Reading Development & Teaching

Morag Stuart and Rhona Stainthorp

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

First published 2016

Library of Congress Control Number: 2015939932

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

Editor: Luke Block
Editorial assistant: Katie Norton
Production editor: Imogen Roome
Marketing manager: Michael Ainsley
Cover design: Wendy Scott
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed and bound by CPI Group (UK) Ltd, Croydon, CR0 4YY

ISBN 978-1-4462-4903-1
ISBN 978-1-4462-4904-8 (pbk)

1

Essential knowledge about language

Summary

In this chapter, you will learn about the phonological system of languages: the ways in which speech sounds are identified and combined to produce and understand spoken language. You will learn about the orthographic systems of languages: the ways in which written languages represent their spoken forms. This is essential knowledge for those of us involved in understanding and teaching reading. The contents of this chapter will equip you with the knowledge you need to read subsequent chapters, and will enable you to answer the question: why is learning to read words in English so much more difficult than in most other languages?

Introduction

A recent study comparing rate of development of word reading skills in children learning to read in nine European languages (Seymour, Aro & Erskine, 2003) demonstrated that learning to read words in English is more difficult than in most other alphabetic languages. For children learning to read in English, these skills developed more than twice as slowly as those of beginners learning to read in Finnish, Spanish or Greek. Of course, difficult is not synonymous with impossible. The vast majority of children in the UK (and other countries where English is the first language) learn to read words accurately and fluently in an acceptable time scale. However, the early stages of learning to read are more taxing in English, and it takes longer for children to become fluent.

The difficulty lies within the *orthography*, which Henderson (1984) defined as the conventional writing system of a language and the way this maps onto spoken language. Each language has its own pattern and rules for writing: its own *orthographic system*. Beginner readers must learn how the orthographic system relates to spoken language in a given writing system. In this chapter, we explain how writing systems work and why the English writing system is so difficult.

We need to introduce you to a number of technical terms: these help us to be precise about our meanings. We will give you some exercises to do to help you feel confident about the terms. Because you are a skilled adult reader, you have already internalized knowledge of English orthography: thus, most of the information we provide will already be part of your implicit understanding. To understand reading development and how to teach children to read, it is useful to transform this implicit understanding into accessible explicit knowledge. We hope that by the time you finish this chapter you will be convinced that, albeit difficult, English orthography can be fascinating, and good teachers can pass on their fascination to their pupils. Writing systems represent the sounds of spoken language, its phonology. Before we consider writing systems themselves, we have to make an extensive detour to describe the English phonological system, which is what English orthography represents.

Phonology: the sound system of language

The sound system of language is called *phonology*, from the Greek *phone* meaning sound and *logos* meaning speech. So the word 'phonology' literally means the sounds of speech. The phonological system is the system of language that uses sounds as its units and these sounds are combined to produce individual words.

We will first list all the speech sounds (phonemes) of English, each paired with its International Phonetic Alphabet (IPA) symbol. Each row in the tables gives you the unique IPA symbol for the phoneme and an example of that phoneme as pronounced in a real English word. Spelling of the phoneme in question is printed in bold. Table 1.1 lists IPA symbols for consonant phonemes, and Table 1.2 lists IPA symbols for vowel phonemes.

Table 1.1 IPA symbols for the 24 consonant phonemes of English

IPA symbol	Example of a word containing the sound
p	**p**at
b	**b**at
t	**t**at
d	**d**og
k	**c**at
g	**g**oat
f	**f**at
v	**v**ote
θ	**th**umb
ð	**th**ey
s	**s**at

IPA symbol	Example of a word containing the sound
z	**z**oo
ʃ	**sh**op
ʒ	trea**s**ure
tʃ	**ch**ip
dʒ	**j**ug
m	**m**at
n	**n**ot
ŋ	si**ng**
l	**l**og
r	**r**at
j	**y**ellow
w	**w**atch
h	**h**ello

Table 1.2 IPA symbols for the 20 vowel phonemes of English

IPA symbol	Example of a word containing the sound
Short monophthong vowels	
ɪ	s**i**t
ɛ	w**e**t
æ	c**a**t
ɒ	pl**o**t
^	d**u**ck
ʊ	p**u**t
ə	b**a**nana
Long monophthong vowels	
i	tr**ee**
ɜ	g**ir**l
ɑ:	f**a**ther
ɔ	s**aw**
u	sh**oe**
Diphthong vowels[1]	
ei	pl**a**y
əʊ	g**o**
ai	s**igh**
aʊ	n**ow**

(Continued)

Table 1.2 (Continued)

IPA symbol	Example of a word containing the sound
ɔi	b**oy**
ɪə	f**ear**
ɛə	th**ere**
ʊə	p**ure**

[1]A diphthong vowel is one where the tongue moves as it is produced so that the sound appears to glide from one vowel to another.

You will see from the tables that there are 44 phonemes in English: 24 consonant and 20 vowel phonemes. These statistics alone alert us to the challenges encountered in English orthography, because English has only 26 letters to represent these 44 phonemes. This imbalance of number of phonemes and number of letters also poses a challenge when writing about the sound system of language. The International Phonetic Alphabet allows us unambiguously to represent each phoneme with a specific symbol. Throughout the book, we will be using IPA symbols when we want to specify particular phonemes.

