

How does word shape influence visual word recognition?

Jon CLENTON and Joanna MOSEDALE

Research (e.g. Haber, Haber & Furlin, 1983; Healy & Cunningham, 1992; Monk & Hulme, 1983; Perea & Rosa, 2002) suggests that it is the shape of a word that is the most important element in recognition, rather than its constituent letters. The suggestion that a word can only be identified as a whole, rather than from its individual letters, is supported in a range of studies (e.g. Cattell, 1886; Reicher, 1969; Wheeler, 1970). Such studies usually stem from the use of a Lexical Decision Task (LDT) in a variety of different formats in order to judge whether words are genuine or not. Formats can relate to size judgements (e.g. Perea & Rosa, 2002), in which letters are presented in alternating size. Similarly, studies (Lété and Pynte, 2003; Content, Mousty & Radeau, 1990) explore the relative frequency of word shape, suggesting that words with a higher frequency shape might be recognised sooner than less frequent alternatives. Alternatively, LDTs have assessed recognition based on the ability to identify items on the basis of letters (inner or outer) being removed or substituted (e.g. Beech & Mayall, 2005; Healy & Cunningham, 1992; Monk & Hulme, 1983; Perea, Comas-Forgas, Soares, & Moret-Tatay, 2012). Such studies appear to demonstrate that LDTs can take a variety of different shapes and format, and support the notion that a word shape appears to dictate recognition.

There are, however, a number of studies that suggest that word shape studies need examining in greater detail. Monk & Hulme (1983) imply that data from mixed-case (i.e. mIxEd CaSe) stimuli counters their own deletion or substitution evidence. Similarly, Paap, Newsome, & Noel (1984) suggest letter effects might account for what some studies describe as word shape effects. Others (e.g. Grainger, 2008) contend that word shape effect studies need examining in far greater detail, to account for the fact that the average reader knows tens of thousands of words.

A key argument in word recognition studies is whether it is the shape of the word as a whole (the envelope), or the features of the individual letters that readers use when reading a word. Studies (e.g. Healy, 1981; Healy, Volbrecht & Nye, 1983; Johnston & McClelland, 1980) appear to propose a hierarchy of features readers use in experiments, with readers mainly using the envelope of a letter to identify it and then secondarily

focusing on additional visual features. More recent studies have also looked at the effects of deleting midsections of letters compared to vertices (corners or intersections) and find that the deletion of vertices has a more detrimental effect on letter recognition (e.g. Lanthier, Risko, Stolz &, Besner, 2009). This latter idea suggests that the envelope and additional visual features are processed in parallel, but readers give a greater weight to the envelope than other features and can therefore use a “sophisticated-guessing decision rule” to guess the word based on the envelope and can tolerate missing features because of this (Healy et al., 1983, p. 528).

With regard to the word shape hypothesis (WSH), Monk and Hulme (1983) carried out a proofreading task where alterations were made to words by either substituting or deleting letters. They found that alterations to word shape were noticed more often than alterations that preserved word shape. In a similar task, Healy and Cunningham (1992) found that all subjects were sensitive to changes in word shape, which they define as an ascender or descender, rather than a neutral letter, being deleted. Haber, Haber &, Furlin (1983) also found proof for the WSH in their experiment in which students were asked to read a passage that ended mid-sentence and were then given a clue for the word that followed. When the clue showed the shape of the word in terms of ascenders, descenders and neutral letters (and length), the chances of guessing the word correctly greatly improved. By way of contrast, Paap, Newsome and Noel (1984) found that participants tend not to detect changes in word shape. They found that substituting one ascender with another is only likely to be detected if it also changed the letter distinctively e.g. changing that to tdat, where word shape is maintained, but h has not been substituted with a potentially confusable letter, is more likely to be noticed than changing that to tbat, therefore asserting that word shape is not what is having an effect on recognition here. In this study, as well as in Monk and Hulme’s (1983) study, Healy and Cunningham (1992) argue that Monk and Hulme’s experiment involved deletions, rather than substitutions, ensuring that the changes in word shape could not be localised at the letter level, because a deleted letter is not able to influence the letter identification process. Monk and Hulme’s (1983) experiment, they argue, was therefore reliable, whereas word shape was not maintained through mutilations involving substitution and the effect of word shape on results in such experiments might therefore be ambiguous. Lété and Pynte (2003) used a computerised word pool (Content, Mousty &, Radeau, 1990) to calculate the frequency of a word’s shape and found that

participants responded faster to words with a low-frequency shape, than those with a high-frequency shape. Having a rare shape speeds up lexical decision making times for lexically low-frequency words. It was also found that lexically high-frequency words benefit more from the priming procedure than do low-frequency ones. This suggests that shape information affects word recognition, deviating from Paap et al's (1984) results. Lété and Pynte argue that in their study, the "physical characteristics...were more advanced [than Paap et al.]" (1984, p. 925) as the sizes of the letters were larger and the ascenders and descenders extended higher or lower above or below the line. Lété and Pynte's (2003, p. 931) second experiment also provides evidence for the WSH. Participants were primed with the shape of a target word before being given the target in an LDT. The results show that when the shape of the upcoming target was primed, "participants rejected legal nonwords more rapidly and with more accuracy than illegal nonwords." However, as Lété and Pynte (2003, p. 931) point out, LDTs do not "necessarily involve the completion of lexical access", as would a naming task, for example.

