The Shape of a Word Can Specify Its Meaning
Author(s): Ralph Norman Haber and Lyn R. Haber
Source: *Reading Research Quarterly*, Vol. 16, No. 3 (1981), pp. 334-345
Published by: Wiley on behalf of the International Literacy Association
Stable URL: http://www.jstor.org/stable/747406
Accessed: 22-05-2018 14:39 UTC

The shape of a word can specify its meaning

RALPH NORMAN HABER
LYN R. HABER
University of Illinois at Chicago Circle

GROFF (1975) HAS CLAIMED that the shape of a high frequency word, defined by its pattern of ascending, descending, and small letters, would specify so few words uniquely that shape information would be useless to readers. The present paper counters Groff's argument by showing that word shape information, when combined with knowledge of the syntactic and semantic structure of the passage being read, more often than not specifies unique words among all of the alternative high frequency words of English. Groff's error was his failure to consider word shape in the context of semantic and syntactic information, which are also available to the reader. The effects of reading all upper case print are also reviewed to support the role of word shape information in normal reading.

La forme d'un mot peut spécifier son sens

GROF (1975) A DÉCLARÉ que la forme d'un mot à haute-fréquence, définie par son modèle de lettres ascendantes, descendantes et minuscules, spécifierait si peu de mots seulement que l'information donnée par la forme du mot serait inutile aux lecteurs. Ce rapport va à l'encontre de l'argument de Groff en montrant que l'information donnée par la forme du mot, en combinaison avec la connaissance de structure syntactique et sémantique du passage lu, spécifie la plupart du temps des mots uniques parmi tous les mots anglais alternatifs de haute-fréquence. L'erreur de Groff consistait dans le fait de ne pas avoir considéré la forme du mot dans le contexte de l'information sémantique et syntactique, qui est aussi offerte au lecteur. Les effets de lecture de toute impression de haut de casse sont aussi revus pour soutenir le rôle de l'information donnée par la forme du mot en lecture normale.

La forma física de una palabra puede precisar su significado

GROFF (1975) MANTUVO que la forma física de una palabra de alta frecuencia, definida por su diseño ascendiente, descendiente, y de pequeñas letras de imprenta, concretaría de una manera única tan pocas palabras, que la información resultante de la forma física no sería de utilidad a los lectores. El presente ensayo contesta el argumento de Groff mostrando que información sobre la forma física de palabras, ligada a conocimiento de estructura sintáctica y semántica del paseje leído, precisa—en más que menos casos—

palabras sin igual entre todas las palabras alternativas de alta frecuencia en inglés. El error de Groff consistió en la falta de considerar la forma física de la palabra en el contexto de información semántica y sintáctica que también está a disposición del lector. Los efectos de lectura de letras de imprenta mayúsculas también se discuten, para corroborar la función de información sobre la forma física de palabras durante la lectura normal.

The case for word shape as a source of visual information in reading has never made much headway against critics who argue that in order to read we ultimately need to process the individual letters. Most of these criticisms are based on logical grounds, or on indirect demonstrations that show that reading easily occurs in the absence of normal word shape information. In the face of such arguments, few proponents have advocated the use of word shape as one of the sources of information in reading, or as something to be taught to beginning readers. The present paper is designed to counter the logical arguments, and suggest reasons why word shape should not be ignored by theorists, reading teachers, or readers.