Above, we equated 'phoneme' with 'speech sound': the box below gives a more complete definition of the phoneme.

What is a phoneme?

A **PHONEME** is the smallest unit of speech sound in a word that changes meaning.

Thus the word <BED> is composed of three phonemes /bɛd/. If we change the first phoneme /b/ to /r/ we get /rɛd/ <RED>. This is a different word with a different meaning.

[It can sometimes be very confusing to represent the sound and the look of a word in texts. In this book, where a word has been printed in brackets like so: < >, we are representing the letters. Where it is printed in slash marks like so: //, we are representing the sounds. Where the topic relates to the word *per se* and the contrast is not between the orthography and the phonology, we will just present the word in upper case letters, like so: WORD.]

There is now considerable evidence that the average adult is not explicitly aware of phonemes in words (Moats, 1994; Stainthorp, 2004). When asked, 'How many sounds are there in the word RUST?', adults are just as likely to say 'two' or 'three' as 'four'. When asked to explain their decision, one person may say that the two sounds are /r/ and /ʌst/; whereas another might say the three sounds are /r/ /ʌ/ and /st/.

The correct answer is **four phonemes**: /r/ /ʌ/ /s/ /t/. We can see how this works if we use the definition of the phoneme given above.

We start with the word /rʌst/ meaning 'iron oxide'. Changing the first phoneme from /r/ to /d/ gives us the word /dʌst/ (DUST) meaning 'tiny particles lying on a surface'. By changing one phoneme for another we end up with a word with a different meaning, so the initial phoneme and the substituted one must both be phonemes of English.

Changing the second phoneme from /ʌ/ to /ɛ/ gives us /rɛst/ (REST) meaning 'to stop work'. We can swap the final two phonemes round so instead of /rʌst/ we have /rʌts/. Again we get a different word. We can also say /rʌst/ without the /s/ phoneme and then we get yet another word /rʌt/.

You might like to try specifying the number of phonemes in this set of words:

STRAIGHT

ENOUGH

TAX

KISSED

BATTED

(Answers on the next page.)

You might also try systematically changing the phonemes in each of these five words to make new words that differ by one phoneme. By playing with the phonemes like this, you will begin to raise your level of phonemic awareness. This is important because in order to teach children their letter-sound correspondences for phonics, teachers need to be confident that they have a fluent ability to identify and manipulate the phonemes in words. Educational psychologists assessing causes of reading difficulties also need these skills. Remember to focus on the phonemes of the words and not the letters.

In the IPA table the phonemes are arranged into two primary groups: consonants and vowels, but we have not yet defined the terms consonant and vowel. These definitions are shown in the next box.

Defining consonant and vowel phonemes

CONSONANT phonemes are those sounds where there is a degree of constriction of the air as it flows out of the mouth or nose.

VOWEL phonemes are those sounds where the air flows out of the mouth without any constriction.

Counting phonemes

	Number of phonemes
STRAIGHT	5 = /s/ /t/ /r/ /ei/ /t/
ENOUGH	4 = /ɪ/ /n/ /ʌ/ /f/
TAX	4 = /t/ /æ/ /k/ /s/
KISSED	4 = /k/ /ɪ/ /s/ /t/
BATTED	5 = /b/ /æ/ /t/ /ɪ/ /d/

In everyday language, the term consonant normally refers to the 21 letters B, C, D, F, G, H, J, K, L, M, N, P, Q, R, S, T, V, W, X, Y, Z, and the term vowel refers to the five letters A, E, I, O, U. The letter Y is generally considered to be a consonant, although much of the time it is used to represent a vowel phoneme, as in BY, FLY, CRY, etc. Because this distinction between letters and sounds can cause confusion, throughout this book we will refer to *consonant letters* and *vowel letters* when considering the written form, and *consonant phonemes* and *vowel phonemes* when considering the spoken form. As already stated, in English there are not enough letters for a one-to-one match between letter and phonemes. This is particularly the case for the vowel phonemes, with only five vowel letters (plus Y) to represent the 20 vowel phonemes in English: thus, learning letter–sound correspondences for vowel phonemes poses an extra challenge for learners. As spelling has developed over the centuries and incorporated orthographic features from different languages, combinations of letters have become used to represent all the different vowel phonemes.

We have covered the smallest unit of phonology that impacts on the writing system: this is the phoneme. At the largest level there are the words themselves. These are sequences of phonemes that are blended together and that carry meaning. An additional useful technical term in the realm of meaning is *morpheme*, defined in the next box. The word comes from the Greek *morphe*, meaning form.

What is a morpheme?

A **MORPHEME** is the smallest grammatical unit of language that has meaning. Each morpheme constitutes either a word, or a meaningful part of a word.

BED is a word composed of one morpheme. It means a flat surface on which one lies to sleep (when it is used as a noun). BEDS has two morphemes, it means more than one bed.