More recently, researchers (e.g. Forster, 1976; McClelland & Rumelhart, 1981; Morton, 1969; Murray & Forster, 2004; Rumelhart & McClelland, 1982) have mostly found evidence against the WSH and many of the models they have developed to explain visual word recognition are 'analytical' models, which propose that words are initially formed from individual letters, rather than an envelope. Models of word recognition generally fall under the category of either parallel or serial. In parallel models all letters in a word are processed simultaneously, whilst in serial models, letters are processed individually. Morton (1969) developed the parallel word recognition model; the Logogen model. This model describes a mental dictionary containing all the words in an individual's vocabulary. The individual receives one or more pieces of input (e.g. visual, auditory, sentence context). A word is recognised when enough information has been received from the input to reach the 'recognition threshold'. This threshold varies as "common words require less information to make them produce an output" (Morton, 1969, p. 260). This explains the frequency, but not neighbourhood effects, both of which are discussed below. This model accounts for the frequency effect as the more exposure a person has had to a word, the lower the logogen threshold will be for it. However, the model cannot account for the neighbourhood effect as there is no interaction between the different recognition units. The interactive activation model of word recognition is similarly a parallel model and was developed by McClelland

and Rumelhart (1981) and Rumelhart and McClelland (1982) and consists of three levels of input: visual features (i.e. vertical line, curve etc.); individual letters and whole words. This model accounts for the word superiority effect identified by Cattell (1886) and demonstrated by Reicher (1969) and Wheeler (1970) as the letter unit in a whole word receives activation from two sources, but only from one source if on its own.

Grainger (2008) refers to the interactive activation model of word recognition and claims that the WSH cannot account for the problem of the invariance in the shape. Grainger argues that as the average skilled reader knows between 30,000 and 50,000 words, “the computation involved in solving shape invariance for each word is going to be a lot more costly than solving it for each letter of the alphabet” (2008, p. 3). Supporters of the interactive activation model account for the word superiority effect by saying letters presented in a word stimulus benefit from an additional activation input compared with letters presented in a non-word context. The Search model (Forster, 1976; Murray & Forster, 2004) is a serial search model, in which, in contrast to parallel models, words are checked in the mind against a candidate set (which depends on the task at hand) of words one at a time until the correct one is found. As high frequency words are checked first, this model accounts for the frequency effect. Lively, Pisoni & Goldinger (1994) criticise this model as they say making such a large number of serial comparisons would take longer than it takes to recognise a word. Forster (1994) later addressed this by suggesting that all the ‘bins’ that are searched in his model, are searched simultaneously. The model also does not account for neighbourhood effects very easily, as it would be expected that if a reader has to ‘search’ through a lexicon, having a large neighbourhood would mean it would take longer to search through all of the words in that neighbourhood, slowing down the word recognition process, which has been found to be the opposite of how neighbourhood size affects recognition (Andrew, 1989, 1992). In terms of a word’s shape affecting recognition, serial search models suggest we look at one object at a time, and therefore do not agree that words are recognised by their envelope, but by their individual letters. Conversely, Healy (1981) describes a ‘hierarchical feature test’ that proof-readers employ which involves giving highest priority to resolving letter envelope and second to other visual features of letters. This is supported by Healy and Cunningham (1992), who carried out two experiments and argued that the results showed that word shape does play a role in word recognition. On their first experiment involving proofreading passages containing words

with deleted letters, they argue that: “the differences between the two types of deleted letters [ascenders or descenders] cannot be attributed to their location in the word, their pronunciation, the length or frequency of the words containing them, or any other linguistic variables not pertaining to their visual features, because the pattern of proofreading errors differs markedly for lowercase and uppercase print.” (1992, p. 147).

In addition to arguments that word shape plays a large part in visual word recognition, researchers have found that other characteristics of words also play a role. Almost every study relating to word recognition has found effects of word frequency, neighbourhood size and context, discussed below. Whilst it is the effects of word shape the current study focuses on, most studies looking at word recognition take these into account, and so it is important to understand their effects. It has been argued that of all the factors that influence performance in visual word recognition tasks, the frequency of a word, i.e. how familiar a word is, appears to have the largest influence on reaction times to a stimulus (Beech & Mayall, 2007), and that “frequency effects have been found in just about every task that could reasonably be classified as a lexical processing task” (Murray & Forster, 2004, p. 721). Murray and Forster further argue that although other stimulus variables i.e. length, regularity, homophony and neighbourhood density, affect recognition time, they “appear to do so only for a restricted range of frequencies or for some tasks and not others”. It has been found that word shape and word frequency have an effect on each other in that the shape of a word can have more of an effect on recognition of high-frequency words (Lété and Pynte, 2003; Perea & Rosa, 2002). Beech and Mayall (2007) carried out an experiment involving priming participants with either the inner or outer features of a word, before having to guess the word being shown. They found that there is a difference in response times for outer and inner primes i.e. the word shape, and the word frequency, namely that in the case of high-frequency words, outer primes led to significantly faster reaction times compared to inner primes, whilst they do not find this with the low-frequency words. Perea and Rosa’s (2002) experiments also find that the effect of word shape is different for high and low-frequency words, but found the opposite to Beech and Mayall (2007). They found “the effect of size alternation was greater for low-frequency than for high-frequency words in a lexical decision task” and that “the effect of case type (lowercase vs. UPPERCASE) occurred for low-frequency words, but not for high-frequency words.” (Perea & Rosa, 2002, p. 785). The current study, therefore, uses control for word frequency, ensuring all words

