Word Recognition, Psycholinguistics and Teaching Second Language Reading

Mick Randall
Dr Mick Randall is Senior Lecturer in TESOL and Dean of Faculty of Education at the British University in Dubai. Formerly, he was a Senior Fellow in the English Language and Literature Division at the National Institute of Education in Singapore, and he also taught at University College Chichester in the U.K. He holds a Masters in Second Language Learning and Teaching and a PhD for his study of lexical access in English and Arabic, both from the University of London. Dr. Randall’s current interests include the psycholinguistics of word recognition and reading. His latest publication is Memory, Psychology, and Second Language Learning published by John Benjamins.
Approaches to the teaching of reading over the last thirty years have been couched within a general information processing framework which looks at the integration of top-down versus bottom-up information. Reading has been described in terms of the reader's own knowledge of the world interacting with the information provided by the text. Approaches have emphasised the importance of what the reader brings to the text as well as the information which is derived from the text itself. This has led to more concentration on the importance of meaning than on the lower-level skills of decoding. On a micro level, an important concept is the idea promulgated by Goodman (1976) that reading is a “psycho logistic guessing game,” where the recognition of words and text is determined by using prediction derived from context and grammar, an approach which relies heavily on top-down processes. This concept has played an important role in many theoretical approaches to initially, first language reading, and, by extension, second language reading. It is the purpose of this paper to explore some of the psycholinguistic evidence for the way that word recognition and reading take place in native speaking readers and then to examine these from the point of view of the second language reader.

One of the most generally accepted frameworks for describing the cognitive processes involved in reading is that of Information Processing. The process of extracting information from the written text is represented diagrammatically in Figure 1.

Information processing

There are a number of implications which can be drawn from this general framework. The first concerns the role of the working memory in the process. It is generally accepted that working memory has a limited capacity (Miller’s (1956) famed 7 plus or minus 2 bits of information). Thus there is a limited cognitive space for the use of both decoding strategies and wider schematic knowledge. This is important for the first language reader, but even more so for the second language reader. For all readers there is a high degree of language knowledge involved in the decoding of written language and this will consume cognitive space within the working memory unless there is a high degree of automatisation of these skills (in which case the skills will occupy less cognitive space within the working memory). If the reader’s attention is consumed with the lower-order processes of text decoding, there will be less cognitive space for the use of general knowledge about the wider context. The main mechanism by which meaning is derived from the text is through the content words contained in the text. Schematic knowledge of the world is also operationalised in terms of semantic networks consisting of interconnected webs of words. This then places word recognition at the centre of the reading process. It is the arena where meaning and form meet, and thus it is essential, in any information processing model of reading, for such skills to become highly automatic if any degree of reading fluency is to be achieved.

The importance of content words in the process of reading can be seen in the study of eye movement research. It has long been recognised that fluent readers do not read through texts in a uniform word by word manner, but scan through a text with the eye resting at certain places
in the text and jumping on to others. The resting points are known as fixations and the jumps as saccades. Eye movements are not universally unidirectional with 10-15% being regressive. However, when native speaking readers read texts, the majority of fixations are on the content words rather than the function words and, more importantly, in careful reading of texts it is found that all content words receive attention as testified by the fixation points. Thus, fluent readers pay little attention to function words but derive the meaning from the content words. It is this group of words which ‘drive’ the reading process. The presence of some regressive saccades provides some evidence for the use of a ‘prediction and confirmation’ process implied by the psycholinguistic guessing game model of Goodman, but this is clearly not the main process involved. Readers gain meaning through paying attention to content words (which can be recognised very quickly by skilled readers) and thus rapid and automatic word recognition skills form the basis of successful reading.

If we accept the importance of automatic word recognition skills for successful reading, then we need to consider the processes involved in word recognition and the evidence which exists for different processes involved.

One of the most powerful models of word recognition in English is the dual route theory (Coltheart et al, 2001). Almost certainly due to the orthographical inconsistency of English orthography, it is generally accepted that there exist two separate and distinctive routes by which words are recognised in English; a whole word (lexical) route and a sounding out, grapheme-to-phoneme

Figure 2. An example of fixations and saccades in reading a text (Rayner & Castelhano, 2007).

