

MONITOR

THE SCIENCE OF WORD RECOGNITION

By Kevin Larson

This presentation, made at the 2003 ATyPl conference in Vancouver, provoked much interest and debate.

Evidence from the last twenty years of work in cognitive psychology indicates that we use the letters within a word to recognise the word. Many typographers and text enthusiasts insist that words are recognised by the outline made around the word shape. Some have used the term 'bouma' as a synonym for word shape. (The term bouma, which comes from papers written in the 1970s by H. Bouma, appears in Paul Saenger's 1997 book *Space Between Words: The Origins of Silent Reading*.)

My paper is written from perspective of a reading psychologist. The data from dozens of experiments all come from peer-reviewed journals where the experiments are well specified, so that anyone can reproduce the experiment and expect to achieve the same result. My goal is to review the history of why psychologists moved from a word shape model of word recognition to a letter recognition model, and to help others to come to the same conclusion.

I will start by describing three major categories of word recognition models: the word shape model, the serial model, and the parallel model of letter recognition. I will present data that was used as evidence to support each model, and will evaluate the models in terms of their ability to support the data. Finally I will describe some recent developments in word recognition and a more detailed model that is currently popular among psychologists.

Model no. 1. Word shape

The word recognition model, which says that words are recognised as complete units, is the oldest model in the psychological literature. The idea is that

we see words as a complete patterns, rather than the sum of letter parts. (See Figures 1 and 2.)

In 1886, James Cattell was the first psychologist to propose the word shape model of word recognition. He presented letter and word stimuli to study participants for a very brief period of time (five to ten milliseconds), and found that people were more accurate at recognising the words than the letters. This finding is now called the Word Superiority Effect. Cattell concluded that we use whole words for word recognition because of their advantage over individual letters.

The second piece of experimental data that supports the word shape model is that lowercase text is read faster than uppercase text. On average people read lowercase text five to ten per cent faster than uppercase text. This supports the word shape model because lowercase text has unique patterns of ascending, descending and neutral characters, while uppercase text has less variance in text size and shape.

The patterns of errors that are missed while proof-reading text provides the third key piece of experimental evidence to support the word shape model. Study participants were asked to read a passage of text for comprehension and to mark any misspelling they found. The passage had some misspellings that were consistent with word shape, and some that were inconsistent with the intended word's shape. Haber and Schindler (1981) found that misspellings consistent with word shape were twice as likely to be missed as misspellings inconsistent with word shape. The word shape model predicted that consistent word shape misspellings would be caught less often, because of their similarity with the intended word. (See Table 3.)

TABLE 3. MISSPELLINGS THAT ARE CONSISTENT WITH WORD SHAPE ARE MISSED MORE OFTEN

Intended word: test	Error rate
Consistent word shape (tesf)	13%
Inconsistent word shape (tesc)	7%

Model no. 2. Serial letter recognition

The serial letter recognition model says that recognising a word is analogous to looking up a word in a dictionary. You start off by finding the first letter, then the second, and so on until you recognise the word. This model is appealing because it is easier to test than the word shape model. However it was quickly discarded because it couldn't explain the Word Superiority Effect.

Evidence for the serial letter recognition model comes from letter recognition speed and the word length effect. Letters can be recognised one at a time at rate of ten to twenty milliseconds per letter, which is consistent with a typical reading rate of 300 words per minute (wpm). The serial letter recognition model also correctly predicts that words with fewer letters are recognised more quickly than words with many letters, while the word shape model expects longer words with more unique patterns to be easier to recognise.

Model no. 3. Parallel letter recognition

The model that most psychologists currently accept as the most accurate is the parallel letter recognition model. This model says that the letters within a word are recognised simultaneously, and the letter information is used to recognise the word. This is an active area of research and there are many

specific models that fit into this general category of model.

Much of the evidence for the parallel letter recognition model comes from the eye movement literature. With the advent of fast eye trackers and computers, a great deal has now been learned about how we read. We now have the ability to make changes to text in real time while people read, which has provided useful insights into reading process that weren't previously possible.