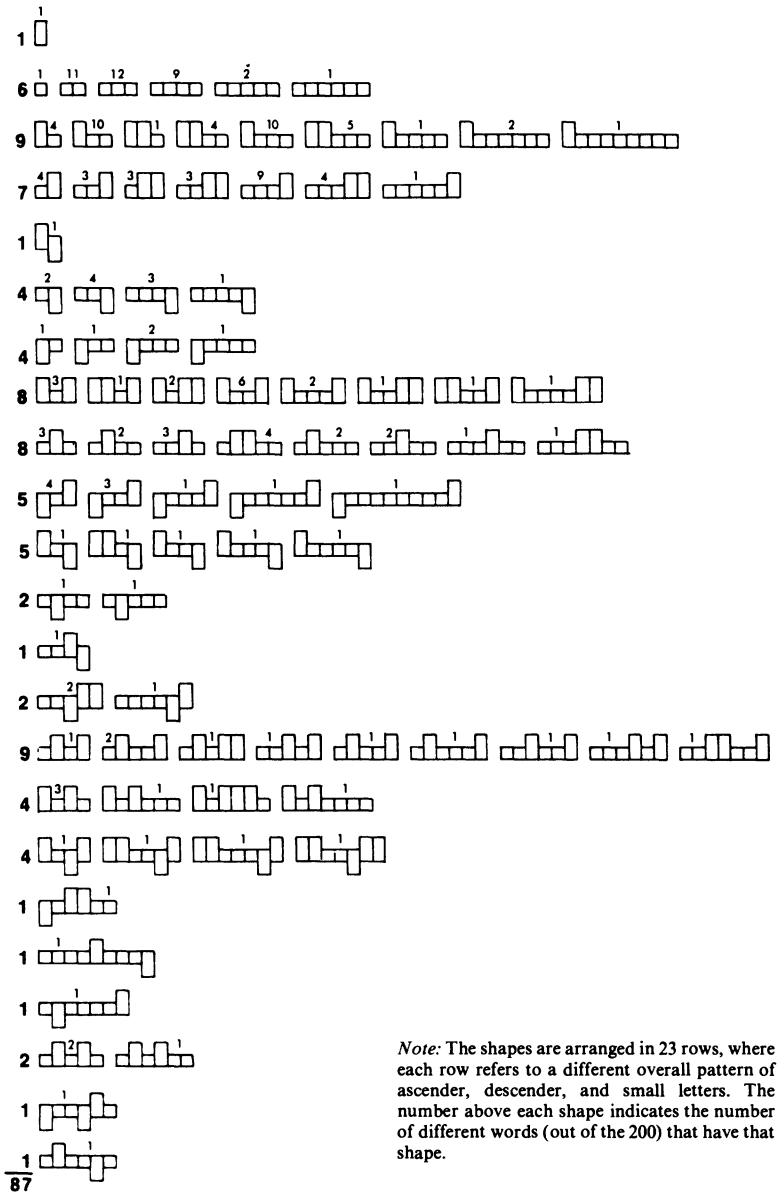
The most explicit case against the use of word shape as a cue to word recogniton has been made on logical grounds by Groff (1974, 1975). As an example, he took two word lists (Hillerich, 1974; Kucera & Francis, 1967) and examined the shapes of the 238 and the 200 most frequent words on each list respectively. Using a definition of word shape based upon the length and pattern of lower case ascending, descending, and small letters, he found that no more than 20% of the words on either list had unique shapes and that as many as 15 words shared some shapes. How then, Groff asks, can a reader use word shape information to distinguish among words, when it is such an unreliable predictor.

Groff's argument would be reasonable if word shape information had to be used in isolation from other sources of information. This is never the case for word shape, nor for any other single source of information. Readers have access to many sources of information at many levels of processing, word shape being only one source. The following analysis shows that word shape is capable of distinguishing among the individual high frequency words that Groff studied with the addition of syntactic and semantic information from the passage.

Table 1 lists, in order, the 200 most frequent words from the Kucera and Francis (1967) count. Figure 1 displays the 87 different word shapes (as defined by the heights of their lower case letters) that fit these

Table 1 The 200 most frequent words in printed English as tabulated by Kucera and Francis (1967)

the	this	we	only	me	because	here	since	home	war
of	had	him	other	even	each	between	against	small	until
and	not	been	new	most	just	both	go	found	always
to	are	has	some	made	those	life	came	thought	away
a	but	when	could	after	people	being	right	went	something
in	from	who	time	also	how	under	use	say	fact
that	or	will	these	did	too	never	take	part	though
is	have	more	two	many	little	day	three	once	water
was	an	no	may	before	state	same	states	general	less
he	they	if	then	must	good	another	himself	high	public
for	which	out	do	through	very	know	few	upon	put
it	one	so	first	back	make	while	house	school	think
with	you	said	any	years	would	last	use	every	almost
as	were	what	my	where	still	might	during	don't	hand
his	her	up	now	much	own	us	without	does	enough
on	all	its	such	your	see	great	against	got	far
be	she	about	like	way	men	old	place	united	took
at	there	into	our	well	work	year	american	left	head
by	would	than	over	down	long	off	around	number	yet
I	their	then	man	should	get	come	however	course	government



Note: The shapes are arranged in 23 rows, where each row refers to a different overall pattern of ascender, descender, and small letters. The number above each shape indicates the number of different words (out of the 200) that have that shape.