are within a similar range on the Zipf scale (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014; and, Zipf, 1949). The Zipf scale is a modern measure of word frequency that looks at a corpus of 51 million words and measures how often a word occurs in this corpus. It assigns a number between one and seven to each word based on where it is on a Likert scale of how often it appears.

As with word frequency effects, the neighbourhood size of a word, although not affecting shape, does have an effect on recognition. The neighbourhood size refers to the number of words that can be made by changing all but one letter of the target word, with all the letters staying in the same position (Andrews, 1989) e.g. neighbours of the word *cat* include *bat*, *cot* and *car*. It has been found that a large neighbourhood size facilitated speed of recognition in LDTs of low-frequency words, but not high-frequency words, but showed inhibitory effects for other tasks such as progressive-demasking (Andrews, 1992; Carreiras, Perea & Grainger, 1997). Pollatsek, Perea & Binder (1999) also found that the frequency of the words in the neighbourhood has an impact on recognition, with words with high-frequency neighbours being recognised more slowly. As shape is the main focus of the current study, it is difficult to carefully control for shape, frequency and neighbourhood effects, when manipulating shape, but neighbourhood size will be taken into consideration and controlled to a small extent. In addition to manipulating words using deletions and substitutions, as have been discussed and are used in the current study, other studies have manipulated the shape of words in other ways. To determine the effect of word shape on recognition, studies have manipulated aspects including size (Perea & Rosa, 2002) and case (Fisher 1975; Paap et al., 1984) and have focused particularly on the effects of ascenders, descenders and neutral letters (Beech & Mayall, 2007; Lété & Pynte, 2003; Monk & Hulme, 1983). Several studies focusing on the effect of word shape on visual word recognition have looked at how manipulating the case of written words affects recognition. This technique is used as it “distorts the characteristic shape of words printed in lowercase and, therefore, should disrupt performance to the extent that the word-shape feature normally contributes to word recognition” (Paap, Newsome & Noel, 1984, p. 415). Indeed, a number of studies have found evidence that manipulation of word case affects the efficiency of readers in various tasks. Fisher (1975) found in two experiments on reading and searching respectively, that alternating between upper-and lowercase letters (*aLtErNaTiNg*) most significantly decreased efficiency, followed by all letters in uppercase. Other studies provide

more evidence for this phenomenon (Mayall & Humphreys, 1996; Monk & Hulme, 1983). Several studies have also found that when text was typed in alternating sized letters (alternating), regardless of case, the efficiency of speed of reading and searching for targets significantly decreased (Perea & Rosa, 2002; Smith, 1969 and Smith, Lott & Cronnell, 1969). This technique of alternating large and small letters, rather than alternating between cases removes the disruption to normal text processes that may occur with case alternation. Others have argued (e.g. Mayall & Humphreys, 1996; Mayall, Humphreys & Olsen, 1997; Oden, 1984) that alternating between cases in word shape experiments is not necessary as case-mixing is not natural in everyday written text and might disrupt normal reading processes. The current study therefore does not include case-mixing as it has already been agreed that it does affect word recognition, but is not a natural occurrence in text. It is however interesting to note the effect, as it is evidence for the WSH. We begin with two central hypotheses:

1. Subjects will recognise genuine words more quickly than non-words.
2. The responses in terms of reaction times and correct decisions will be in the order of (from slowest/least errors to quickest/most errors):

Set 1: A letter substituted for another letter with no similarities in shape e.g. a neutral to an ascender

Set 2: A letter substituted for another letter with a similar shape e.g. an ascender to an ascender

Set 3: A letter substituted for a similar, easily confusable letter e.g. e to o

Set 4: Two adjacent letters swapped over

On the basis of the literature presented above, the current investigation ask the following three research questions:

1. Out of the four different criteria for non-words (see hypotheses), which has the biggest effect on recognition times of words in an LDT?
2. Out of the four different criteria for non-words, which has the biggest effect on error rates in an LDT?
3. How does word shape influence processing times in an LDT on visual word recognition?

Method

The experiment was carried out in a quiet room with one participant at a time. The test was created using the PsychoPy software (Peirce, 2009), which presented the stimuli to the participants and measured their reaction times. The stimuli were presented on a computer screen with a resolution of 1366 x 768 pixels, and the text was at a height of 2, which refers to the height relative to the screen dimensions and was found to be a clear and easy to read size in the pilot studies. The stimuli were presented for 300 ms in white Ariel font on a grey background and positioned in the middle of the screen, as it was found in the pilot studies that black on a white background was too hard to read because of the high contrast as the words flashed up so quickly. During the test, the screen was completely taken up by the test and so there were no distractions.