Some morphemes are called *free morphemes*. These are words in their own right. Each word carries meaning and has its own syntactic (grammatical) status, such as noun, verb, adjective, etc. In language, words are combined together in a rule governed way to form phrases in order to convey meaning. (We cover these aspects of language more extensively in Chapter 6, when we discuss language comprehension.)

Some morphemes are called *bound morphemes*. Below the level of the word are morphemes that carry meaning, but which cannot stand on their own. Hence the term 'bound': they have to be bound to other morphemes. In the box above, the morpheme /z/ (spelled <S>), which conveys plurality, is a bound morpheme. This is because it cannot stand on its own. It has to be affixed to another morpheme. Other examples of bound morphemes include the past tense ending, -ED /ɪd/ (WANT → WANT**ED**); the present progressive tense ending – -ING /ɪŋ/ (WANT → WANT**ING**); the -ER /ə/ ending depicting an agent (FARM → FARM**ER**), or a comparative (HAPPY → HAPPI**ER**); and the -EST /ɛst/ ending depicting a superlative (HAPPY →HAPPI**EST**). The new Programme of Study for English in the revised National Curriculum (Department for Education, 2014) requires most of the bound morphemes given as examples here to be taught to Year 1 pupils in England.

In between the phoneme level and the word level there are other units of phonology. One of these is the *syllable*, defined in the next box.

What is a syllable?

A **SYLLABLE** is a unit of spoken language formed of one obligatory vowel phoneme and possibly preceded by or followed by optional consonant phonemes.

BED /bɛd/ is a single syllabic word composed of the vowel phoneme /ɛ/, preceded by the consonant phoneme /b/ and followed by the consonant phoneme /d/.

BEDROOM is a bisyllabic word composed of the syllable /bɛd/ and the syllable /ru:m/.

Though most adults are not explicitly aware of phonemes, they are much more comfortable with syllables. The syllable seems to make intuitive sense to English speakers and people can happily clap to the 'beat' of language. In effect, what happens when we clap to the beat is that we clap on each syllable. When we do this, we are clapping on each vowel phoneme.

The final set of terminology about the phonological system relates to the structure of the syllable, illustrated first in 'The structure of the syllable' box below.

The structure of the syllable

The syllable is composed of three segments:

ONSET, NUCLEUS (or PEAK) and CODA

The **ONSET** is the consonant phoneme or consonant phoneme cluster at the start of the syllable. This is optional in English. The NUCLEUS is the vowel phoneme. This is compulsory. The CODA is the final consonant phoneme or consonant phoneme cluster. This is also optional in English.

Here are some examples:

<OWE> /əʊ/ is a syllable that is a word with just the nucleus;

<OWN> /əʊn/ has a nucleus + coda;

<GO> / gəʊ / has an onset + nucleus;

<GOES> / gəʊz / has an onset + nucleus + coda.

Try building up sets of words like this from single vowel phonemes that also happen to be words, like ARE, AIR, EAR, I, YOU. You have to focus on the phonemes and not the spelling. With a bit of perseverance, you can get up to a single syllable word that has up to three consonant phonemes in the onset and in the coda.

You are likely to find that the spelling of the word influences you too much at first. For example, you might decide not to add the phoneme /b/ to the beginning of AIR, because the spelling BAIR is not a real word in English – but the sound pattern, /bɛə/ is, and it is word sounds we are playing with here. Conversely, if you add the letter G to the word OWN you do get a real word GOWN. But the letters OW in GOWN represent the phoneme /aʊ/ not the phoneme /əʊ/, which is in the source word, OWN. So again, spelling has misled you; /gəʊn/ is not a real word in English. However, if you go on to insert the letter R to GOWN after the letter G, the nucleus reverts to the phoneme /əʊ/ as in the word GROWN or indeed GROAN. Spelling sometimes works!

It is when we begin to reflect upon how spelling and sound interact in this way that we recognize how difficult English orthography can be.

In relation to the syllable, there are just two further terms we need to define: terms that describe how the three phonemic elements (onset, nucleus and coda) can be clustered together. These terms are defined in 'The subsyllabic units of body and rime' box.

The subsyllabic units of body and rime

The **BODY** is the onset and nucleus clustered together, e.g. the BEA- of BEAT.

The **RIME** is the nucleus and coda clustered together, e.g. the -EAT of BEAT.

When two words share the same rime element, they are said to *rhyme*. Thus FOX, BOX, SOCKS all share the rime /ɒks/. They differ only in their onset phoneme. Rhyming seems to be intuitively easy for English speakers: people can happily generate strings like B**ED**, S**AID**, F**ED**, H**EAD** when asked to generate rhyming words. Try it yourself with CAT and TOE. Remember to ignore spelling! Playing rhyming games with young children can be fun, and helps them develop their insights into the sound structure of words.

The body seems to make less intuitive sense to English speakers. We can generate strings of words sharing the same body (e.g. **BE**D, **BE**T, **BE**CK, **BE**ND), but this seems to demand more attention and trips off the tongue less readily than rhyming strings.