If we accept the importance of automatic word recognition skills for successful reading, then we need to consider the processes involved in word recognition and the evidence which exists for different processes involved.

One of the most powerful models of word recognition in English is the dual route theory (Coltheart et al, 2001). Almost certainly due to the orthographical inconsistency of English orthography, it is generally accepted that there exist two separate and distinctive routes by which words are recognised in English; a whole word (lexical) route and a sounding out, grapheme-to-phoneme

Figure 3. The Dual Route being used to read a word aloud (after Coltheart et al 2001).
route. The two routes in operation in the process of reading aloud a word can be seen in Figure 3. Evidence for the existence of such routes is derived from studies of dyslexia, and particularly acquired dyslexias where patients have suffered some sort of brain lesion which has left them with a reading impairment. Broadly, there are two types of dyslexics, phonological and surface dyslexics. Phonological dyslexia is a condition where patients can lose the ability to read pseudo-words (e.g. KRAIN or DRIGHT) while retain the ability to recognise all real words including ‘irregular’ words (e.g. TRAIN or PLIGHT). Surface dyslexics, on the other hand, lose the ability, to some extent, to read ‘irregular’ words, but retain the ability to read regular words and pseudo-words (e.g. they can read TRAIN and KRAIN but would struggle with words like LIGHT and FEIGN).

The fact that both of these routes exist in English is obviously the basis of the pedagogic ‘wars’ which have often been fought between the ‘Phonics’ and ‘Look and Say’ approaches to first language reading pedagogy. The fact that both seem necessary to successful reading is, perhaps, the basis for a truce. However, it does raise an interesting issue as to whether both are also important to the second language reader, issues which we shall return to later.

Another recent finding from research into the way that native English speakers read concerns the structure of the syllable and the size of letter ‘chunks’ that readers use to recognise words. In many ways this is a sub-strategy of the lexical, whole word approach. There is increasing evidence that whole words (or whole syllables) are recognised by breaking the syllable into the ONSET and the RIME. Thus ‘catch’ would be ‘chunked’ as:

\[
\text{ONSET} \quad + \quad \text{RIME}
\]

\[
c + \text{atch}
\]

To those of us (like myself) who often struggle with spelling, this seems an intuitively important way to go about trying to remember the different spellings of vowels and their following consonants. If the spelling is analysed along these lines a lot of very useful patterns begin to emerge. Processing of different sized chunks has been dubbed “psycholinguistic grain size” (Goswami, Ziegler, Dalton, & Schneider, 2003) and she and her associates have demonstrated that different languages seem to have different grain sizes, a factor which can be detected in young children learning to read in those languages.

Another issue which has received attention in the psycholinguistics of word recognition is that of letter ‘salience.’ It has long been recognised that not all letters within a word are equally salient. Bruner and O'Dowd (1958) demonstrated using partial anagrams of words such as AVIATOR, that the reversal of letter pairs in different places within words had different disruptive effects on word naming. Thus AVIATOR was more difficult to unscramble than AVIATRO which was more difficult than AVAITOR. This has led to the robust Initial – Final – Medial order of saliency, described by Aitchison (1987) as the ‘Bathtub effect’ (Figure 4).
different types of words and saliency, is the following. Look at the text below and then count the number of Fs there are:

FINISHED FILES ARE THE RESULT OF YEARS OF SCIENTIFIC STUDY COMBINED WITH THE EXPERIENCE OF MANY YEARS

Most readers report only three Fs and it takes some time before all six Fs are seen. One clear explanation for this is that readers do not pay attention to the function words and thus miss the Fs in the function words. They also take in the three occurrences of the function word ‘of’ as whole words and they are not broken down into individual letters. Also, in two of the content words, theFs are in highly salient word initial letter position and the third is in a highly salient initial onset position within the final syllable of the word.

The above represent some of the processes used by native speakers. Most of them relate to the processes used by competent readers who are themselves fluent oral users of the language. L2 readers are in a very different position. They are not necessarily fluent speakers and thus have poor knowledge of the language. Equally important, these readers are often already literate in their first language and thus have developed unconscious strategies for word recognition in their own language. These strategies derive from the intrinsic nature of their own first language or from literacy training in their first language (which, again, may well derive from the orthographic and linguistic structure of their L1). There has often been an assumption in psychological studies that patterns observed are free of linguistic influences (many of the psychological studies are conducted in the US and the UK with English speaking students), but studies are emerging that this may be a false assumption and that there may be very different processes for word recognition in different languages. Therefore, unless we assume that processes of reading and word recognition are common to all languages, we need to examine:

a. the degree to which these processes are common across different languages and
b. how native speaker processes may need to be taught to second language learners.

