaeilge/National Corpus of Irish (Vol. 1). Dublin: ITÉ.

Juel, C. (2006). The impact of early school experiences on initial reading. In D. Dickinson & S. Neuman (Eds.), Handbook of early literacy research (Vol. 2, pp. 410–426). New York: Guilford Press.

Koda, K. (2004). Insights into second language reading. New York, NY: Cambridge University Press.

Lyddy, F. (2005). Celtic biliteracy. Literacy Today, 43, 1–2.

Mac Cárthaigh, M. (2006). Literacy and L1 speakers of Irish. In T.M. Hickey (Ed.), Literacy and language learning: Reading in a first or second language (pp. 52–64). Dublin: Reading Association of Ireland.

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. Applied Linguistics, 26(1), 32–47.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes, 63(1), 59–82.
National Institute of Child Health and Human Development (NICHD). (2000). Report of the National Reading Panel. Teaching children to read (NIH Publication No. 00-4769). Washington, DC: US Government Printing Office.
Ó Faoileáin, D. (2006). Developing a phonics programme in Irish for the teaching of reading in the Gaeltacht. In T.M. Hickey (Ed.), Literacy and language learning: Reading in a first or second language (pp. 65–73). Dublin: Reading Association of Ireland.
Ó Murchú, M. (1977). Successes and failures in the modernization of Irish spelling. In J. Fishman (Ed.), Advances in the creation and revision of writing systems (pp. 267–289). The Hague: Mouton.
Ó Sé, D. (1990). Súil ghéar ar litriú na Gaeilge (A sharp eye on the spelling of Irish). Teangeolas, 28, 16–18.
Ó Siadhail, M. (1995). Learning Irish (2nd ed.). New Haven, CT: Yale University Press.
O'Sullivan, J., & Goosney, J. (2007). Get ready, get set, get going: Learning to read in northern Canada. Thunder Bay: Centre of Excellence for Children and Adolescents with Special Needs, LakeheadUniversity. Retrieved from http://www.coespecialneeds.ca/uploads/docs/getreadygetsetenglish.pdf
Parsons, C., & Lyddy, F. (2009). Early reading strategies in Irish and English: Evidence from error types. Reading in a Foreign Language, 21(1), 22–36.
Pikulski, J., & Chard, D. (2005). Fluency: Bridge between decoding and reading comprehension. The Reading Teacher, 58(6), 510–519.
Seymour, P., Aro, M., & Erskine, J. (2003). Foundation literacy acquisition in European orthogra- phies. British Journal of Psychology, 94(2), 143–174.
Sheils, J., & Ó Sé, D. (2004). Teach yourself Irish (3rd ed.). London: McGraw-Hill.
Stuart, M., Dixon, M., Masterson, J., & Gray, B. (2003). Children's early reading vocabulary: Description and word frequency lists. British Journal of Educational Psychology, 73(4), 585–598.
Stuart, M., Masterson, J., & Dixon, M. (2000). Spongelike acquisition of sight vocabulary in beginning readers. Journal of Research in Reading, 23(1), 12–27.
Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. Psychological Bulletin, 131(1), 3–29.

Appendix: Top 101 Words in Early Readers[1]