The set of tree diagrams in Figure 1.1 show you the possible ways of deconstructing the syllable. SCHOOL is used as the exemplar word.

You will note that in the word <SCHOOL> /skul/, the onset has two phonemes /s/ and /k/ as an initial cluster /sk/. A *consonant cluster* is where two consonant

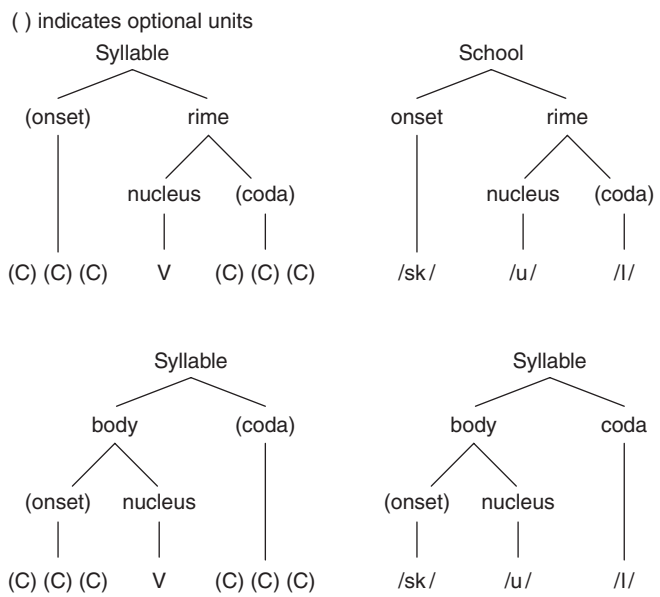

Figure 1.1 Possible ways of deconstructing the syllable

phonemes are pronounced together without any intervening vowel phoneme. Languages vary greatly in terms of the sequences of phonemes allowed in the different positions in the syllable; English permits a number of different consonant phoneme clusters in the onset position. All the consonant phonemes except /ŋ/ can occupy the onset position as single phonemes. Twenty-three different two-phoneme consonant phoneme clusters can also occupy the onset position (can you work out what they are?). English even permits three-phoneme consonant cluster onsets: <STRING> <STRETCH>. Notice that <THREE> does not have a three-phoneme consonant cluster because the letters <TH> represent the phoneme /θ/ (check the IPA list) so the three letters in the onset position represent just two phonemes: /θ/ and /r/.

Similarly, all the consonant phonemes except /j/ and /h/ can occur as the coda. Also, some consonant phoneme clusters can occur as the coda in syllables and some cannot. There is no rationale for why some consonant clusters can appear as onsets or codas in English and why some cannot. For example, the cluster /ks/ cannot appear as an onset but it can occur as a coda – e.g. <TACKS> /tæks/. It is certainly nothing to do with whether or not we can articulate them. All we can say is that some can and some cannot.

Writing systems

Having introduced you to a number of important concepts relating to the phonological aspects of language, it is now time to turn to the orthography itself.

Writing is one of the most important inventions of the human mind. Through writing, the thoughts and feelings of people long dead can speak to us down the centuries. Their voices were never recorded but, through their texts, these voices can be 'heard'. Writing systems evolved from spoken language and capture spoken language in visual form. However, whereas spoken language is a biologically determined behaviour that is common to humans as a species, writing systems are culturally determined and only found in those societies that have created and adopted them. This is what we mean by 'invention'.

Writing systems are graphic. They all have some characteristics in common. They make use of a limited set of strokes configured in different ways to represent language. They differ in terms of the stroke patterns that they use, and also in the linguistic units that are represented by the graphic units. All existing modern writing systems can be broadly categorized on the basis of the way in which the language units are mapped onto orthographic units (Gelb, 1963). The orthographic units can represent *phonemes*, *syllables* and *words*. These are all concepts with which you are now familiar. Writing systems that represent language at the level of the largest unit – the word or concept – are called *logographic* or *ideographic* orthographies. Those that represent language at the level of an intermediate size of unit – the syllable – are called *syllabaries* or *alphasyllabaries*. And finally, those that represent language at the level of the smallest unit – the phoneme – are called *alphabets*, which will be discussed later in this chapter.

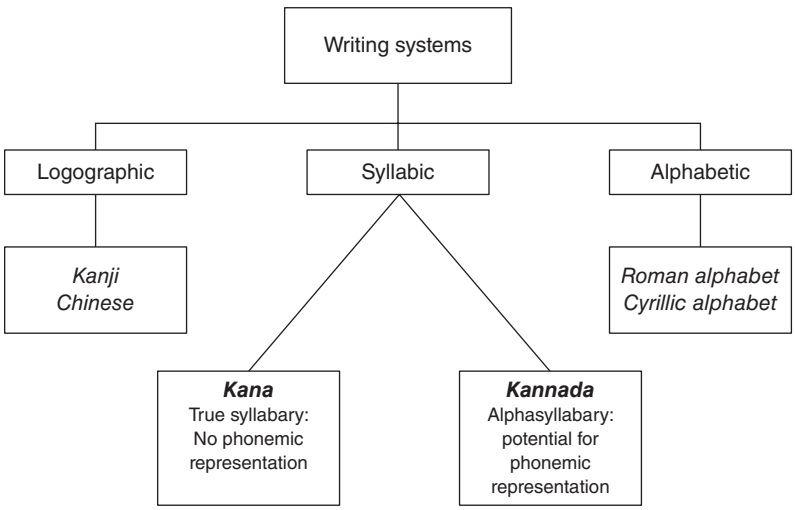

Figure 1.2 Basic categories of writing systems (from Coltheart, 1984)

