Journal of Cognitive Psychology

Do serifs provide an advantage in the recognition of written words?

Carmen Moret-Tatay [a] [b] & Manuel Perea [a]

[a] Universitat de València, Valencia, Spain

[b] Universidad Católica de Valencia, Valencia, Spain

Available online: 09 Aug 2011

PLEASE SCROLL DOWN FOR ARTICLE

Do serifs provide an advantage in the recognition of written words?

Carmen Moret-Tatay[1,2] and Manuel Perea[1]

[1]Universitat de València, Valencia, Spain
[2]Universidad Católica de Valencia, Valencia, Spain

A neglected issue in the literature on visual-word recognition is the careful examination of parameters such as font, size, or interletter/interword spacing on reading times. Here we analysed whether serifs (i.e., the small features at the end of strokes) play a role in lexical access. Traditionally, serif fonts have been considered easier to read than sans serif fonts, but prior empirical evidence is scarce and inconclusive. Here we conducted a lexical decision experiment (i.e., a word/nonword discrimination task) in which we compared words from the same family (Lucida) either with a serif font or with a sans serif font—in both a block list and a mixed list. Results showed a small, but significant advantage in response times for words written in a sans serif font. Thus, sans serif fonts should be the preferred choice for text in computer screens—as already is the case for guide signs on roads, trains, etc.

Keywords: Lexical decision; Word recognition.

A wealth of research has shown that the process of reading a printed word is extremely efficient (Rayner & Pollatsek, 1989). Despite variations in font, size, or cAsE, the average reader is able to identify the appropriate lexical entry among thousands of other entries in around 150–250 ms. In the past decades, myriads of experiments in cognitive psychology have been conducted on the sublexical/lexical factors that influence the recognition of printed words (i.e., word frequency, familiarity, word length, age of acquisition, orthographic/phonological neighbourhood, regularity, etc.; see Andrews, 2006, for a review). However, there has been less research investigating the role of perceptual factors in the recognition of written words, such as the choice of typeface (e.g., see Slattery & Rayner, 2010), the size of the print (Chung, Mansfield, & Legge, 1998), or the spaces between letters/words (Tai, Sheedy, & Hayes, 2006).

Here we examine whether the speed of lexical access is affected by one critical aspect of every font: the presence or absence of "serifs" (i.e., the small appendages at the end of the strokes; e.g., compare the serif font Times New Roman with the sans serif font Arial). Serif fonts have their origins in the inscriptional capitals of the Roman script, and have traditionally been considered easier to read than sans serif fonts—as such they are the recommended fonts in the *Merriam-Webster's Manual for Writers and Editors* (2003). Indeed, the vast majority of books (including e-books) are printed in a serif font, and the *Publication Manual of the APA* (2009) also recommends serif fonts when submitting manuscripts for publication.

The question under scrutiny in the present study is whether serifs are just a historical artifact that does not help identifying words or whether serifs play a significant role in lexical access. The advocates of serif fonts claim that serifs may provide an extra cue to the location of letter strokes (Rubinstein, 1998) or that serifs make letters perceptually more unique and identifiable (e.g., see McLean, 1980). If that is the case, serifs would facilitate the process of letter identification, and thus identification times would be faster for the words written in a serif font rather than in a sans serif font. The story is more complicated, though. Leaving aside that none of the previous claims on the advantage of serifs has any firm empirical support (see Arditi & Cho, 2005, for a review), there are arguments that support a potentially deleterious role of serifs in visual-word recognition. First, serifs are not an inherent feature of letters. In terms of signal detection theory, serifs may merely act as visual noise (i.e., they reduce the clarity of the letters; see Woods, Davis, & Scharff, 2005). For instance, in the (biologically plausible) model of local combination detectors (Dehaene, Cohen, Sigman, & Vinckier, 2005), words are identified via their constituents (i.e., the letters) on the basis of neural levels of increased complexity along the ventral stream: from neurons responding to letter fragments (around the extrastriate visual cortical area V2), letter shapes (around V4), abstract letter detectors (V8), and substrings or small words (around the left occipitotemporal sulcus). In such a model, neurons at the lower level of vision may fire at some letter contours (i.e., serifs) that are not informative, thereby adding noise (rather than signal) to the recognition of the visually presented words. Second, because of the ornaments, serif fonts reduce slightly—all other thing being equal—the space between letters (e.g., compare computer vs. computer). This, in turn, may have two potentially harmful effects: (1) letter position coding may be hindered (Gomez, Ratcliff, & Perea, 2008), and (2) lateral masking across internal letters may increase (see Bouma, 1970). The net effect is that words written in a sans serif font may lead to faster identification times than words written in a serif font. It is important to note here that sans serif fonts are becoming increasingly popular (e.g., a sans serif font such as Helvetica has been used in many commercial wordmarks, and even films [Helvetica, 2007, by G. Hustwit] have been dedicated to this font). Furthermore, sans serif fonts are the preferred choice for guide signs on roads, trains, subways, etc. (In passing, note that the title page of the *Journal of Cognitive Psychology* employs a sans serif font.)

