

The Effects of Irregular Orthography on the Processing of Words in a Foreign Language

Andrés Suárez and Paul Meara

Santiago de Compostela University, Spain and Birkbeck College, London University

Research on word-recognition in English suggests that, because of the irregularity of English spelling, readers may make use of both a phonological route, and a direct visual route in matching words to their mental lexicon. In Spanish, on the other hand, with its highly regular spelling system, it is assumed that readers can rely entirely on the phonological method. This paper reports an experiment designed to investigate Spanish speakers' methods in recognising English words. Three hypotheses were entertained: (1) that they would behave like English speakers, using both routes; (2) that they would rely on the direct visual route; (3) that they would rely entirely on the phonological route. The experiment involved subjects' reading four types of words: regular words, regular pseudo words, exception words, and exception pseudo words. The results do not provide conclusive evidence in favour of any of the three hypotheses. The Spanish speakers performed comparatively worse than predicted on all kinds of non-regular words. These results suggest that they might be relying on the phonological route, with a reliance on direct visual access for a few high-frequency items, but more research is needed.

INTRODUCTION

Everybody knows that English spelling is a serious source of difficulty for non-native speakers of the language. Although the importance of the problem is widely acknowledged, there is surprisingly little research into just why orthography should be such a serious source of difficulty, and what psycholinguistic processes are brought into play when people read in an orthography that they cannot process automatically. This paper is a preliminary attempt to look at how native speakers of Spanish internalise the orthographic structure of English, and how it affects their ability to read words in English.

Many models of word-recognition in English assume that there are two basic ways of identifying a word from a written stimulus. Coltheart (1978 and elsewhere) refers to these routes as the phonological route and the direct visual access route. Other terminology is used by other researchers, (e.g. Kavanagh and Mattingly, 1972; Humphreys and Evett, 1985) but the basic idea is very similar, and the arguments in favour of the dual route models are essentially the same, despite the superficial differences. Basically, the argument goes something like this. For many words, it is possible to access the mental lexicon by converting a written form into a phonological code by means of a set of grapheme-phoneme correspondence rules.

Andrés SUÁREZ works in the Department of Educational Research at the University of Santiago de Compostela in Spain. Before returning to Galicia, he worked in the Universidad Complutense in Madrid, and the Colegio Bilingüe Cañada Blanch in London. He is currently looking at ways of assessing vocabulary knowledge in native Spanish speaking schoolchildren. Paul MEARA is Head of the Department of Applied Linguistics at Birkbeck College, London University (London WC1 H0PD). He is best known for his work on vocabulary acquisition in an L2 and has published extensively in this field.

Once a written form has been converted in this way, it is then possible to “look up” the corresponding phonological form in one’s internal lexicon. However, in English, a large number of irregular exception words exist which make this strategy inefficient, and this has led people like Coltheart to suggest that there exists an alternative direct visual access route to the mental lexicon, which does not rely on phonological encoding. Strong evidence for this view comes from the fact that certain types of dyslexia can be accounted for in terms of this model. In particular, so-called surface dyslexics perform as if they can access words via a phonological route, but not via a direct lexical access route, while phonological dyslexics perform as if they can operate a direct lexical access route, but are unable to make use of phonological clues to identify words (cf Patterson, Marshall and Coltheart 1985).

Despite some recent criticisms of this view (e.g. Humphreys and Evett, 1985), the arguments in favour of a dual access approach to word recognition in English are very strong. However, it is not at all obvious that the same arguments apply in other languages. Consider Spanish, for example. Spanish has an orthography which is almost entirely regular. There are, admittedly, a few irregularities. Three letters (B/V/W) all map onto allophones of the /b/ phoneme; H is not pronounced; in some dialects C, S and Z all map onto /s/. However, these irregularities cause few difficulties for spelling, and they do not affect the pronunciation of a word: grapheme-phoneme correspondences in Spanish are almost completely regular (Mata and Cormand 1977). This situation means that it is theoretically possible for Spanish speakers to recognise all words via a phonological route. Direct lexical access is not necessary.