Logographic writing systems

One of the most well-known examples of a logographic writing system is the hieroglyphic system used in Ancient Egypt. But we do not need to go back into the depths of history to find such a system. Contemporary Chinese and Japanese Kanji scripts are examples of logographic writing systems. Relative to English, they seem to represent language at the level of the concept or word. The graphic units are called *characters*. There are phonetic elements in some characters, which give Chinese readers insight into the sound as well as the meaning of the character. Reading in Chinese requires recognition of thousands of unique characters, with each character standing for a different concept. This poses a heavy strain on memory and, not surprisingly, it takes a long time for children to become fluent readers of Chinese and Japanese Kanji script.

Syllabic writing systems

The two Japanese Kana scripts represent language at the level of the Mora. A *mora* in Japanese is similar to a syllable in English. There are 48 different characters in each Kana script to represent the 48 different mora in the spoken language. So, to learn to read the Kana scripts, children have to learn the 48 mora-kana character correspondences in each script and then they can identify the words on the page. Most children enter school in Japan having already learned to identify the Kana characters, but they then have to learn close to 3,000 separate Kanji characters during their school years in order to be fluent word readers of everyday written Japanese.

The Kannada script, which is one of a number of different scripts used in India, is called an alphasyllabary because it represents language at the level of the syllable, but has the potential to represent phonemes as well. Children have to learn more than 400 individual characters and continue to learn these long into their primary school years (Nag & Snowling, 2012). The memory requirements are considerably more than for learning the Roman alphabet used for English, but the system is more predictable than English. Predictability is an important characteristic when learning to read.

Alphabetic writing systems

Alphabetic writing systems are the most common systems in use today. They represent language at the level of the phoneme. These writing systems largely derive from the Phoenician system, which was in use at least 3,500 years ago. The original Phoenician script just represented the consonant phonemes; the Greeks later devised letters for the vowels as well. This characteristic of representing only consonant phonemes in the written form is still found today in Hebrew and Arabic.

Modern European languages, including English, use alphabetic writing systems that represent both consonant phonemes and vowel phonemes. They do not all use the same alphabet: Russia uses the Cyrillic Alphabet, Greece has its own Greek alphabet and English uses the Latin alphabet. Across different languages, alphabets vary in the predictability with which they represent sounds. Some languages use accents (e.g. French: café, scène) or diacritic marks (e.g. Spanish: mañana, French: garçon) to modify letters: this increases the number of letters and shrinks the gap between number of letters and number of phonemes.

One of the newest writing systems is Turkish, reformed following a conference in 1929. The Latin alphabet was adopted, and a rational, consistent orthographic system was designed with one letter to represent each of the phonemes in modern Turkish. It is one of the simplest alphabetic systems in use today. The 26 letters in the Latin alphabet have to represent the 29 phonemes in modern spoken Turkish (21 consonant phonemes; 8 vowel phonemes). Some of the Latin letters were therefore augmented with accents, giving 29 letters that provide consistent one-to-one correspondences with the 29 phonemes. Thus, the Turkish language has a completely regular and consistent orthography, both when reading and spelling. Children learn to read words very quickly in Turkish (Babayigit & Stainthorp, 2007). They just have to learn the letter-sound correspondences and then by applying these consistently they have a system for identifying the phonemes in each word.

Let us play with using the Turkish alphabet for English.

Use Table 1.3 to translate each of the letters below into the phoneme that it represents.

Table 1.3 Subset of Turkish consonant and vowel letters and phonemes

Letter			
Upper case	**Lower case**	**IPA symbol for phoneme**	**Pronounced approximately as**
B	b	/b/	*b* in ***b****oy*
D	d	/d/	*d* in ***d****og*
E	e	/ɛ/	*e* in *r****e****d*
F	f	/f/	*f* in ***f****ar*
I	ı	/ɯ/	Roughly as *i* in *cous****i****n*
L	l	/ɪ/	*l* in ***l****ove*
O	o	/o/	*o-e* in *m****o****r****e***
Ö	ö	/ø/	*ir* in *b****ir****d*
P	p	/p/	*p* in ***p****in*
R	r	/ɾ/	*r* in ***r****at*
Ü	ü	/y/	*ue* in *cl****ue***
Y	y	/ɟ/	y in **y**es

Dü yü prıfö red o blü?

You should find that it reads as a rather mundane sentence. And, once children have learned the 29 letter-sound correspondences, it is difficult to go wrong. But think about the same sentence spelled conventionally in English orthography: Do you prefer red or blue?