The test and what participants needed to do was explained before it began, and participants were instructed via text on the screen before the test began, as to how to respond to stimuli and told to react as quickly as possible. The test began once they pressed the space bar, after reading these instructions. The first five stimuli presented were trials, as pilot studies showed that it took about three or four responses before participants got used to the test. Participants were required to press the f key for stimuli they believed to be words and j for those they thought were non-words. These keys were found to be the easiest for participants to use in this type of experiment in the pilot study as they are on the same row near each other but not right next to each other. The next stimulus was presented as soon as participants had made their decision. The words and non-words from each of the four sets were mixed up randomly but the stimuli were presented in the same order for each participant. The session lasted approximately 10 minutes.

Subjects

Thirty-one participants over the age of 16 took part in the experiment. All had at least an average score in a speed-reading test (ReadingSoft, 2012) and were first language (L1) users of English.

Materials

Four sets of words and non-words were used in the experiment. Each of the four sets contained seven words and seven non-words. All of the word and non-word stimuli were

five letters in length and the non-words were created by mutilating the interior letters of the words in some way, as the interior letters affect the envelope of the word to a larger extent. All of the words were between 5.01 and 5.9 on the Zipf scale of word frequency (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014) and had a neighbourhood size of between 2 and 10 words (Davis, 2005). It is important to control for word frequency and neighbourhood size as they have been found in many studies on word recognition to have a large effect (see e.g. Beech & Mayall, 2005; Murray & Forster, 2004; Perea & Rosa, 2002). All of the words, apart from those in Set 4 (see below) were orthographically and phonologically possible in English. All of the four sets were created by mutilating letters of these five letter high-frequency words in different ways.

Set 1: One letter was substituted for another with no similarity in shape e.g. a neutral for an ascender, or an ascender for a descender etc. as in sorry becoming sorby.

Set 2: One letter was substituted for another with a similar, but not easily confusable shape e.g. an ascender for another ascender, as in stuff becoming shuff.

Set 3: One letter was substituted for another very similar, easily confusable letter e.g. t to l or e to o where they have the same exterior (see Beech and Mayall, 2005), as in along becoming atong.

Set 4: Two adjacent interior letters were swapped over as in other and ohter. Unlike the other sets, the non-words in Set 4 were not orthographically or phonologically possible English words.

Results and Discussion

The reaction times of each individual for each word or non-word in the experiment, as well as whether or not their response was correct, were automatically recorded in milliseconds once the experiment was over. The reaction times for each word, then each set of non-words, was calculated. Errors made in response to words i.e. responding to a word as a non-word, were not included in the calculation of the averages. The average reaction times and number of errors has been calculated throughout analysis using the mean, as this is considered the most accurate measure of central tendency, although any unusually large or small numbers can distort the answer.

Overall, there were 164 errors made out of 1,736 responses, which means that 9.4% of all responses were errors. In Perea and Rosa's (2002) experiment, 5.7% of responses were

errors and in Perea et al. 2012) only 3.8% and 3.9% of the responses in their two experiments were errors. This suggests that in the current study, more errors were made than would be expected in an LDT focusing on word recognition. The reason for this is not clear, and no literature condemns a high percentage of errors, although reactions involving errors were not counted when calculating average recognition times. On average, each participant made 11 errors in their responses, although the number of incorrect responses varied greatly between individuals from 0 to 21 errors. As with the reaction times, there was a difference in the number of errors made on words and non-words. There were significantly more errors made on the non-words than on the words, with a difference of at least two errors. The study hypothesised and the results demonstrated that words are recognised more quickly than non-words, as has been consistently found to be true in other studies. For all four sets of words, the average reaction time was quicker for words than non-words, with around a 100ms difference between the average reaction time for words (595ms) and non-words (694ms). It was expected that there would be a significant difference between the average reaction times of the four non-word sets, but there was only 27ms between the sets with the fastest and slowest reaction times.

We hypothesised that the responses in terms of reaction times and errors, from quickest to slowest and most to least number of errors would be: set 3 (a letter substituted for a similar, easily confusable letter); set 2 (a letter substituted for another letter with a similar shape); set 1 (a letter substituted for another letter with no similarities in shape); then set 4 (two adjacent letters swapped over). This is because set 3 made the smallest difference to the shape of the original word and so according to the WSH, should be reacted to more quickly as participants will believe it is the corresponding real word, but therefore make more errors, with the opposite being true for set 4. The results however did not match this hypothesis. The number of errors and the reaction times did not correlate with each other in terms of the four sets of words. The reaction times from fastest to slowest were set 1, 4, 2, then 3, and the number of errors from least to most was set 4, 3, 1, then 2.

The results for the individual sets are outlined as follows.

Set 1: as the mutilations to the non-words in set 1 produced the second biggest change in shape out of the four sets, it was hypothesised that these words would produce the second slowest reaction times and second least number of errors. The results show that the non-words in this set actually produced the second most number of errors and the fastest

reaction times.