Figure 1
The 87 different outline shapes that describe all of the 200 most frequent English words.

200 words. The 87 shapes are arranged in 23 groups, each group sharing the same overall pattern. Thus there are six different word shapes made up of all small letters (line 2 of Figure 1), differing only in the number of letters. There are nine shapes of the pattern of one or more tall letters followed by one or more small letters. The number above each shape stands for the number of different words among the top 200 that have that particular shape. The largest number of words sharing the same shape is 12 for three small letters. Of the 87 different shapes, only 8 of them are shared by five or more words, while 51, or 25%, have a unique shape. Up to this point, these are the same data that Groff used.

In normal reading, whether beginning or mature, the reader can use information about the syntactic structure of the sentence and the semantic content of the passage to help make choices among words. Without arguing whether this actually represents how reading occurs, we can easily show in theory how much more information is provided by word shape when shape information is combined with syntactic and semantic knowledge.

As a demonstration of this combination of word shape with syntax, Figure 2 presents a brief passage taken from a second-grade Basal reader (Harris, Creekmore, & Greenman, 1972) in which each of its words from the most frequent 200 has been replaced by its outline shape. The first word of each sentence is not included, because the initial capitalization alters the familiar shape. It is certainly possible to read and understand the story when presented as in Figure 2. Table 2 presents an analysis confirming the impression that the story is quite easy to understand when presented in this version, and shows why. Each of the 39 words portrayed by their shape in the story is listed.

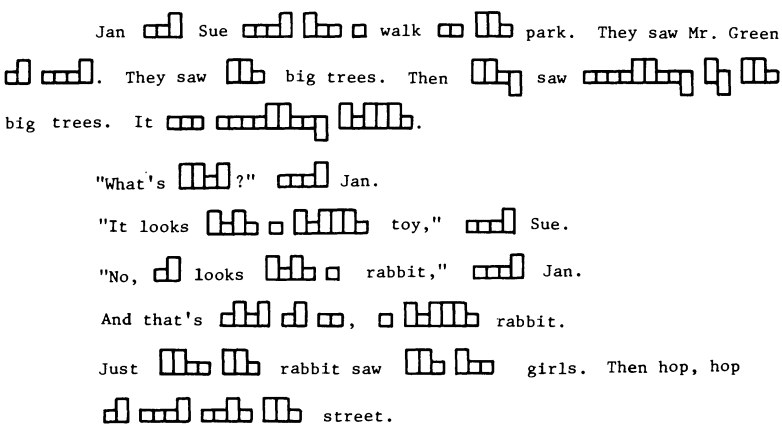


Figure 2

A brief Basal reader story in which each high frequency word is replaced by its outline shape.

Table 2 Number of words with the same shape as each high frequency word that appeared in the story in Figure 2

High Frequency Words	Number of Words Sharing Envelope	With Context Restriction
and	3	1
went	9	6
for	10	3
a	1	1
in	11	5
the	1	1
at	4	2
work	9	3
the	1	1
they	1	1
something	1	1
by	1	1
the	1	1
was	12	1
something	1	1
little	1	1
that	1	1
said	9	2
like	3	1
a	1	1
little	1	1
said	9	1
it	4	2
like	3	1
a	1	1
said	9	1
what	1	1
it	4	1
is	11	1
a	1	1
little	1	1
then	4	2
the	1	1
the	1	1
two	10	1
it	4	1
went	9	1
into	3	1
the	1	1
Means	4.1	1.4

Note. The first column lists the number of different words shared by that shape (taken from Table 1). The last column lists the number of words from column 1 that are permissible alternatives given the contextual constraint of the story up to that point.

The first data column in Table 2 tabulates the number of words that share each shape, taken from Figure 1. The mean is 4.1 words per shape. A one-in-four chance of predicting a word correctly is not very high, as Groff pointed out. The second column in Table 2 indicates the number of words from the preceding column that are permitted to

occur in that word position in a sentence, given whatever contextual restrictions have been created by all the preceding words. For example, ten words in the most frequent 200 share the shape of "for," the third word in Table 2. However, six of these cannot reasonably occur as the next word in a story which begins "Jan and Sue went _____ . . . ." Thus, reference to the contextual restrictions obtaining in a real context rule out more than half the permissible words of this one word shape.