The vowel phoneme is exactly the same in DO, YOU and BLUE, but the spelling is different in each case. As we have said, mapping vowel spellings onto vowel phonemes is one of the major challenges in learning to read words in English.

You may be asking yourself why we don't have a conference in the UK to reform spelling, thereby ensuring that British children could learn to read as quickly as Turkish children. The English Spelling Society has been campaigning for this since 1908, but, although it is beyond our remit to explain fully here, the losses would outweigh the gains. In the 1960s some schools introduced the Initial Teaching Alphabet (ITA) for teaching reading. The ITA alphabet consisted of 44 letters, giving a one-to-one mapping between letters and phonemes, and turning English into a regular consistent language. Downing (1967) showed that this was a very efficient system for learning to read, but it had some drawbacks. Children were surrounded by texts that were not written in ITA. The rich children's literature was not available to them to enable them to practise their developing reading skills because there were relatively few books translated into ITA script.

In the end, ITA died out as a means of teaching reading in schools, but the implicit logic of the instructional method lives on. The instructional method was phonics.

English orthography

ITA, as an artificial system, and Turkish, as a living experiment, are modern, deliberate inventions. English orthography is much older and has been influenced by many different languages and writing traditions. It has developed organically without any strategic planning, so spellings of many sounds now seem to be arbitrary. However, it is important to remember that English is an alphabetic language, and the basis of alphabetic systems is that the phonemes of the language are represented by letters. So however strange and illogical the representations might appear to be, each phoneme in a word is always given a graphic representation. The graphic representation is made up of a single letter or group of letters and is called a *grapheme*.

What is a grapheme?

A **GRAPHEME** is the letter or combination of letters that represents a phoneme in a word.

Each word will have the same number of graphemes as phonemes but there may be more letters than phonemes.

We have used the term *regular* to describe the Turkish system. Regularity here means that the graphemes of a word all represent their most usual, frequent phonemes, so that if one knows the correspondences between graphemes and phonemes one can sound out a word. One can decode the word accurately. Examples of regular words are <DOG> <MAT> <BED> <PIN> and <RUG> which are decoded correctly as /dɒg/ /mæt/ /bɛd/ /pɪn/ and /rʌg/. There is no ambiguity about these regular words. There is one grapheme which consistently represents each phoneme. These words are therefore also said to be transparent.

As we have said, English has a complex orthography and it is difficult to make statements about the system which do not have exceptions. A case in point is the statement above. There is an exception to the rule that there is always a grapheme for each phoneme. The letter <X> generally represents two phonemes as in <BOX> /bɒks/, <FIX> /fɪks/, <EXIT> /ɛgzɪt/.

Reading vowels

There are 20 different vowel phonemes in English. The vowel phonemes in the five words <DOG> <MAT> <BED> <PIN> and <RUG> are all short vowels and, much of the time, the spelling of the short vowels is reasonably regular and consistent, as in these words. However, in the spelling of the more plentiful long

vowels, regularity and consistency tend to break down. Various combinations of letters are used to represent these longer vowels.

Where two letters correspond to a single phoneme, the resulting grapheme is called a *digraph*. The words <SEED> <PAID> and <BOIL> all contain regular vowel digraphs. This means that the vowel letters <EE> <AI> and <OI> are parsed together and represent the single long vowel phonemes /i/ /ei/ and /ɔi/. The box below lists all the vowel letter digraphs (two-letter graphemes) for vowel phonemes.

Two-letter graphemes for long vowels composed of vowel letters

<AI>: sail <AU>: maul

<EA>: bead <EE>: feet <EI>: rein <EU>: eulogy

<IE>: pie

<OA>: coat <OE>: toe <OO>: fool <OU>: out

<UE>: due <UI>: fruit

These graphemes make a sort of sense because they are vowel letters parsed together to represent vowel phonemes. However, a further orthographic device is to create graphemes from a vowel letter and a consonant letter. In this instance <Y> is a consonant letter that has to be parsed with the preceding vowel letter. The full list is given in the next box.

Vowel plus consonant letter graphemes for long vowels

<AH>: rah <AL>: calm <AR>: car* <AW>: saw <AY>: play

<ER>: farmer* <EY>: fey <EW>: sew

<IR>: fir*

<OL>: folk < OR>: for* <OY>: boy <OW>: flow

<UR>: fur* <UY>: buy

Because the consonant letters are not sounded in these graphemes, they are not transparent. They are therefore said to be *opaque*. The existence of opaque graphemes is another reason why it is so difficult to learn to read English words. They are also not consistent, which adds another layer of uncertainty, but we will deal with

that problem later. You will notice that we have placed an asterisk against each of the graphemes containing an <R>. This is because in some accents the <R> is sounded. In these accents, the word <CAR> is pronounced /kɑ:r/ with three phonemes.

A further orthographic device for representing some long vowel phonemes is the use of the 'split vowel digraph' (or 'e-controlled vowel' in the USA). This means that the vowel is represented by two vowel letters. The first one is in the position in the word where it is sounded and the second one comes after the succeeding consonant letter. Examples of this are: <MADE>, <THESE>, <PINE>, <HOME>, <RULE>. This orthographic pattern is fairly regular but not totally transparent. Where there is the spelling <Vowel> – <Consonant> – <E> (VCE), the vowel phoneme can be reasonably reliably predicted in single syllabic words – hence it is regular, but the final letter E is not sounded – hence it is not transparent.