Rank	Word	Letter -Sound	Graph-Phon	Pronunciation	Letter/Grapheme-Phoneme Division & Other Notes
2	A	X	X	/ə/	Irregular since in normal speech this is a schwa
35	Ach	X	R	/ax/	A-ch CH=X Digraph
8	Ag	X	X	/eg/	A-g
			R(C,U)	/agəm/ R in C & U	
44	Agam	R	X(M)	but /əgum/ in M	A-g-a-m
			R(C,U)	/agət/) R in C & U	
82	Agat	R	X(M)	but /əgut/ in M	A-g-a-t
			R(C)		
6	Agus	R	X(M,U)	/agəs/ R in Conn	A-g-u-s
83	Air	X	X	/er'/	A-ir
			R(C)	/am/ in Conn	A-m Regular Grapheme-Phoneme correspondence in C, but final 'm' causes regular vowel
100	Am	X	X(M)	/aum/ in M	change in M (not in other dialects).
31	Amach	X	X	/ə'max/	A-m-a-ch X for Letter Sound because of the 'ch', and X for Grapheme-Phoneme for the schwa first vowel syllable schwa
1	An	X	X	/ən/	A-n In normal speech this is schwa
				/a(:)n/ in C & U	A-nn Irregular by Letter-Sound, but regular by Grapheme-Phoneme where the 'nn' effect on the vowel is regular- treating the nn digraph as a regular vowel lengthener (but different
34	Ann	X	R	/aun/ in M	realisations in different dialects)
28	Anois	X	X	/ə'nos'/	A-no-i-s First syllable schwa
87	Anseo	X	X	/ən's'o/	A-n-se-o First syllable schwa eo breaks eo=long vowel rule
75	Ansin	X	X	/ən's'in'/	A-n-s-i-n First syllable schwa
99	Aon	X	R	/e:n/ in C and M /í:n/ in U	Ao-n Ao digraph = /e:/
4	Ar	X	X	/er/	A-r
3	Arsa	X	X	/ersə/	A-r-s-a
64	As	R	R	/as/	A-s
19	Atá	X	X	/ əta:/	a-t-á first syllable schwa

[1] 101 words listed as last 2 had same frequency.

Rank	Word	Let-Sound	Graph-Phon	Pronunciation	Letter/Grapheme-Phoneme Division & Other Notes
51	Beag	X	X	/b'ug/ in M & C /b'eg/ in U	Be-a-g Irreg vowel (pronounced /u/ in two dialects).
80	Béar	X	R	/b'e:r/	B-é-ar
24	Bhfuil	X	R	/wil'/	Bhfu-i-l Bhf=w
14	Bhí	X	R	/v'i:/	Bh-í Bh=v
84	Breá	X	R	/br'a:/	Bre-á
101	Buí	X	R	/bi:/	Bu-í
43	Cá	R	R	/ka:/	C-á
46	Cad	R	R	/kad/	C-a-d
52	Caithfidh	X	X	/kahəg'/ M	C-aith-fidh 'f' silent
94	Cat	R(M&U)	X(C)	/kat/ (M&U) /kut/ (C)	C-a-t
48	Cé	R	R	/k'e:/	C-é
36	Chonaic	X	R	/xonək'/	Ch-o-na-i-c
73	Chuaigh	X	X	/xuəg'/ in M /xuəɣ'/ in U	Final 'gh' pronounced differently in each dialect from its pronunciation syllable-initially
54	Chuir	X	R	/xir'/	Chu-i-r
42	Daidí	X	R	/dad'i:/	D-a-id-í
88	Dearg	X	R	/d'arəg/	De-a-r-g ‘rg' ‘rm' regular as ‘rəg' ‘rəm'
50	Deas	X	R	/d'as/	De-a-s
45	Deir	X	R	/d'er'/	D-e-ir
32	Do	X	X	/də/	D-o Schwa
61	Dtí	X	R	/d'i:/	Dt-í
66	Duit	X	R	/dit'/	Du-i-t
56	Dul	R	R	/dul/	D-u-l
13	É	R	R	/e:/	É
95	Féach	X	R	/f'e:x/	F-é-ach
68	Fear	X	R	/f'ar/	Fe-a-r
38	Féin	X	R	/f'e:n'/	F-é-in
11	Go	X	X	/gə/	G-o Schwa
29	I	R	R	/i/	I
62	Í	R	R	/i:/	Í
76	In	R	R	/in'/	I-n
39	Ina	R	R	/in'ə/	I-n-a
10	Is	X	X	/is/	I-s