There is very little well-controlled research on whether the presence/absence of serifs has an impact on reading times. Given that a serif font and a sans serif font may differ in a number of relevant variables, such as the x-height, the thickness of stems, or character widths, we will restrict ourselves to those studies which have controlled for these factors. In an unpublished study, Morris, Aquilante, Bigelow, and Yager (2002) compared two fonts from the same family (Lucida and Lucida Sans). Importantly, as indicated by Morris and colleagues, these two fonts are identical in the critical factors (e.g., x-height, stem weights, character widths, character spacing, modulation of thick to thin, among others), so that the critical difference was the presence or absence of serifs. Using a rapid serial visual presentation, Morris et al. found a detrimental effect of serif in a small font size, whereas they failed to find a significant effect of serif in a 16-point size. More recently, Arditi and Cho (2005) created several fonts that had a differing degree of serif. In a continuous reading task, Arditi and Cho failed to find any systematic differences as a function of the presence/absence of serif. However, only four normal-reading individuals participated in the Arditi and Cho study, and one must be cautious about making strong conclusions from a null result with a very small sample size.

Clearly, whether or not a serif font produces faster recognition times than a sans serif font has important implications for readers in the digital era (i.e., particularly when reading on computer display screens). We believe that it is important to reexamine this issue by using the most popular task in visual word recognition: the lexical decision task (i.e., a word–nonword discrimination task). This task is highly sensitive to a broad variety of orthographic, phonological, lexical, and semantic effects (Balota et al., 2007). Thus, lexical decision may be more sensitive to subtle manipulations than the more global, continuous reading task employed by Morris et al. (2002) and Arditi and Cho (2005). As Rayner (1998) indicated, ''researchers can have some confidence that results obtained with standard naming and lexical decision tasks generalise to word recognition processes while reading'' (p. 392). Furthermore,

the lexical decision task provides an online measure unlike the rapid serial visual presentation. Because the magnitude of the effect of serif may be small, the number of words per condition in the present experiment was quite high (160 words in a serif font and 160 words in a sans serif font).

To avoid any potential confounds that may result when comparing two different fonts, we selected two fonts from the same family (i.e., the same fonts as in the Morris et al., 2002, study): Lucida Bright (i.e., a serif font) and Lucida Sans (i.e., a sans serif font). These two fonts occupy exactly the same amount of horizontal space (e.g., casino vs. casino), and there is no a priori reason to consider that the serif Lucida font is more/less familiar to the participants than the sans serif Lucida font (see Morris et al., 2002). Word-length (i.e., five-letter stimuli vs. eight-letter stimuli) was also manipulated to explore whether the presence/absence of serifs plays a differential role for shorter and longer words; note, however, that the number of letters does not influence the recognition of words with adult readers—only nonwords show a robust length effect (e.g., see Acha & Perea, 2008, for recent evidence).[1]

One final methodological note: The stimuli in the experiment were presented in serif versus sans serif font using a blocked design. Participants were presented with pure blocks of stimuli in serif (or sans serif) font, and with mixed blocks in which both fonts were presented in a random order (see Perea, Carreiras, & Grainger, 2004, for a similar procedure). This way, we controlled for the potentially deleterious effect of font alternation across trials in the same block.

METHOD

Participants

Twenty students from the University of Valencia participated voluntarily in the experiment. All of them either had normal or corrected-to-normal vision and were native speakers of Spanish. None of them reported having any reading disability.