This position is a radical one, of course. It basically suggests that the structure of the mental lexicon in a Spanish speaker might be fundamentally different from the structure of the mental lexicon in an English speaker. English speakers have two formal representations for each word in their lexicon; Spanish speakers may have only one. A slightly less radical position is to argue that Spanish speakers may well develop a direct lexical access route to their mental lexicons, but there is no reason why they should, and if they didn’t then it would be very hard to detect this deficiency in normal every-day life. Some evidence in favour of this second position comes from a study of a Colombian post-graduate student (FE) who is severely dyslexic in English, but who performs nearly normally in Spanish. FE’s performance in both languages is consistent with the idea that he can only access words via a phonological route. In English this represents a severe linguistic handicap; in Spanish, however, the problem has few if any practical linguistic consequences.

FE’s case raises the possibility that all Spanish speakers normally use a phonologically based code to recognise written words. This question is clearly an important one, and we are currently looking into it (cf Meara 1986). However, this

paper is not concerned with how Spanish speakers recognise Spanish words, but rather with how they handle words in an L2 when the L2 uses an irregular orthography. We are assuming for the sake of argument that Spanish speakers who are relatively fluent in English have developed some form of direct lexical access which enables them to operate in English. The question we now wish to ask is how this system of direct lexical access differs from the direct lexical access system that we believe is available to native English speakers.

One of the classical methods of investigating direct lexical access in English comes from Glushko (1979). Glushko asked American university students to read aloud lists of words and pseudo-words. "Regular" lists contained items which were orthographically regular; "exception" lists were made up of real exception words, or pseudo-words that were orthographically very close to irregular exception words. Glushko observed the number of reading errors made by the subjects and their reading latencies. The pattern of results suggested that all the items, whether words or pseudo-words, were read using the same basic mechanism, and that this mechanism depends heavily on a process of analogy.

The question we ask in this paper is how native speakers of Spanish will react in this task. There are basically 3 possibilities worth considering:

- a) Suppose that Spanish speakers normally develop a direct access route to their L2 lexicon. Under these circumstances, we would expect them to behave in roughly the same way as native English speakers do on Glushko's task by using both phonological and visual access.
- b) Alternatively, we could suppose that English orthography presents such enormous problems for Spanish speakers that they actually abandon the phonological route altogether, and come to rely on the direct lexical access route even more heavily than native English speakers do. In this case, we would expect native Spanish speakers to produce relatively few errors in real words, and to produce relatively more errors on pseudo-words. This should show up as an interaction between L1 and word-type, but the pattern should not be affected by the regularity or irregularity of the words.
- c) The third possibility is that Spanish speakers typically do not develop a direct lexical access route, and continue to rely on the phonological coding route, even though they know it is inefficient and unreliable. In this case, we would expect the results of Spanish speakers to show large numbers of errors on irregular words, irrespective of how well they know them, and irrespective of whether the words read are real words or pseudo-words.

EXPERIMENT

SUBJECTS

Twelve native Spanish speakers were tested. All of them were teachers at the Spanish school in London (Colegio Bilingüe V. Cañada Blanch). All of them had an advanced level of English.

In addition, eight native English speakers were tested as a control. These subjects were students in the Department of Applied Linguistics at Birkbeck College, London University.

All the subjects were volunteers.

METHOD

The subjects were required to read aloud a series of lists of words which appeared one at a time on a computer screen. They were instructed to read the words clearly, as quickly as possible, and to avoid errors. Pressing the space bar caused the next word of the list to appear on the screen. This allowed an approximate measure of reading speed to be obtained, but for technical reasons, only the error data will be used in the analysis that follows. Four lists of words were prepared: regular words, irregular words, regular pseudo-words and irregular pseudo-words. Each list contained 30 words. The words used were a subset of the words used by Glushko, which are very tightly controlled for frequency of letters and overall shape. Glushko describes his stimuli as follows:

“The stimuli were generated from a set of 43 monosyllabic exception words (eg DEAF). Exception words are words with different spelling-to-sound correspondences than most words with the same vowel and terminal consonants A regular word (eg DEAN) was selected that differed from the exception only in its terminal consonant Then, using each pair of an exception and regular word, regular pseudo-words like HEAN and exception pseudo-words like HEAF were constructed that differed by only the initial consonants from the base words.” (p679.)