One powerful indicator that patterns may not be universal can be seen in the study of array scanning (Randall & Meara, 1988; Randall, 2007). Array scanning involves probing the speed by which different letters or shapes presented in arrays can be detected by subjects. In such experiments, subjects are first presented with a single target letter or shape on a computer screen, (e.g. ‘A’) which is then removed and then presented with a five item ‘array’ of letters or shapes (e.g. ‘BSART’) which may or may not contain the target. The subjects are asked to say if the array contains the target or not. By moving the target to different places in the array it is possible to measure how quickly subjects see the target in different array positions. In many experiments carried out with native English speakers, the robust finding was that arrays of letters or digits were seen in quite different ways than shapes. Typical search patterns for native speakers can be seen in Figure 5. What is interesting here is the typical “M”-shaped pattern produced for linguistic targets.
When the same experiments are carried out with native Arab speakers using arrays of English letters, Arabic letters or Hindi digits, an equally robust finding is that such readers always produce the classic "U"-shaped patterns associated with English speakers and shapes (Figure 6). This pattern did not change with increasing fluency and exposure to English. It seemed to be a set pattern with subjects who had had initial literacy training in the Arabic script.

One of the problems with such psychological experiments is that the task has little face validity (the task is a long way from anything we would recognise as reading), and the interpretation of such findings in terms of reading as we know it, is highly speculative. However, it is clear from this that the robust reaction time findings with native English speakers cannot be taken for granted cross-linguistically. Furthermore, what it does show is that unconscious strategies which are generated from different language systems are extremely powerful.

In terms of the search strategies employed by native English speakers, one of the possible explanations of the more rapid recognition of the ends of the arrays could be due to the effects of saliency noted above. English words are marked by spaces and it makes sense that the first

---

**Figure 5.** Array scanning responses for native English speakers (Randall 2007).

**Figure 6.** Array search patterns for Arab speakers with letters (Arabic and Roman scripts), digits and shapes.
thing that a native English reader would do is to isolate words by paying attention to the gaps, and thus the words. If we examine Arabic script, spaces are not the only determining feature of words. Arabic script has spaces both within and between words; spaces being determined by the connectivity restraints of the letters themselves. If we look at a sentence in Arabic and its translation into English, it can be seen that ‘words’ in the Arabic text may have spaces within as well as between them, something which causes the English speaking learner of Arabic considerable problems (Figure 7). In Arabic, the gaps between words are greater than the gaps within words, but words are not as clearly visually distinctive as they are in the Roman alphabet text. I suggest that Arab readers may thus be attending to quite different salient features to extract words from the text than are English readers.

Good, I will tell him the beginning of the first month to the end of the second

Figure 7. A sentence from an Arabic textbook and its translation (note the “gaps” between the letters of the first two words).

As regards eye movement research, Bernhardt (1991) found that foreign language (FL) readers made far more fixations than ‘native speakers’ and held the fixations for longer. She also found that FL readers paused on both content and function words. In many ways this is merely a re-statement of the problem; FL readers need to make more fixations because they do not have the underlying linguistic knowledge within which to place the textual information. This means that they need to fixate on both content and function words. They take longer on each fixation because they do not have the automatic word recognition skills to access the meaning of the word; they are spending cognitive space and energy trying to decode the word. FL readers are not recognising words as quickly as native speakers. On the one hand, they do not have the grammar to predict and thus ignore the function words, but equally they do not have the word recognition skills to rapidly access the content words from the text. What is necessary is for their word recognition skills to be improved by attending to salient features of the words, and in the next section we shall look at the research evidence for some word recognition problems exhibited by L2 readers.