Materials

We selected a set of 320 words from the B-Pal Spanish database (Davis & Perea, 2005). One hundred and sixty of these words were of five letters and the other 160 words were of eight letters. Factors like word frequency (mean word frequency per million: 37.6 and 36.3 for the five-letter and eight-letter words, respectively) and number of orthographic neighbours (Coltheart's N) (mean 0.55 and 0.41 for the five-letter and eight-letter words, respectively) were controlled. For the purposes of the lexical decision task, 320 nonword targets were created (160 of five letters and 160 of eight letters; e.g., *gulca*, *vildo*, *esnicial*, *cilanera*). The nonwords had been created by changing two letters from Spanish words that did not form part of the experimental set. The mean number of neighbours (Coltheart's $N = 0.34$) was similar as that for words. The complete list of stimuli is available at http://www.uv.es/mperea/stimuli-serif.pdf. The procedure employed in the experiments to create the pure/mixed lists mimicked the procedure used in the blocking experiments of Perea et al. (2004). For each participant, there was a pure list with 160 stimuli (80 words and 80 nonwords) written in Lucida Bright (a serif font), and a pure list with 160 stimuli (80 words and 80 nonwords) written in Lucida Sans (i.e., a sans serif font). In addition to these two pure lists, there were two mixed lists composed of an equal number of words/nonwords in serif/sans serif font. Assignment of words to conditions was arranged so that each word occurred in a pure list (both serif and sans serif) and in a mixed list (both serif and sans serif), but not for the same participant. For instance, the word *oasis* was written in Lucida Bright in a pure list for Group 1, it was written in Lucida Sans in a pure list for Group 2, it was written in Lucida Bright in a mixed list for Group 3, and it was written in Lucida Sans for Group 4. The four experimental lists (two pure and two mixed) were written in a random order to all participants in one group, but within each list, a different random ordering of the stimuli occurred for each participant. There were five participants in each of the four lists. For each participant, the ordering of the four blocks was randomly chosen by the computer program.

---

[1]It has been shown that ClearType improves the legibility and the reading times of words when presented on a computer LCD screen (Slattery & Rayner, 2010). For that reason, we chose ClearType in the present experiment. (ClearType was designed by Microsoft to improve the appearance of text on computer display screens.)

Procedure

Participants were tested individually in a quiet room. Presentation of the stimuli and recording of response times were controlled by DMDX (Forster & Forster, 2003) on a Windows-based computer. The stimuli were presented on an LCD screen. On each trial, a fixation point (" + ") was presented at the centre of the screen for 500 ms. Then, the target stimulus, always in lowercase, remained on the screen until the participant's response. The intertrial interval was 1.5 s. The target stimulus was written in 14-point Lucida font (either Lucida Bright or Lucida Sans). Participants were instructed to press one of two buttons on the keyboard to indicate whether the letter string was a Spanish word or not. This decision was to be made as rapidly and as accurately as possible. Reaction times were measured from the onset of the letter string until the participant's response. There was a small break after each experiment block (160 trials). Each participant received a total of 20 practice trials prior to the experimental phase. The session lasted approximately 25 min.

RESULTS

Incorrect responses (3.4% of the data for words and 4.8% for nonwords) and lexical decision times less than 250 ms or greater than 1500 ms (0.77% of trials) were excluded from the latency analysis. The mean response times and error percentages from the subject analysis are presented in Table 1. For word and nonword stimuli, ANOVAs based on the participant and item response latencies and error percentage were

conducted based on a 2 (font: serif, sans serif) $\times 2$ (length: 5 letters, 8 letters) $\times 2$ (block: pure, mixed) design.

Word stimuli

The ANOVA on the latency data showed that, on average, words written in a sans serif font were responded to 19 ms faster than the words with the words written in a serif font, $F1(1, 19) = 5.35$, $MSE = 2945$, $p < .05$; $F2(1, 318) = 20.22$, $MSE = 5083$, $p < .01$. None of the other effects/interactions approached significance (all $ps > .15$).

The ANOVA on the error data showed that participants committed more errors to five-letter words than to eight-letter words, $F1(1, 16) = 13.70$, $MSE = 17.3$, $p < .005$, $F2(1, 318) = 17.19$, $MSE = 110.6$, $p < .001$. In addition, participants committed more errors on words written in a serif font than on words written in a sans serif font, although the effect was not significant, $F1(1, 19) = 1.82$, $MSE = 12.4$, $p = .19$, $F2(1, 318) = 2.61$, $MSE = 68.9$, $p = .10$. None of the other effects/interactions approached significance (all $ps > .16$).