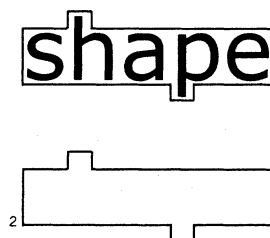
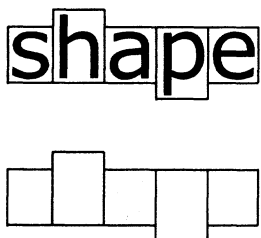
It has been known for over 100 years that, when we read, our eyes don't move smoothly across the page. Rather they make discrete jumps from word to word. We fixate on a word for a period of time, roughly 200-250 milliseconds, then make a ballistic movement to another word. These movements are called saccades and usually take 20-35 milliseconds. Most saccades are forward movements from seven to nine letters, but ten to fifteen per cent of all saccades are regressive or backwards movements. Most readers are completely unaware of the frequency of regressive saccades while reading.

The location of the fixation is not random. Fixations never occur between words, and usually occur just to the left of the middle of a word. Not all words are fixated; short words, and particularly function words, are frequently skipped. (Figure 4, opposite, shows a diagram of the fixation points of a typical reader.)

During a single fixation there is a limit to the amount of information that can be recognised. The fovea, the clear centre point of our vision, can only see three to four letters to the left and right of fixation at normal reading distances. Visual acuity decreases quickly in the parafovea, which extends out as far as fifteen to twenty letters to the left and right of the fixation point.

Eye movement studies indicate that there are three zones of visual identification. Readers collect information from all three zones during the span of a fixation. Closest to the fixation point is where word recognition takes place. This zone is usually large enough to capture the word being fixated and often includes smaller function words directly to the right of the fixated word. The next zone extends a few letters past the word recognition zone, and readers gather preliminary information about the next letters in this zone. The final zone extends out to fifteen letters past the fixation point. Information gathered out this far is

FIGURE 1. WORD SHAPE RECOGNITION USING THE PATTERN OF ASCENDING, DESCENDING AND NEUTRAL CHARACTERS.
FIGURE 2. WORD SHAPE RECOGNITION USING THE ENVELOPE AROUND THE ENTIRE WORD.



WITH THE ADVENT OF FAST EYE TRACKERS AND COMPUTERS, A GREAT DEAL HAS NOW BEEN LEARNED ABOUT HOW WE READ

TABLE 5. RELATIONSHIP BETWEEN LETTERS AVAILABLE IN WINDOW AND READING RATE.

used to identify the length of upcoming words and to identify the best location for the next fixation point.

There are two experimental methodologies that have been critical for understanding the fixation span: the moving window paradigm and the boundary study paradigm. These methodologies make it possible to study readers while they're engaged in ordinary reading. Both rely on fast eye trackers and computers to perform text manipulations while a reader is making a saccade. While making a saccade, the reader is functionally blind. He or she will not perceive that text has changed, if the change is completed before the saccade has finished.

wpm. If the reader is given nine letters past the fixation point, they will see the entire word 'experiment', and part of the word 'was'. With nine letters, the reading rate is moderately slowed. However if the reader is given fifteen letters past the fixation point, reading speed is just as fast as if there was no moving window present. Up to fifteen letters, there is a linear relation between the number of letters that are available to the reader and the speed of reading. (See Table 5.)

From this study we learned that our perceptual span is roughly fifteen letters. This is interesting, as the average saccade length is seven to nine letters, or roughly half our

Window Size	Sentence	Reading Rate
3 letters	An experimxxx xxx xxxxxxxxxxx xx	207 wpm
9 letters	An experiment wax xxxxxxxxxxx xx	308 wpm
15 letters	An experiment was condxxxxx xx	340 wpm

MOVING WINDOW STUDY

In the moving window technique we restrict the amount of text that is visible to a certain number of letters around the fixation point, and replace all of the other letters on a page with the letter 'x'. The reader's task is simply to read the page of text.

McConkie and Rayner (1975) examined how many letters around the fixation point are needed to provide a normal reading experience. Table 5 shows a snapshot of what a reader would see if they were reading a passage and fixated on the second 'e' in 'experiment'. If the reader is provided three letters past the fixation point, then they won't see the entire word for 'experiment', and their average reading rate will be a slow 207

perceptual span. This indicates that, while readers are recognising words closer to the fovea, we're using additional information further out to guide our reading.

The moving window study demonstrates the importance of letters in reading, but it is not airtight. The word shape model of reading would also expect that reading speed would decrease as word shape information disappears. The word shape model would make the additional prediction that reading would be significantly improved if information on the whole word shape were always retained. This turns out to be false.

Table 6 shows the reading rate when three letters are available. It is roughly

TABLE 6. FULL WORD INFORMATION DOES NOT IMPROVE READING RATE.