The VCE pattern generally stands for a long vowel phoneme rather than a short one. Though not completely reliable, there is a generality that the split vowel pattern 'makes the vowel say its name'. Indeed, digraphs and trigraphs for vowel phonemes generally code long monophthongs and diphthongs. However, this is not totally reliable. <A> in <APRON> is a single letter grapheme for the long vowel phoneme /eɪ/. And the single letter <O> codes /əʊ/ <GO>. But not always. It codes /u/ in <TO>.

The further we get into describing the orthography, the more exceptions to regularity we find. And it is this unstable predictability which makes English such a difficult language. The split vowel digraph format is a case in point.

There are two highly frequent and important words that do not have a regular pronunciation of the V-E grapheme: <HAVE> and <GIVE>. From their orthography, they should be pronounced to rhyme with <RAVE> and <FIVE>. But English orthography does not permit the letter V at the end of a word. This overrides the phonological representation of the short vowel phoneme /æ/ in <HAVE> and /ɪ/ in <GIVE>. If you are an avid Scrabble player, you might know there are twelve permitted words that all end in <V>. However, nine of these are non-English words: e.g. MAZELTOV; two are recent shortened forms: PERV and IMPROV; and one is a relatively newly coined word: SPIV. A newly coined word is called a neologism. One recent neologism ending in <V> (CHAV) has not yet made it to the Scrabble list.

Beyond the two letter digraphs, English has some graphemes that are even longer, composed of three or even four letters. The next box gives you all the remaining graphemes for vowels.

Remaining graphemes for vowels

<AIGH>: straight <AIR>: fair <ARE>: fare <AUGH>: aught

<EAR>: ear <EAU>: beau <EER>: deer <EIGH>: eight <ERE>: here

<IGH>: sigh

<OUGH>: bought <OUL>: could <OUR>: flour

The peculiar case of <Y>

<Y> is generally classed as a consonant letter, which is rather misleading, particularly for young children learning about letters and sounds, because most of the time it represents a vowel phoneme. Whether it is a consonant letter or a vowel letter depends on its position in the word. In word-initial position, it represents the consonant phoneme /j/ as in <YELLOW> and <YES>. It also appears as a consonant letter in syllable-initial position in a few words: <ROYAL> <LOYAL>. However, as we saw above, it appears in the vowel graphemes <AY> <EY> <OY> and <UY>. It also stands for a vowel phoneme in its own right as in <BY> <MY> <FLY>, or as part of the split vowel digraph Y-E as in <TYPE> and <STYLE>. For each of these words you will notice that <Y> is representing the phoneme /ai/.

Reading consonants

We have so far focused on where vowel phonemes are spelt with at least two letters that have to be parsed together. However, this orthographic device also occurs for consonant phonemes, as can be seen in the last box. Those graphemes shown with a - before them in the next box can only be found in the coda position.

Two- and three-letter graphemes for consonant phonemes

<-CK>: deck

<-DGE>: fudge

<GH>: ghost <GN>: gnat <GU>: guide

<KN>: knee

<-MB>: comb <-MN>: autumn

<-NG>: sing

<PH>: phone

<SC>: scent <SH>: ship

<TH>: the <-TCH>: watch

<WH>: what <WR>: write

And finally, some consonant graphemes are *geminates*. These are double letters. The term comes from the Latin *geminus* meaning twin. The geminate consonants are: <BB> <CC> <DD> <FF> <GG> <LL> <MM> <NN> <PP> <RR> <SS> <TT>. In terms of reading, the doubling of the consonant letters is mostly redundant because they make the sound of the single letter (but they 'preserve' the sound

of the preceding vowel letter when a word is affixed: e.g. <BAT> <BATTING>). However, <CC> sometimes stands for two phonemes /ks/: as in <ACCENT> and <ACCIDENT>.

The presence of digraphs and trigraphs as a characteristic of English orthography is one of the reasons that English is so difficult to learn to read. Children have to learn to recognize these orthographic patterns and also learn that they represent a single phoneme. Because consonants are used in vowel digraphs, there is a lack of transparency, so there is a lot for young children to learn.

Unpredictability

As we have said, the presence of digraphs poses challenges for reading English words because letters have to be parsed together to identify the phoneme they represent. The use of consonant letters as part of digraphs to stand for vowel phonemes is particularly challenging. But digraphs like these are not unique to English. For example, modern German has many. The ones we present here all relate to phonemes which are common to both English and German: the grapheme <EI> stands for the phoneme /ai/; <IE> stands for /i/; <EU> stands for /ɔi/; <SCH> stands for /ʃ/; and <TSCH> stands for /tʃ/. The difference between English and German is that though both languages employ multi-letter patterns for phonemes, the graphemes in German are consistent. It is easy to learn the grapheme-phoneme correspondences because they are predictable.