To provide a measure of contextual restriction, 18 adult native English speakers were asked to rate the acceptability of each of the alternative words of each shape in each position for the sentences in this story. Each rater was handed 39 cards. The first card contained the word "Jan _____" followed by the three words from the 200 most frequent whose shape corresponds to the word actually used in the Basal reader: "not, and, out." Raters were told to circle all of the alternatives which they could use to construct a reasonable English sentence at the level that might appear in an early grade Basal reader. Then they went to the next card, which had: "Jan and Sue _____; said, such, most, must, much, each, work, used, went." They continued through the deck of 39 cards. Two criteria were used to tabulate the ratings. The first counted only alternatives circled by at least 80% of the judges as contextually acceptable, while the second counted all alternatives circled by 50% of the judges. The data from the second, lax criterion are entered in the last column of Table 2. The mean number of contextually acceptable words that fit the word shape for each position is 1.43, nearly a third of that for word shape alone shown in Column 2 ($t=8.3$, $p<.001$). For 31 of the 39 words, or 80%, there is only one permissible alternative—more than three times the 25% found in word shape alone. If the more conservative criterion is used, the mean number of acceptable alternatives is 1.15, with 36 of the 39 blanks having only one alternative.

As a final comparison, we asked the judges to indicate which alternative they thought had actually appeared in the story, even if there were several they thought were acceptable. The judges were unanimous for 37 of the 39 words, picking the correct alternative in each case. One they failed to agree on was the second word to be guessed: "Jan and Sue_____." Tense had not yet been defined, much less topic or tone of the story, so that "said, went, and work" were equally reasonable choices.

Imagine the analysis turned the other way. For each of the high frequency words in the story, how many alternative words could be substituted without regard to shape and length, but preserving context? Even restricted to the 200 high frequency words as choices, the numbers

are substantially larger. For example, the last word in the first paragraph of the passage, "little," could be replaced with any other adjective, and, of course, with any singular animate noun—about 25 words out of the 200. Knowing the shape of the word, however, restricts the choice to a single word. This supports the great power of word shape for restricting alternative possibilities among high frequency words in any well-formed sentence.

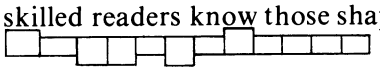
The reader is invited to try this analysis on any set of materials desired—other Basal reader passages, the daily newspaper, scientific writing, or even the present article. The numbers themselves may change slightly, but the substantial reduction in alternatives for any one word is always demonstrable when syntactic and semantic information are used in addition to shape and length.

The present analysis is a theoretical one. It is designed to counter the theoretical one presented by Groff. Groff claimed that shape did not sufficiently differentiate among words to be useful. We have shown that when context is added to shape, only a very, very few of the frequent words in a passage could be misguessed.

What evidence do we have that readers use shape information, singly or in combination? Rayner (1978), in a recent review of research of the eye movements made by mature readers, reports a number of studies demonstrating that readers do attend to word shape information in normal reading and that removing such information reduces reading performance. Wildman and Kling (1978-1979) and Haber and Haber (1980) review the use of visual information in reading more generally and also describe a number of studies supporting the utilization of word shape in the reading process. As one example from our own laboratory, Haber, Haber, and Furlin (Note 1) asked college students to read passages in which, after several full sentences were presented, a sentence ended in the middle. the reader had to guess the next word. Then some more of the text was presented, again ending in the middle of a sentence, and the next word had to be guessed. This procedure continued to the end of each of a large number of passages. In one condition, no information about the word to be guessed was given; in another, its length was indicated; and in a third, its length and shape. The accuracy of guessing increased substantially across these three conditions, reaching 50% for some passages.

There is no claim made here that word shape, used in conjunction with syntactic and semantic knowledge, is always sufficient and that no processing of individual letters ever occurs. Intuitively, one

important use of word shape information is for identification of the highly familiar and frequent words in print, ones with which readers have had extensive encounters. In support of this intuition, Haber and Schindler (1981) showed that misprints that preserved word shape were less detectable than those that altered shape only for the very high frequency function words. There were no differences in detection of misprints as a function of shape preservation for lower frequency content words. Therefore, word shape information is utilized, at least in a proofreading task, for only the high frequency words; for the remainder, processing requires other sources of information, presumably letter features.