Window Size	Sentence	Reading Rate
3 letters	An experimxxx xxx xxxxxxxxxxx xx	207 wpm
1 word (3.7 letters)	An experiment xxx xxxxxxxxxxx xx	212 wpm
9 letters	An experiment wax xxxxxxxxxxx xx	308 wpm
2 words (9.6 letters)	An experiment was xxxxxxxxxxx xx	309 wpm
15 letters	An experiment was condxxxxx xx	340 wpm
3 words (15.0 letters)	An experiment was conducted xx	339 wpm

equivalent to the reading rate when the fixated word is entirely there. That is true even though the entire word has 0.7 more letters available on average. When the fixated word and the following word are entirely available, reading rate is equivalent to when nine letters are available. Reading rate is also equivalent when three words or fifteen letters are available.

This means that reading is not necessarily faster when entire subsequent words are available; similar reading speeds can be found when only a few letters are available.

BOUNDARY STUDY

The boundary study (Rayner, McConkie and Zola, 1980) is another innovative paradigm that eye trackers and computers made possible. With this we can examine what information the reader is using inside the perceptual span (fifteen letters), but outside of the word that is being fixated.

Figure 7 illustrates what the reader sees in this kind of study. While reading the words 'The old captain' the reader

will be performing ordinary reading. When the reader reaches the word 'put', the key word of interest becomes available within the reader's fixation span. In this example the key word is 'ebovf'. When the reader saccades from 'put' to 'ebovf', the saccade will cross an invisible boundary which triggers a change in the text. Before the saccade finishes, the text will change to the correct text for the sentence, in this case 'chart'. The reader will always fixate on the correct word for the sentence. (See Figure 7.)

The critical word in this study is presented in different conditions, including an identical control condition ('chart'), a dissimilar word shape with some letters in common ('chyt'), and similar word shape with no letters in common ('ebovf'). Readers were faster when some letters were in common than when the word shape was similar. This demonstrates that letter information is being collected within the fixation span, even when the entire word is not being recognised. (See Table 8.)

Rayner, McConkie and Zola (1980)

FIGURE 4. SACCADIC EYE MOVEMENTS.

Roadside joggers endure sweat, pain and angry drivers in

the name of fitness. A healthy body may seem reward ...

FIGURE 7. DURING THE SACCADIC, 'EBOVF' CHANGES TO 'CHART'.

The old captain put the ebovf on the ...

The old captain put the chart on the ...

WORD SHAPE IS NO LONGER A VIABLE MODEL OF WORD RECOGNITION

TABLE 8. RELATIVE SPEED (RECOGNITION TIME) OF BOUNDARY STUDY CONDITIONS.

further investigated what happens with a capitalised form of the critical word CHART. This eliminates the role of word shape, but retains perfect letter information. They found that the fixation times were the same as the control condition. This demonstrates that it isn't visual information about either word shape or letter shape that is being retained from saccade to saccade, but rather abstracted information about which letters are coming up.

Evidence for word shape revisited
What about the evidence that supported the word shape model? Does any of it disprove the parallel letter recognition model?

The strongest evidence for the word shape model was the Word Superiority Effect, which showed that letters can be more accurately recognised in the context of a word than in isolation. This was a logical claim until McClelland and Johnson (1977) demonstrated that

rather the existence of regular letter combinations.

The weakest evidence in support of word shape is that lowercase text is read faster than uppercase text. This is entirely a practise effect. Most readers spend the bulk of their time reading lowercase text, and are therefore more proficient at it. When readers are forced to read large quantities of uppercase text, their reading speed will increase to the rate of lowercase text.

Earlier I reported that readers were twice as likely to fail to notice a misspelling that is consistent with word shape. Unfortunately this study confounded word and letter shape, comparing misspellings with similar letter and word shape to misspellings with different letter and word shape. Paap, Newsome and Noel (1984) determined the relative contribution of word shape and letter shape, and found that the entire effect is driven by letter shape. There are many more errors

critical word	condition	recognition time
chart	Identical word (control)	210 ms
chyft	Dissimilar word shape Some letters in common	280 ms
ebovf	Similar word shape No letters in common	300 ms

pseudowords also show a Word Superiority Effect. For example, pseudowords such as 'mave' and 'rint' are not words in the English language and should not have a familiar word shape, but they do have the phonetic regularity that makes them easily pronounceable. Therefore, the reason for the Word Superiority Effect isn't the recognition of word shapes, but

when the replaced letter has the same basic shape (replacing 'n' for 'h') than when the replaced letter retains the same pattern of ascenders and descenders (replacing 'd' for 'h'). (See Table 9.)