Examples of each type are shown below in Table 1.

Table 1: Examples of stimulus types used in this study

See text for explanation of the categories.

regular word	regular pseudo-word	exception word	exception pseudo-word
BLEED	DREED	BLOOD	DROOD
CODE	GODE	COME	GOME

Subjects were given a practice list of 10 mixed regular and irregular lists before they began to read the experimental lists. Each subject then read the regular words, the exception words, the regular pseudo-words and the exception pseudo-words, in that order. Within each list, the words were randomly ordered. The entire session was tape-recorded, and phonetic transcripts were made from these recordings by an experienced phonetician.

RESULTS

The basic data to be analysed consists of the attempts each subject made at reading the written words aloud. For each subject an accuracy score was calculated, being the number of correct pronunciations produced in the course of each word list. "Correct" here is interpreted to mean the pronunciation given in the OUP *Shorter English Dictionary*. Errors were counted whenever the given pronunciation deviated significantly from this norm. For pseudo-words, we adopted the same criterion as Glushko used: i.e., an error was recorded if the subject's pronunciation differed from the pronunciation predicted by the normal spelling-to-sound correspondences of English. Thus an error would be recorded if HEAF was pronounced to rhyme with DEAF; a pronunciation that rhymed with LEAF would be recorded as correct.

This data is summarised in Table 2 and Figure 1.

Table 2: mean percent words read correctly (rounded to whole %)

	RW	RPW	IW	IPW
English L1	99	97	97	91
Spanish L1	92	87	72	67

These data were subjected to an analysis of variance in which the main effects were L1 (English or Spanish), Word-type (real or pseudo-words) and Regularity (regular or exception words). The analysis revealed a significant L1 effect [Spanish Ss make more errors than English Ss: $F(1,18)=74.5, p<.001$]; a significant Word-type effect [real words produce fewer errors than pseudo-words: $F(1,18)=9.5, p<.01$]; and a highly significant regularity effect [$f(1,18)=53.4, p<.001$]. These results are very much in line with those reported by Glushko (1979), except that our Native Speaker subjects produced relatively few errors in any category, and our Spanish subjects produced relatively high numbers of errors on exception words. Overall, our Spanish subjects produced more errors than the English speakers, and this difference is particularly marked in the case of exception forms, irrespective of whether they are real words or pseudo-words. This shows up as a significant interaction between L1 and Regularity [$F(1,18)=22.9, p<.001$]. None of the other interactions is significant.

DISCUSSION

The main finding to be discussed here is that the native Spanish speakers we tested seem to perform very badly on exception forms relative to regular forms, while still performing at quite a high level overall. This result does not coincide easily with our predictions, and it is not immediately clear how the finding should be interpreted. If the Spanish speakers had consistently got exception words wrong, and if their errors had all been regularisation errors, then it would clearly be possible to claim that they were relying entirely on a phonological strategy for handling these words. In the event, the data is not that clear cut. Performance levels for exception words are in the 70% range i.e. more than two thirds of these items are read correctly. On the face of it, this finding rules out the possibility that the subjects were relying on a phonological strategy, since the majority of these words were read correctly. On the other hand, the data shows very clearly that the subjects do not have a highly developed direct lexical access route: if they did, we would expect the irregular exception words to be much more accurately read than the irregular pseudo-words, since by definition these latter cannot have representations in the mental lexicon. In fact, all exception items were relatively badly handled, irrespective of whether they were real words or pseudo-words. Two thirds of the exception pseudo-words were read as if they were regular pseudo-words.