Rank	Word	Let-Sound	Graph-Phon	Pronunciation	Letter/Grapheme-Phoneme Division & Other Notes
27	Isteach	X	X	/əs't'ax/	I-s-te-a-ch Initial schwa
96	Lá	R	R	/la:/	L-á
23	Le	R	X	/l'ə/	L-e
59	Leaba	X	R	/l'abə/	Le-a-b-a
63	Leat	X	R	/l'at/	Le-a-t
98	Léi	R	X	/l'e:i:/ in M but /l'e:/ /l'e:i/ /le:j/l & /le:hi/) elsewhere	L-é-i
81	Léim	X	R	/l'e:m'/	L-é-im
26	Leis	X	R	/l'es'/	Le-i-s
47	Liam	R	R	/l'iəm/	L-i-a-m
25	Liom	X	X(C)	/l'um/ in C /l'om/ in U and M	Li-o-m
22	Maith	X	X	/mah/	M-a-ith 'th' is silent
40	Mamaí	X	R	/mami:/	M-a-ma-í
9	Mé	R	R	/me:/	M-é
77	Mhaith	X	X	/wah/	Mh-a-ith 'th' is silent
97	Mhamaí	X	R	/wami:/	Mh-a-ma-í Mh =/w/ regularly consistent but not across dialects
37	Mise	R	R	/mis'ə/	M-i-s-e
17	Mo	X	X	/mə/	M-o Schwa
49	Mór	R	R	/mo:r/	M-ó-r
60	Múinteoir	X	R	/mu:n't'o:r'/	M-ú-in-te-o-ir 'In' =/n'/ 'te' here=/t'/ 'ir' here=/r'/
12	Na	X	X	/nə/	N-a Schwa
58	Ná	R	R	/na:/	N-á
78	Nach	X	R	/nax/	N-a-ch
18	Ní	R	R	/n'i:/	N-í
30	Níl	R	R	/n'i:l'/	N-í-l
89	Nuair	X	R	/nuər'/	N-ua-ir Treating the ua digraph as reg for the diphthong
41	Ó	R	R	/o:/	Ó
90	Ocras	R	R	/okrəs/	Oc-r-a-s
69	Orm	R	R	/orəm/	O-r-m
55	Raibh	X	X	/rev'/	R-aibh Vowel 'a' = /e/(M) /o/(C) 'Bh' silent in C

70	Rith	X	X	/rih/	R-i-th X=treating final th as silent
53	Róisín	X	R	/ro:s'i:n'/	R-ó-is-í-n
15	Sa	X	X	/sə/	Schwa
57	Sásta	R	R	/sa:stə/	S-á-s-ta
74	Sciob	X	X	/sc'ub/ in C and M	Sci-o-b Vowel /u/
7	Sé	R	R	/s'e:/	S-é
21	Seo	X	X	/s'o/	Se-o Breaks eo=o: rule
16	Sí	R	R	/s'i:/	S-í
67	Siad	R	R	/s'iəd/	Si-a-d
85	Sibh	X	R(M,U)	/s'iv'/ in M =R /s'ib/ in C	S-ibh Bh=/v/ regular
33	Sin	R	R	/s'in'/	S-i-n
91	Siopa	X	X	/s'upə/ in C & M /s'ip'ə/ in U	Si-o-p-a
79	Sise	R	R	/s'is'ə/	S-i-s-e
71	Suas	R	R	/suəs/	S-ua-s Ua diphthong reg
5	Tá	R	R	/ta:/	T-á
72	Teach	X	R	/t'ax/	Te-a-ch
86	Tháinig	X	R	/ha:n'əg'/	Th-á- in-i-g th =h digraph
65	Thosaigh	X	X	/hosig'/ in M /hosi:/ in Conn	'Gh' as in 'Chuaigh'
92	Thug	X	R	/hug/	Th=h digraph
93	Trí	R	R	/t'r'i:/	T-r-í
20	Tú	R	R	/tu:/	T-ú