Nonword stimuli

The ANOVA on the latency showed that responses to eight-letter nonwords took longer than the responses to five-letter nonwords, $F1(1, 19) = 15.38$, $MSE = 3113$, $p < .005$, $F2(1, 318) = 37.88$, $MSE = 11404$, $p < .001$. In addition, participants responded slightly faster to nonwords written in a sans serif font than to nonwords written in a sans serif font, although the effect

TABLE 1.

Mean lexical decision times (RTs, in ms; with standard deviations in parentheses) and percentages of errors (ERs) for word and nonword targets in the experiment

	Type of list							
	Pure				*Mixed*			
	Serif		*Sans serif*		*Serif*		*Sans serif*	
	RT	ER	RT	ER	RT	ER	RT	ER
Words								
5-letter	535 (92)	5.1 (5.0)	513 (104)	4.0 (4.4)	534 (88)	5.5 (5.4)	516 (93)	4.0 (3.1)
8-letter	549 (115)	2.1 (2.3)	524 (120)	2.5 (3.3)	540 (108)	2.5 (3.8)	525 (118)	1.8 (1.8)
Nonwords								
5-letter	595 (111)	4.1 (4.2)	583 (128)	6.4 (6.6)	585 (114)	5.8 (3.7)	582 (117)	5.5 (5.1)
8-letter	627 (125)	4.4 (6.7)	620 (161)	4.6 (5.7)	622 (148)	3.3 (2.9)	619 (153)	4.1 (3.7)

was not significant, $F1 < 1$; $F2(1, 318) = 2.75$, $MSE = 6176$, $p = .09$. None of the other effects/interactions approached significance (all $ps > .15$).

The ANOVA on the error data showed that participants committed more errors to five-letter nonwords than to eight-letter nonwords, $F1(1, 19) = 8.50$, $MSE = 8.49$, $p < .01$, $F2(1, 318) = 2.82$, $MSE = 205.6$ $p = .09$. None of the other effects/interactions approached significance (all $ps > .15$).

DISCUSSION

The results of the present experiment are straightforward: There was no reading benefit for words written in a serif font. On the contrary, there was an advantage of around 19 ms for words written in a sans serif font. This effect was independent of stimulus length (short vs. long) and block (pure vs. mixed). Furthermore, there was no reading cost of font alternation. It may be important to note that, for word stimuli, the advantage of sans serif over serif fonts occurred for a large majority of readers: 16 out of 20 participants (80%). Nonetheless, the median size of the effect on a subject-by-subject basis was rather small (around 8 ms). We ran a replication of the experiment reported in this paper using a completely randomised list of stimuli. This replication (that will not be reported in full) also revealed a significant advantage of around 8–11 ms for the stimuli written in a sans serif font. Taken together, the present findings reveal that there is a small advantage of processing for words written in a sans serif font—using a standard size font (14 point).

What is the reason why the words written in a sans serif font produced faster identification times than the words written in a serif font? First, as indicated in the introduction, serifs may just act as visual noise on the letters (i.e., the serif contour does not constrain any alternatives among the potential base letter; e.g., it could be an *m*, an *r*, an *f*, etc.). As a result, the process of word recognition may be smoother for items written in a sans serif font (Woods et al., 2005). Bear in mind that the neurons in the visual system responsible for abstract letter representation are font-invariant and absent from any gratuitous ornament. This implies that serifs (if anything) may reduce the signal/noise ratio (see Dehaene et al., 2005, for a hierarchical model of visual-word recognition along the ventral stream). Second, because of the little ornaments, there is a slight increase in the separation between letters in a sans serif font than in a serif font (e.g., compare hand vs. hand). Prior research has shown that a small increase in interletter spacing (e.g., casino vs. casino) produces faster reading times than the default interletter spacing (see Arditi & Cho, 2005; Tai et al., 2006). This may be so because of a reduction in lateral inhibition and/or as a result of a more precise stage of letter position coding (i.e., less perceptual uncertainty; see Gomez et al., 2008). But whatever the specific cause is, the point here is that serifs do not seem to play a beneficial role in visual-word recognition—beyond being a decorative burden. This finding has practical implications: Sans serif fonts should be the preferred choice for text in computer screens—as is already the case for most guide signs on roads, trains, or museums.