Set 2: whilst it was hypothesised that the non-words in set 2 would produce the second quickest reaction times and second highest number of errors, they actually produced the second slowest reaction times and most number of errors. The non-words in set 2 were mutilated to some extent, but did not change the shape drastically or completely retain the shape and the results did not differ hugely from the hypothesis, so cannot be used to draw a conclusion on the WSH as strongly as can sets 1 and 3. However, the fact that set 2 did not come between sets 1 and 3 based on reaction times and number of errors, despite coming in the middle of these three sets in terms of changes to shape, goes against the theory of word shape having an effect on visual word recognition.

Set 3: as a result of the shape of the words and non-words in set 3 (a letter substituted for a similar, easily confusable letter) were so similar, it was hypothesised, under the WSH (e.g. Cattell, 1886; Haber, Haber & Furlin, 1983) that participants would make the most errors on these non-words, as they would mistake them for words. We also hypothesised that the non-words in this set would produce the fastest reaction times as they look very similar to real, high frequency words. In terms of reaction times, the results go completely against the hypothesis, as the non-words in set 3 produced the slowest reaction times (707ms). This is evidence against the WSH as it took on average 104ms longer for participants to make a decision on non-words in this group than their corresponding words. If participants were using the shape of the word above anything else to recognise words, the difference in reaction times between the words and non-words in this group should have been a lot smaller as the shape changed so little. The average number of errors made in set 3 also went against the hypothesis and produced the second fewest number of errors (5.71) of all sets, just slightly fewer than set 4 which produced the least number of errors (5.57). This suggests that the shape of the word did not confuse the participants as much as other aspects of the non-words, into thinking that they were their corresponding real words.

Set 4: swapping two adjacent letters over as in set 4 creates the biggest change in word shape of all the four sets of non-words and it was therefore hypothesised that participants would make the fewest errors and react slowest to these words, as it is most obvious that they are not real words and non-words are always reacted to more slowly. Additionally, some are not phonologically possible in English. The number of errors made matched the hypothesis, with participants making the most number of errors, which was on

average 7.71 errors, but did not produce the slowest reaction times, although there was only 17ms difference between set 3 (which produced the slowest reaction times) and set 4. We might argue that this is evidence against the WSH, as the set which changed the words' shapes so dramatically did not produce the fastest reaction times. However, it did not support the hypothesis in terms of number of errors, although this might have been because some of the words were not phonologically possible in English and were therefore more obviously non-words. Because this set of words, unlike the others, does not gradually change the shape of the words, it is not as clear whether or not it is the shape affecting recognition times.

Individual words: as well as comparing the word/non-word sets, there are also some interesting results from individual words. Although there were not always significant differences between the average reaction times of the whole sets, some of the individual words produced significantly much slower reaction times than other words. All of the words which were reacted to most quickly were real words, whilst all but five of the 20 words reacted to most slowly were non-words. We began by hypothesising that the non-words in set 4 would all be on the list of the slowest words, but there are in fact words from all four sets. The non-word *locel* was in set 3 and therefore did not change the envelope of the shape, but produced the slowest reaction time of any non-word. One reason for this could be because it looks so much like *local*, but it is obvious that there is something wrong with it and it took participants longer to make a decision. Although only a single word, this strongly goes against the WSH, as according to that, readers would see the shape and assume it is the word *local*, which is a high-frequency word and so would react quickly to it. Nine people did however make an error on this word. The non-words *birng*, *clær* and *watns* were the only non-words from set 4 to appear in the list of the 20 words/non-words that produced the slowest reaction times. It might be assumed that more words from this set would appear as they make the biggest change to word shape and are most clearly non-words, as many are not phonologically possible. This could be why they were reacted to more quickly; because it is more obvious that they are non-words. *birng* however, is possible in English, which could explain why it was reacted to more slowly. This goes against the WSH as it seems that other factors such as whether it is legal in English are more important. (see tables 1 -3 below).

Table 1. Fastest Average Reaction Times of All Words/Non-Words

Word/non-word	Average reaction time (ms)
clear	503
hello	505
start	522
sorry	543
along	543
found	547
small	549
watch	549

Table 2. Slowest Average Reaction Times of all Words/Non-Words

Word/non-word	Average reaction time (ms)	Non-word set
taben	660	3
stage	661	-
rould	663	1
given	668	-
sorby	675	1
spaze	683	2
goven	685	2
short	697	-
watns	722	4
plyce	730	1
sbort	731	3
while	734	-
shuff	735	2
price	744	-
poice	754	2
claer	775	4
spart	784	1
pover	795	2
staqe	802	3
birng	808	4
locel	830	3

Table 3. Most Number of Errors Made Overall on Non-Words

Number of errors	Word	Non-word Set
7	Atong	3
7	Walch	3
7	Bring	4
7	Clear	4
8	Plyce	1
8	goven	2
8	spaze	2
9	Nirth	2
9	Shuff	2
9	Local	3
9	Other	4
10	Poice	2
13	Spart	1

We conducted an ANOVA on the data with reaction time as the dependent variable and non-word set as the independent variable. The ANOVA on the data showed no significant effect for change of set [$F(3,24) = 0.189, p > 0.05, MSe = 0.005$]. 'F' is the ratio of the variance between groups to within groups, 'p' is the significance and MSe is the mean squared error and is calculated by dividing the sum of squares within groups by the number of degrees of freedom (number of degrees of freedom are calculated by subtracting one from the number in each group - in this case there were seven words, minus one, in each group - and multiplying by the number of groups - in this case four, as there were four sets of words). The significance was 0.903. We also conducted an ANOVA with the number of errors as the dependent variable and non-word set as the independent variable. The ANOVA on the data again showed no significant effect for change of set [$F(3,24) = 0.981, p > 0.05, MSe = 6.893$]. The significance was 0.418. As neither the reaction times nor the number of errors of the participants were found to be significantly affected by the different sets, it might be concluded that in this study, word shape did not significantly affect participants' reactions.