Further examination of the evidence used to support the word shape model has demonstrated that the case for the word shape model was not as strong

TABLE 9. WORD SHAPE AND LETTER SHAPE CONTRIBUTIONS TO PROOF-READING ERRORS.

	Same word shape	Different word shape
than		
Same letter shape	t h an – 15% missed	t n an – 19% missed
Different letter shape	t d an – 8% missed	t m an – 10% missed

as it seem7ed. In the next section I will describe an active area of research: using computers to model parallel letter recognition.

Computer models of parallel letter recognition

The human brain is made up of billions of neurons, each with 4000 synapses on average. Each of these neurons is a summation machine that adds (and subtracts) input from other neurons. We can use computer simulations to create models that have biological plausibility which can explain complex behaviours with simple mechanisms. (See Figure 10.)

The first well known computer simulation of reading was McClelland and Rumelhart's Interactive Activation model (1981).

Figure 11 shows how this model works. The reader here is processing the letter 'T' in the first position in a word. The flow of information here starts at the bottom where there are visual feature detectors. The 'T' is the most activated letter node because it has the most incoming excitatory activation coming from the feature level. The connections between each node is either excitatory (represented with an arrow at the end of the connection) or inhibitory (represented with a circle at the end of the connection). The letter node for 'T' will send excitatory activation to all the words that start with 'T' and inhibitory activation to all the other words. As word nodes gain in activation, they send inhibitory activation to all other words and excitatory activation back to letter nodes from letters in the word. Letters in positions other than the first are needed in order to figure out which of the words that start with 'T' is being read.

The most important benefit of neural network modelling is that it's specific enough to be programmed into a computer and tested. The Interactive Activation model is able to explain many human behaviours that it was not specifically designed for. For example when a human is shown the degraded stimulus in figure 12, it is very easy to figure out that 'WORK' is the degraded word, but the computer simulation of

this model can also solve this problem.

There has been great progress in developing neural network models of reading that can account for more human reading behaviours (see Plaut, McClelland, Seidenberg and Patterson, 1996). Neural network models are now able to account for how reading develops and generate correct word pronunciations without the use of specific word nodes.

Conclusions

Given that all the reading research psychologists I know support a version of the parallel letter recognition model of reading, how is it that all the typographers I know say we read by matching whole word shapes?

It appears to be a grand misunderstanding. The paper by Bouma that is most frequently cited does not support a word shape model of reading. Bouma (1973) presented words and unpronounceable letter strings to subjects away from the fixation point and measured their ability to name the first and last letters. He found that:
A) Subjects are more successful at naming letters to the right of fixation than to the left of fixation.
B) When distance to the right of the fixation point is controlled, subjects are better able to recognise the last letter of a word than the first letter of word. This is why we tend to fixate just to the left of the middle of a word. Bouwhuis and Bouma (1979) extended the Bouma (1973) paper by not only

OUR PERCEPTUAL SPAN IS ROUGHLY FIFTEEN LETTERS

finding the probability of recognising the first and last letters of a word, but also the middle letters. They used this data to develop a model of word recognition based on the probability of recognising each of the letters within a word. They conclude that 'word shape ... might be satisfactorily described in terms of the letters in their positions.' This model of word recognition clearly influenced the McClelland and Rumelhart neural network model discussed earlier, which also used letters in their positions to probabilistically recognise words.

Word shape is no longer a viable model of word recognition. The bulk of scientific evidence says that we recognise a word's component letters, then use that visual information to recognise a word. In addition to perceptual information, we also use contextual information to help recognise words during ordinary reading, but that has no bearing on the word shape versus parallel letter recognition debate.