This is not the case with English: there is a high degree of inconsistency in grapheme-phoneme correspondences. This inconsistency is found much more in vowel phoneme spelling than in consonant spelling, but not exclusively. So we will begin with a consonant grapheme.

The grapheme <CH> usually denotes the phoneme /tʃ/. When children are learning to read they first have to be taught the grapheme phoneme correspondence <CH> = /tʃ/. However, as they expand their reading experiences they will learn that there are a considerable number of words where the grapheme <CH> stands for the phoneme /k/ (e.g. ACHE, CHEMIST, CHOIR, MONARCH, ARCHITECT, CHARACTER, TECHNICAL). You might notice that each of these words has a number of semantically related neighbours and in each of these the <CH> grapheme corresponds to the phoneme /k/ (e.g. CHOIR, CHORAL, CHORISTER). This means that there is additional learning to be done, but this is quite generative. One way of counteracting inconsistency is to learn about word families.

The examples above are of semantically related word families, but inconsistency can be countered by learning word families even when they are not semantically related – just orthographically related. The orthographic pattern <OUGH> will serve to illustrate this.

BOROUGH, BOUGH, BOUGHT, BROUGH, BROUGHT, COUGH, DOUGH, DROUGHT, ENOUGH, OUGHT, ROUGH, SOUGHT, THOROUGH, THOUGH, THOUGHT, THROUGH, TOUGH.

These words are in alphabetic order but we can sort them in other ways.

For example COUGH, ENOUGH, ROUGH, TOUGH and the place name BROUGH all have the orthographic pattern <OUGH> where the <GH> stands for the phoneme /f/. However, the <OU> grapheme stands for either /ɒ/ in COUGH or /ʌ/ in ENOUGH, ROUGH, TOUGH and BROUGH.

In the rest of the words <OUGH> is a vowel grapheme. It can be:

- /əʊ/ as in DOUGH and THOUGH
- /ɔ/ as in BOUGHT, BROUGHT, OUGHT, SOUGHT and THOUGHT
- /u/ as in THROUGH
- /ʌ/ as in BOROUGH and THOROUGH,
- /aʊ/ as in BOUGH and DROUGHT

There is no way of predicting how the <OUGH> grapheme will be pronounced. But it is possible to learn to read these words accurately. As you will see when we cover the development of reading, the way to learn to read these words accurately is to remember them as whole visual units.

<OUGH> is a very obvious inconsistent grapheme and one which is clearly opaque. However, when reading words containing this orthographic pattern, not all aspects are inconsistent or unpredictable. The onsets of all the <OUGH> words are completely regular and transparent. This means that, with a knowledge of the individual letter-sound correspondences, even if the word is not known, a first initial attempt at reading the word can give some useful information: <DR> and <B> at the beginning always represent /dr/ and /b/.

Let us consider some other types of word families. As we have already said, rhyme seems to be very salient to English speakers. Familiarity with rimes can help with developing accurate word reading skills: ATE, BATE, DATE, FATE, etc. form a completely consistent rime family. Playing with word families like this one can be very exciting for young children, and help to increase their vocabularies.

However, because English has such an inconsistent orthography, some apparent rime families are highly unpredictable and inconsistent. The <Vowel> <Consonant> <E> orthographic pattern was completely regular and consistent for the <ATE> rime, but let us consider the <OVE> family. As we said above the VCE orthographic pattern regularly codes long vowel phonemes. This means that if <OVE> were consistent it would stand for /əʊv/. It does in COVE, ROVE, STOVE and WOVE. But that leaves DOVE, GLOVE, SHOVE and LOVE where the vowel phoneme is /ʌ/; and MOVE and PROVE where the phoneme is /u/. This gives us ten base words with the same orthographic rime (-OVE) but only four of these have a regular pronunciation. This makes predicting from rime patterns both helpful but challenging at the same time.

What about the other element of the syllable: the body? Body patterns can also help with learning words. Let us take the coding of the sound /ɒ/ as in the regular word <PLOT>. What about <WHAT>? In words that have the body /wɒ/ or /swɒ/ the grapheme for /ɒ/ is more likely to be <A> than the regular <O>: WAS,

WANDER, SWAN, SWALLOW, etc. Of course you will not be surprised that there are always exceptions: WOBBLE, WOK. This makes predicting from bodies again both helpful and challenging.

Closing words

At this point we hope to have raised your awareness of the level of complexity of English orthography which makes learning to read words more difficult than in other more transparent, regular and consistent alphabetic languages. We have focused almost entirely on words and their sounds. And the examples of words have in the main been monosyllabic and mono-morphemic. You should now have the information necessary to understand about the development of word reading and how this can best be taught. In Chapter 6 we will also cover more complex aspects of words as they are encountered in phrases, sentences and texts. This will involve thinking about words with bound morphemes and complex syllabic structures. Things can only get more exciting. We hope that all professionals working with young children will find English orthography a source of enjoyment and intellectual entertainment as well as a source of frustration. But remember – you have been able to read this chapter – so you cracked the code. It is not impossible.