We now present our discussion in light of our three discussion questions.

1. Out of the four different criteria for non-words, which has the biggest effect on recognition times of words in a lexical decision making task?

Set 1, where a letter was substituted for a similar, easily confusable letter, produced the slowest reaction times of all of the four sets and therefore had the biggest effect on word recognition. However, there was only a difference of 27ms between the sets with the fastest and slowest reaction times, suggesting that the four different criteria did not have a big effect on recognition times.

2. Out of the four different criteria for non-words, which has the biggest effect on error rates in a lexical decision making task?

In this particular experiment, the error rates produced more varied results than the reaction times and so can be used more effectively to compare the four sets of words. Set 2, where a letter was substituted for a letter with a similar, but not easily confusable shape, produced the most errors and therefore had the biggest effect on error rates. The average error rate in this set was 7.71, compared with 5.57 in set 4, a difference of 2.14, whilst there was a difference of 1.43 errors between the set of real words which produced the most errors and the set which produced the fewest errors. This is quite significant, as there should be little or no difference between the sets of real words.

However, out of the 31 participants, 7.71 making an error on at least one of the words in set 2 is significant and shows that the participants were more likely to make errors on non-words than words, perhaps because they were using the shape of the word to make their decision. If the participants were relying most heavily on the shape, however, it would be predicted that most errors would be made on the non-words in set 3, as these preserved the shape of the real words the most. There was a difference of 2.57 errors between the words and non-words in this group, which is quite significant, but was not the largest difference (which was 4.85 in set 2).

3. How does word shape influence processing times (in a lexical decision making task on visual word recognition)?

The overall results suggest that it is not a whole word's shape, or envelope, that is most important to readers in word recognition. The fact that the key change in all of the mutilations affected the shape of the words and that these non-words produced slower recognition times than the real words strongly suggests that the shape of words is important in recognition. It has been found in countless other studies that non-words produce slower recognition times than words (see, e.g. Cattell, 1886; Warren, 2013; Whitney, 1998), as readers process the word and have to search for it in their mental lexicon, which

takes longer if it is a non-word as all words - high and low-frequency - must be searched before rejecting the non-word. It could therefore be argued that that is what has happened in the current experiment. However, because the only changes to the real words to turn them into non-words was a change in shape, the difference in reaction times could be said to be evidence that shape is at the top of the hierarchy of features readers use in visual word recognition.

In terms of the results of the individual sets of non-words, these do not support the WSH. The fact that there was virtually no difference in the reaction times between the four sets suggests that the WSH is wrong, as the differences in shape do not produce differences in reaction times. However, the individual sets did produce slightly more significant differences in error rates. These error rates, however, did not match the hypothesis of what would be expected if it was the word's shape having an effect either; that the biggest change in word shape would produce the fewest errors and vice-versa. Although there was some difference between the average reaction times and error rates on the four sets, they were not in the order that was hypothesised and so this cannot be evidence for the WSH.

Conclusion

Different types of tasks have had different results concerning the effect of a word's shape on recognition. For example, Lété and Pynte (2003) found that the lexical-frequency effect is much larger in an LDT than in a naming task. Lexical decision making tasks do not require lexical access to the stimuli, only recognition of the words. It could therefore be argued that unlike other tasks used to investigate word recognition, LDTs do not provide results that are relevant to 'real life', as in real life, readers have to access the meaning of a word, and it is not enough to simply know whether or not a word is a real word. The majority of studies carried out on the effect of shape on visual word recognition have used LDTs, but it is worth noting that there has been some argument over the validity of using such tasks and some discrepancies found between types of task.

One limitation to consider is the argument that word shape is not maintained through mutilations involving substitutions, as in the current study, as the results are ambiguous (Healy & Cunningham, 1992). It has also been found that changes in word shape are only likely to be detected if the letter distinctively changes (as in set 1) (Paap, Newsome and Noel, 1984), which would suggest that the shape of the word is not changing in any of the

other sets. Although these are only theories, the fact that the results in the current study found no correlation between the error rates and reaction times is a limitation as it makes the results harder to interpret. It was expected that the slower the reaction time, the fewer errors would be made, as participants would react more quickly to a word they thought they knew, but would make more errors, as the shape of the word would confuse them. The results have been discussed in line with the original hypotheses, which are based on other studies and theories, but because there is so little difference between the reaction times of the four non-word sets, no correlation between error rates and reaction times and the idea that word shape is not maintained through mutilations, the results could be said to be ambiguous. This is confirmed by the lack of significance found on the effect of word shape on error rates and decision-making times in the statistical analysis, showing that there is not sufficient evidence to support or reject the hypotheses.