Visual Features of Letters

One of the issues that needs to be investigated is that of visual features involved in recognising letters. Many studies have found processing speeds dependent on fluency and familiarity with orthography (see Bernhardt, 1991). Not all features are equally salient and Arabic and English are very different in terms of saliency. My own personal experience with Arabic, where the presence of a small dot will change the letter, requires a very different search process than that of the Roman alphabet with its high degree of redundancy. Similarly, observation of young Arab learners of English copying all the features of a typewritten font, regardless of their significance, confirms the reverse difficulty that learners have with deciding what is salient in printed Roman letters with all the serifs which are not salient.

Graphemes

Studies with L1 and L2 Lexical Decision Tasks using ‘orthographic neighbours’ (words which are close orthographically) indicate that such effects are present in L1 but not in L2 (van Heuven, 2005). Orthographic neighbours (such as ‘train’ and ‘grain’) will produce interference effects on word decision tasks with fluent native speakers, but not with speakers in their L2. Thus, even good L2 speakers have difficulty in responding to words that are close in the L2, but they can in
L1. Haynes and Carr (1990) compared visual matching among novice and advanced Taiwanese L2 learners of English and L1 English speakers. Results showed a clear contrast between L1 and L2 students. Ryan and Meara found L2 Arab subjects much less sensitive to inserted vowel changes/omissions in English words which they attribute to ‘vowel blindness’ in the Arab students. Arabic does not mark short vowels in normal script and thus students are not used to looking for vowels.

Phonological Decoding

Several studies have found L2 readers attend to graphemic features which override attention to meaning. Support for the use of phonological processing in L2 word recognition is found in a number of studies (Bernhardt, 1991; Koda, 1989; Bialystock, Majumder, & Martin, 2003). Koda found differences in logographic/phonological processing in Japanese. Bialystock et al. found differential phoneme awareness across different literacy languages, with Chinese children performing worse on phoneme segmentation tasks. Current research on spelling errors with Arab students shows problems associated with phonemes as a major factor (Sadhwani, 2007). Onset, Rime and Saliency

Goswami et al. (2003) have demonstrated difference in phonological awareness between L1 Greek and German (low grain size) and L1 English (high grain size) readers reading English depending on psycholinguistic grain size. Meara (1984) demonstrated difference between L1 English and L1 Spanish readers in initial and final letter saliency.

Syntactic and Semantic Knowledge

Studies have shown that vocabulary knowledge is a key factor in increasing reading comprehension, but syntactic complexity of the passage has not been shown to effect comprehension (See Koda, 2005). However, syntactic knowledge of the reader per se has not been investigated except that increasing proficiency has been shown to correlate with increasing comprehension (for a discussion see Bernhardt, 1991).

Implications for the Teaching of Second Language Reading in English

Effective reading is clearly connected to word recognition processes. Approaches to reading and word recognition in psychology have largely been determined by research on users of the Roman alphabet, and, within that, users of English. Cross-linguistic evidence is beginning to show that processes may be quite different in different languages and this is amplified with users of different scriptal systems. Such differences are largely based on the properties of the of the L1 writing system. Therefore, in second language reading more attention needs to be paid to establishing rapid word recognition procedures. Such procedures need to be based on a ‘Cognitive Contrastive Analysis,’ where the two languages are compared, not simply on a linguistic level (as was the case with traditional Contrastive Analysis) but from the perspective of the way that the different language system effects the processing of the language. This is essentially the subject of the growing interest in the field of Cognitive Linguistics.

In practice this means that at least one of the aims of reading texts and the exercises that follow, should be to establish rapid word recognition strategies. This requires an emphasis on word access rather than vocabulary extension. Also, vocabulary exercises should be sensitive to word recognition strategies as well as semantic factors. Word associations should be emphasised to build up associative networks and vocabulary presented and practised in
phrases (not sentences) to establish ‘chunks.’ There should be lots of practice with simple, comprehensible texts.

Rapid word recognition can be enhanced by the use of reading aloud, particularly the teacher reading aloud while the students follow the text. It forces the student to recognise the word quickly, in order to keep up with the reading and it enhances the recognition process by providing dual modality (visual and aural) for recognition to take place. This not only reinforces the grapheme-phoneme correspondences, but it also reinforces the content/function word distinction and effective ‘chunking’ through the teacher’s use of stress patterns.

None of this is rocket science. All of these techniques are within the normal arsenal of the teacher. What it does require is a re-assessment of their use and a re-focusing on the essential problems faced by the second language reader of English who is struggling to read in a different script.

References