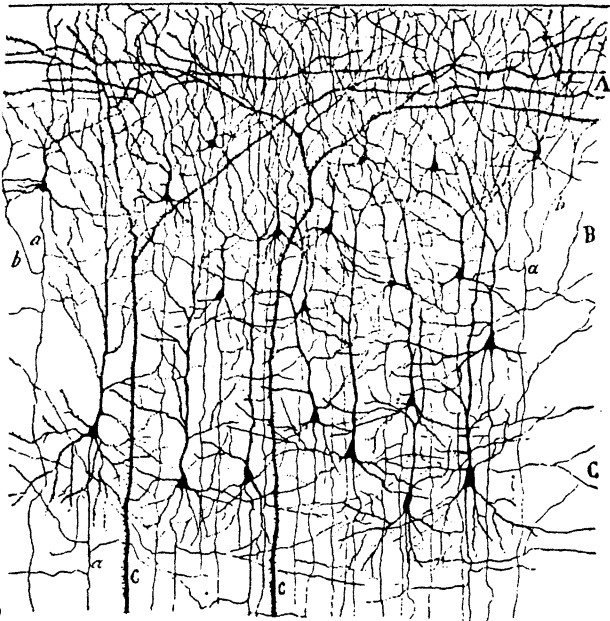
I hope that it is clear that the readability and legibility of a typeface should not be evaluated on its ability to generate a good word shape.

Kevin Larson is a psychologist who has been working for Microsoft in different capacities since 1996. In 2000 he completed his PhD in cognitive psychology from the University of Texas at Austin, studying word recognition and reading acquisition. He joined Microsoft's ClearType team in 2002 to help get a better scientific understanding of the benefits of ClearType and other reading technologies.

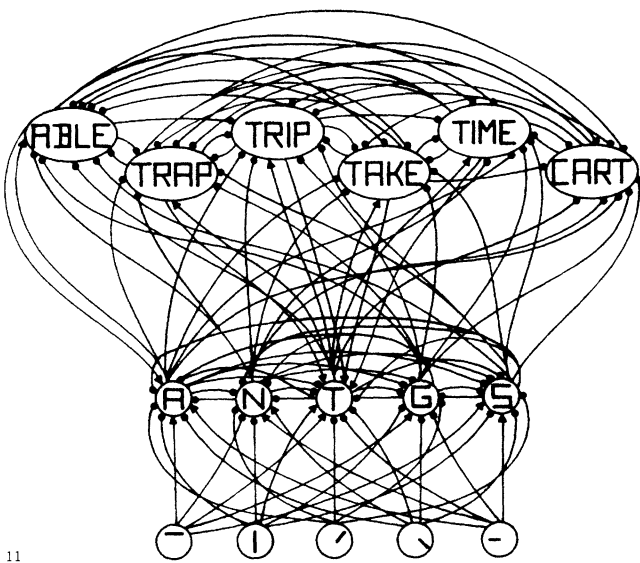
REFERENCES

- Bouma, H. (1973). Visual Interference in the Parafoveal Recognition of Initial and Final Letters of Words, *Vision Research* no. 13, pp.762-782.
- Bouwhuis, D. & Bouma, H. (1979). Visual word recognition of three letter words as derived from the recognition of the constituent letters, *Perception and Psychophysics* no. 25, pp.12-22.
- Cattell, J. (1886). The time taken up by cerebral operations. *Mind* no. 11, pp.277-282, 524-538.
- Haber, R. N. & Schindler, R. M. (1981). Errors in proof-reading: Evidence of syntactic control of letter processing? *Journal of Experimental Psychology: Human Perception and Performance* no. 7, pp.573-579.
- McClelland, J. L. & Johnson, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception and Psychophysics* no. 22, pp.249-261.
- 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* no. 88, 375-407.
- 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* no. 10, pp.413-428.
- Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* no. 103, pp.56-115.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Review* no. 124 (3), pp.372-422.

This is an abridged version of Kevin Larson's paper.

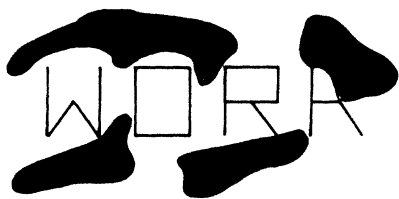


10



11

FIGURE 10. A FIELD OF NEURONS AND SYNAPSES IN THE CEREBRAL CORTEX. FIGURE 11. MCCLELLAND AND RUMELHART'S INTERACTIVE ACTIVATION MODEL. A FEW OF THE NEIGHBOURS OF THE NODE FOR THE LETTER 'T' IN THE FIRST POSITION IN A WORD, AND THEIR INTERCONNECTIONS. FIGURE 12. THIS DEGRADED STIMULUS IS EASILY READ AS 'WORK' BY HUMAN READERS.



12