The aim of this study was to determine the extent to which a word's shape plays a role in visual word recognition. It expected to support the WSH, in that it is the envelope of a word, rather than the individual letters that readers use to recognise a word. The hypothesis that the bigger the change in a word's shape, the slower the response time and the fewer the number of errors was proposed as a result of a large number of past studies finding that word shape has a large effect on recognition (see e.g. Beech & Mayall, 2007; Healy, 1981; Healy & Cunningham, 1992, Lété & Pynte, 2003). However, the results found in the current experiment tend to contradict these findings, and suggest that it is not the whole shape of a word that readers use. Overall, there was very little difference between the recognition times of the participants for the four sets of non-words and the error rates did not match the hypothesis that the bigger the shape mutilation, the fewer the number of errors. The results of the participants' average reaction times were barely significant as the difference was so small, but this in itself is evidence against the WSH, as it would be expected that there would have been a much more significant difference between the four sets, if word shape was what was affecting recognition. In terms of models of word recognition, the current study supports the interactive activation model and the serial search model, both discussed in the literature review. Both of these models argue that readers do not recognise whole words, which has been found in the results from the current experiment. The results further suggest that the hierarchical feature test proposed by Healy (1981) is not correct, as it appears that the participants did not give highest priority

to resolving the words' envelopes. There are many studies that focus on the effects of word length (New, Ferrand, Pallier & Brysbaert, 2006), word frequency (Beech & Mayall, 2007; Murray & Forster, 2004) and neighbourhood size (Andrews, 1989, 1992). The fact that these were all controlled in the current study, and the only change in all of the non-words was substituting one letter from a real word with another to affect the shape of the word as a whole, suggests that word length, frequency and neighbourhood size are all more important in word recognition than shape, as there was virtually no change in the reaction times between the non-word sets. Although there were some differences in the error rates, they were not what would be expected if it was word shape that was having the biggest effect on recognition.

On the basis of the arguments over whether or not LDTs can be used to test the effects of word shape on word recognition, it would be useful to carry out a similar experiment where the non-words are manipulated in the same way, but using a naming task. Monk and Hulme (1983) carried out a proofreading task using very similar mutilations to the ones in the current study, and found that the shape of the words had an effect on readers. It would therefore be interesting to see if the same effect is found in a naming task, as it appears that the type of task does have an effect on results. Another argument mentioned in the limitations is that word shape is not maintained through mutilations involving substitutions. It would therefore be interesting to carry out a similar LDT, but with words mutilated in different ways, involving deletions, for example. Beech and Mayall (2005), for instance, deleted the exterior of a word, leaving only the interior parts of each letter. A study similar to the current one, but using the same process as Beech and Mayall could show that the WSH hypothesis is in fact correct, and that it was simply that the words' shapes were not maintained in the current experiment, or alternatively could further prove the findings of this study. Carrying out a study very similar to the current one, but mutilating the interiors of letters, rather than the envelope, would provide further insight into whether it is the envelope of words or interior features that is more important in word recognition. This was also quite a small-scale study and so carrying it out on a larger scale could produce different results. Most of the participants were students between the ages of 16 and 18, which is not representative of the population as a whole. In the studies already discussed that found evidence for the WSH, many more participants were tested: Beech and Mayall (2005) had 128 participants, Healy and Cunningham (1992) had 120, Lété and Pynte

(2003) had 50, and Monk and Hulme (1983) had 60. However, Perea and Rosa (2002) had only 24, which is fewer than the 31 in the current experiment, and still found evidence for the WSH. Whether or not it would make a difference to the findings cannot be concluded unless the study was carried out, but the more participants, from a wider range of backgrounds (e.g. different ages or education levels), the more representative of the population the results would be.

In conclusion, the results of the present experiment illustrate that the outline of a shape is not the most important aspect in visual word recognition. The results suggest that the definition of a word's outline consisting of only ascenders, descenders and neutrals is over-simplified, as no effect was found by substituting one letter for another, but does not completely discount the theory that a word's shape plays an important role in recognition. The study further proved that words are reacted to more quickly than non-words in LDTs and that neighbourhood size, word length and word frequency are important in word recognition. Written language is made up by individual letters and putting them together to make words, and so shape is of course vital to recognising a letter or word. The current study suggests that it is not simply the outline, or envelope, of a word or letters that is top of the hierarchy of features that readers use. This therefore suggests that other features of the shape of the words, such as their inner features or individual letters have more of an effect on recognition.

References

- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 802-814.
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 234-254.
- Beech, J. R. & Mayall, K. A. (2005). The word shape hypothesis re-examined: evidence for an external feature advantage in visual word recognition. *Journal of Research in Reading*, 3, 302-319.
- Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighbourhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 23 (4), 857-871.
- Cattell, J. M. (1886). The influence of the intensity of the stimulus on the length of the reaction time. *Brain*, 9, 512-514.