The Spanish speakers perform worse than the English speakers in all conditions, but their performance deficit is much greater where the items to be read are irregular. It is possible that the relatively poor performance found with irregular real words may be partly due to the subjects not knowing these words, and in effect treating them as pseudo-words. However, all our Spanish subjects were highly fluent in English, and in any case, Glushko's irregular words are in fact considerably more frequent than his regular words (mean frequencies according to the Kucera-Francis norms were 52 and 20 respectively). These two considerations seem to allow us to rule out explanations of our finding based on familiarity.

However, if we compare our results to the results obtained with native speakers of English in earlier experiments, a slightly different picture emerges. Most published studies of this type show that native speakers of English are more likely to make errors on exception words than on regular words. The proportion varies from one study to another; Baron and Strawson (1976) report a ratio of 1 regular error to 9 exception ones; Glushko (1981) reports a ratio of 1 regular error to 7 exception ones; Masterson (1983) reports a ratio of 1 regular error to 6 exception ones. Our data does not fit this pattern. Our Spanish speaking subjects make 1 regular word error for every 3.5 exception word errors. This ratio is quite close to the sort of figures which have been quoted for severely abnormal native speakers of English. Masterson (1985) for instance, quotes three cases of surface dyslexics, whose performance on a task similar to ours produced ratios of errors in regular words to

errors in exception words of 1 to 2.7, 1 to 3, and 1 to 4.25 respectively. This data neatly brackets our data, and suggests that our Spanish speakers are performing more like English-speaking dyslexics than normal English speakers.

To summarize, then, the formal analysis of the error data suggests that our subjects perform not extremely badly, but certainly somewhat worse than we would expect on irregular words. The simplest account of this result is that the subjects rely on a phonological coding mechanism – in much the same way as surface dyslexics do – and that this strategy breaks down frequently when it is applied to exception words. However, this interpretation is confused by the fact that our subjects make relatively large numbers of errors in regular words too. Again, the absolute numbers are not high, and a close inspection of the data suggests that these errors are rule-based, but arise because the wrong rules have been applied.

CONCLUSION

In the introduction to this paper, we suggested that it was possible for Spanish speakers to perform adequately in their native language with a word recognition system that was very different from the word recognition system we believe to be operating in native English speakers. We suggested that native Spanish speakers learning English might have to develop new ways of recognising words if they are to perform adequately in English. The data reported here suggests that Spanish learners of English may continue to rely on a phonological access route, and that a direct visual access route operates only for a few highly frequent words. Clearly the data we have reported here does not force this interpretation, but it does suggest that a more extensive study would be in order.

REFERENCES

- Baron, J. and C. Strawson. Use of orthography and word specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 2(1976), 386-393
- Coltheart, M. Lexical access in simple reading tasks. In: G. Underwood (ed.) *Strategies of Information Processing*. London: Academic Press. 1978.
- Glushko, R. The organisation and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 5(1979), 674-691.
- Glushko, R. Principles for pronouncing print: the psychology of phonography. In: A. Lesgold and C. Perfetti (Eds.) *Interactive Processes in Reading*. Hillsdale, NY: Erlbaum. 1981.

- Humphreys, G. and L. Evett. Are there independent lexical and non-lexical routes in word processing? An evaluation of the dual-route theory of reading. *The Behavioural and Brain Sciences*, 8(1985), 689-740.
- Kavanagh, J. and I. Mattingley. (Eds.) *Language by Eye and Ear*. Cambridge, Ma.: MIT Press. 1972.
- Masterson, J. Surface dyslexia and the operation of the phonological route in reading. PhD Thesis, University of London. 1983.
- Masterson, J. On how we read non-words: data from different populations. In K. Patterson, J.C. Marshall and M. Coltheart. 1985.
- Mata, M. and J.M. Cormand. *Cuadernos de Fonología Castellana para la Enseñanza de la Lengua y Escritura*. Barcelona: Bibliograf, S.A. 1977.
- Meara, P. Psycholinguistic relativity: Is Spanish different? In: V.J. Cook (Ed.) *Experimental Approaches to Second Language Learning*. Oxford: Pergamon Press. 1986.
- Patterson, K., J.C. Marshall and M. Coltheart. (Eds) *Surface Dyslexia*. London: Erlbaum. 1985.