Clearly, more research is required to examine whether the observed pattern of data goes beyond the individual words and is actually influencing normal reading. This can be achieved by monitoring the participants' eye movements during normal silent reading (see Rayner & Pollatsek, 1989, for a review of the literature on eye movements and reading). At a foveal level, the pattern of word data should presumably mimic that observed here (see Davis, Perea, & Acha, 2009, for a recent comparison of lexical decision times and eye movement data). Nonetheless, one remaining question is whether at a parafoveal level the linking strokes of serif fonts could be less affected by lateral masking than the small gaps created by sans serif fonts. One second area of interest for future research is children's reading materials. Children books are usually printed in a serif font (see Woods et al., 2005); it may be important to examine whether sans serif fonts also facilitate visual-word recognition for beginning/intermediate readers—as occurs with adult readers. Finally, additional experimentation on the presence/absence of serifs should also be performed with readers with low vision. In a review on typographical factors with low vision readers, Russell-Minda et al. (2007) indicated that "there appears to be a subjective preference among readers with low vision for sans serif fonts" (p. 411). Carefully controlled studies are necessary to determine whether this preference goes accompanied by faster reading times.

In sum, the choice of critical parameters in a text (e.g., individual text characteristics, print size, line/character/word spacing) in publishing

companies up to the present has been set without much empirical/theoretical basis (see Woods et al., 2005). This study represents a preliminary, modest effort to set some standards in this field. The conventional view that supports the use of serif in written texts is mostly based on historical grounds (i.e., the inscriptional lettering of Latin words during the Roman Empire) and anecdotal evidence (e.g., "serifs help the eyes to stick to the horizontal line") rather than on well-controlled research (see Arditi & Cho, 2005, for review). The present data demonstrate that serifs do *not* facilitate the process of visual-word identification; instead, the presence of serifs may (if anything) hinder lexical access. This, together with the growing popularity of sans serif fonts, may be taken as a signal that, as happened with Gothic fonts in the twentieth century, serif fonts may eventually fall into disuse in the twenty-first century—note that the default font in Microsoft Word is no longer a serif font (Times New Roman) but a sans serif font (Calibri).

REFERENCES

Acha, J., & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: Evidence with beginning, intermediate, and adult readers. *British Journal of Psychology*, 99, 245–264.

American Psychological Association. (2009). *The Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Andrews, S. (Ed.). (2006). *From inkmarks to ideas: Current issues in lexical processing*. Hove, UK: Psychology Press.

Arditi, A., & Cho, J. (2005). Serifs and font legibility. *Vision Research*, 45, 2926–2933.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226, 177–178.

Chung, S. T. L., Mansfield, J. S., & Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Research*, 38, 2949–2962.

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37, 665–671.

Davis, C. J., Perea, M., & Acha, J. (2009). Re(de)fining the orthographic neighbourhood: The role of addition and deletion neighbours in lexical decision and reading. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1550–1570.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences*, 9, 335–341.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers*, 35, 16–124.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review*, 115, 577–601.

McLean, R. (1980). *The Thames and Hudson manual of typography* (Vol. 1). London, UK: Thames & Hudson Ltd.

Merriam-Webster's manual for writers and editors. (2003). Springfield, MA: Merriam-Webster.

Morris, R. A., Aquilante, K., Bigelow, C., & Yager, D. (2002, May). *Serifs slow RSVP reading at very small sizes, but don't matter at larger sizes.* Paper presented at the SID Digest of Technical Papers symposium, Boston, MA.

Perea, M., Carreiras, M., & Grainger, J. (2004). Blocking by word frequency and neighborhood density in visual word recognition: A task-specific response criteria account. *Memory and Cognition*, 32, 1090–1102.

Rayner, K. (1998). Eye movements in reading and information processing. *Psychological Bulletin*, 124, 372–422.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rubinstein, R. (1988). *Digital typography: An introduction to type and composition for computer system design*. Boston, MA: Addison-Wesley.

Russell-Minda, E., Jutai, J. W., Strong, J. G., Campbell, K. A., Gold, D., Pretty, L., & Wilmot, L. (2007). Typeface legibility for readers with low vision: A synthesis of the research evidence. *Journal of Visual Impairment and Blindness*, 101, 402–415.

Slattery, T. J., & Rayner, K. (2010). The influence of text legibility on eye movements during reading. *Applied Cognitive Psychology, 24*, 1129–1148.

Tai, Y.-C., Sheedy, J., & Hayes, J. (2006). Effect of letter spacing on legibility, eye movements, and reading speed. *Journal of Vision*, 6, 994.

Woods, R. J., Davis, K., & Scharff, L. F. V. (2005). Effects of typeface and font size on legibility for children. *American Journal of Psychological Research*, 1, 86–102.