- Content, A., Mousty, P. & Radeau, M. (1990). Brulex: Une base de données lexicale informatisée pour le lexique français écrit et parlé [Brulex: A computerised lexical database for oral and written French lexicon]. *L'Année Psychologique*, 90, 551-566.
- Davis, C. J. (2005). N-Watch (Version 10.1) [Computer software]. Retrieved from <http://www.pc.rhul.ac.uk/staff/c.davis/Utilities/>
- Fisher, D. F. (1975). Reading and visual speech. *Memory & Cognition*, 3(2), 188-196.
- Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. C. T Walker (Eds), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27-85). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Grainger, J. (2008). Cracking the orthographic code: An introduction to the special issue on orthographic processes in reading. *Language and Cognitive Processes*, 23, 1-35.
- Haber, L. R., Haber, R. N., & Furlin, K. R. (1983). Word length and word shape as sources of information in reading. *Reading Research Quarterly*, 18 (4), 165-189.
- Healy, A. F. (1981). The effects of visual similarity on proofreading for misspellings. *Memory & Cognition*, 9, 453-460
- Healy, A. F. & Cunningham, T. F. (1992). A developmental evaluation of the role of word shape in word recognition. *Memory & Cognition*, 20, 141-150.
- Healy, A. F., Volbrecht, V. J., & Nye, T. R. (1983). The effects of perceptual condition on proofreading for misspellings. *Memory & Cognition*, 11, 528-538.
- Johnston, J. C. & McClelland, J. L. (1980). Experimental tests of a hierarchical model of word identification. *Journal of Verbal Learning and Verbal Behaviour*, 19, 503-524.
- Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J., & Wahlen, W. H. (2011). What lexical decision and naming tell us about reading. *Reading and Writing*, 25(6), 1259-1282.
- Lanthier, S. N., Risko, E. F., Stolz, J. A., & Besner, D. (2009). Not all visual features are created equal: Early processing in letter and word recognition. *Psychonomic Bulletin and Review*, 16, 67-73.
- Lété, B. & Pynte, J. (2003). Word-shape and word-lexical-frequency effects in lexical-decision and naming tasks. *Visual Cognition*, 10, 913-948.
- Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition: Research and theory. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 265-301). New York: Academic Press.
- Mayall, K. A. & Humphreys, G. W. (1996). Case mixing and the task sensitive disruption of lexical processing. *Journal of Experimental Psychology: Learning, memory and Cognition*, 22, 78-294.
- Mayall, K., Humphreys, G. W. & Olsen, A. (1997). Disruption to word or letter processing: The origins of case-mixing effects. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 22, 278-294.
- McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 483-524.
- Monk, A. F. & Hulme, C. (1983). Errors in proofreading: Evidence for the use of word shape in word recognition. *Memory & Cognition*, 11, 16-23.

- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Murray, W. S. & Forster, K. I. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, 111 (3), 721–756.
- New, B., Ferrand, L., Pallier, C. & Brysbaert, M. (2006). Re-examining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, 13(1), 45–52.
- Oden, G. C. (1984). Dependence, independence, and emergence of word features. *Journal of Experimental Psychology: Human Perception & Performance*, 10, 394–405.
- Oxford Dictionaries (n.d.). Retrieved from: <http://www.oxforddictionaries.com/>
- Paap, K. R., Newsome, S. L. & Noel, R. W. (1984). Word shape's in poor shape for the race to the lexicon. *Journal of Experimental Psychology: Human Perception and Performance*, 10 (3), 413–428.
- Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Front Neuroinformatics*, 2 (10). doi: 10.3389/neuro.11.010.2008
- Perea, M., Comesana, M., Soares, A. P., & Moret-Tatay, C. (2012). On the role of the upper part of words in lexical access: Evidence with masked priming. *The Quarterly Journal of Experimental Psychology*, 65 (5), 911–925.
- Perea, M. & Rosa, E. (2002). Does 'whole-word shape' play a role in visual word recognition? *Perception & psychophysics*, 64, 785–794.
- Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1142–1158.
- ReadingSoft. (2009). Speed reading test online. Retrieved from: <http://www.readingsoft.com/index.html#results>
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275–280.
- Rumelhart, D. E. & McClelland, J. L. (1982). An interactive activation model of context effects in letters perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89 (1), 60–94.
- Smith, F. (1969). Familiarity of configuration vs. discriminability of features in the visual identification of words. *Psychonomic Science*, 14, 261–262.
- Smith, F., Lott, D., & Cronnell, B. (1969). The effect of type size and case alternation on word identification. *American Journal of Psychology*, 82, 248–253.
- Stanley, P. (2005). Typophile: X-height. Retrieved from: <http://typophile.com/node/12028>
- Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67, 1176–1190.
- Warren, P. (2012). Experiments: Simple word-nonword. Retrieved from: http://www.intro2psycholing.net/experiments/visual/word_nonword/wnw_index.php

Jon CLENTON and Joanna MOSEDALE

Warren, P. (2013). *Introducing Psycholinguistics*. Cambridge: Cambridge University Press.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1, 59-85.

Whitney, P. (1998). *The psychology of language*. Boston: Houghton Mifflin Company.

Zipf, G. (1949), *Human Behaviour and the Principle of Least Effort..* Reading MA: Addison-Wesley.