All words printed in lower case type have a shape, so that in theory, readers have that shape information available. However, for long words with presumably unique shapes, it seems unlikely that even skilled readers know those shapes—know, for instance, that the shape is  hippopotamus. But that does not mean that shape information about low frequency words is useless. For these words, given their length and unusual shape, readers can readily reject nearly all words in English as possibilities from their shape alone. There fore, one use of word shape information for low frequency words might be to eliminate possible incorrect alternatives, rather than only providing information for the identification of the single correct alternative. Further, if the story being read is about animals, especially jungle animals, the shape may be sufficient for a unique identification even if a hippopotamus has not appeared in the story before. More generally, we expect that the semantic content of a passage primes the reader to expect many words with related semantic fields. Thus, a story about cats leads the reader to expect particular words such as fur, whiskers, tails, purr, mischief, and the like. These expectations should prime not only specific spellings, but also specific shapes of words; while whisker is a low frequency word, once expected, its shape is known, and can be used to help identify the word.

The above comments suggest that word shape information may be useful for words of all types, but that it may function in different ways for different classes or frequencies of words, as well as for different reading tasks. A proofreading task, such as Haber and Schindler (1981) used, may not be a sufficiently sensitive indicator of those differences.

It is, of course, possible to read print in which familiar word shape information is deleted, as with text printed in all capital type. Since all the letters are scaled to the same size and location on the line,

the shape patterns, except for length, are removed. Fisher (1975) and Smith (1969) both report a small reduction in reading speed when normal lower case is switched to all upper case—between a five and ten percent loss. When both shape and length are removed, either by removing the spaces (Fisher, 1975) or by filling each with an X (Spragins, Lefton, & Fisher, 1976; Hochberg, Levin, & Frail, Note 2), reading speed drops much more. Hence, shape and length information both appear to be important in normal reading, even though reading can still occur without them. And just because people can read upper case print easily does not imply that word shape is not useful when available.

Conversely, we expect that reading time is increased when trying to read passages printed as is the one in Figure 2, at least for readers unpracticed on such passages. We are collecting data on this question currently. If true, then it shows that even for the high frequency words, the individual letters do make a contribution to their processing. One factor that has yet to be controlled is possible dimensions of letter shapes as they affect word shapes. Consider the differences in shape between *did* and *but*, in which the center of gravity is to the right for the former and to the left for the latter, even though both have the same overall shape, as defined in this paper. We do not know of evidence that readers do attend to such configurations, but it is certainly a potential source of information, again at least for highly familiar words.

There is much less literature on the use of word shape by beginning readers. Part of its absence may be explained by the belief, fostered by Groff's work, that shape could not be a predictive source of information so why bother to study it. It is also true that some theorists of the reading acquisition process (e.g., Calfee, 1977) believe that beginning readers so concentrate on various decoding skills that they cannot also attend to larger sources of information such as whole word shape or syntactic rules.

In early studies with beginning or immature readers, Marchbanks and Levin (1956) and Williams, Blumberg, and Williams (1970) report data to show that shape cues are not important. More recently, Rayner (1976; Rayner & Hagelberg, 1975) has criticized the earlier studies for failing to properly control the features of the individual letters, concluding that word shape is a useful cue in first grade and beyond. Hochberg, Levin, and Frail (Note 2) and Spragins, Lefton, and Fisher (1976) examined the differences between adults, fifth-, and third-grade readers in their use of length and shape information. Both found that adults suffered more disruption when word length or word

shape information was removed than did children, suggesting that adults used that information more in their normal reading. Even so, subjects of all ages were affected by removal of shape and length, so even the third-grade readers were using that information.

We badly need more data about the different visual sources of information during the acquisition process. Such data presumably will show that beginning readers do rapidly become sensitive to information about words conveyed by their shapes. But such results do not tell us how to take advantage of this sensitivity. We do not suggest that children, or adults, should be taught the shapes of words. Such learning should probably occur fairly automatically, especially for the high frequency words of the language. Rather, given the potential usefulness of shape information, children should be encouraged to use the shape information. Thus, when a child misreads a word in which the erroneous response does not match the stimulus word in shape, the teacher can point out that mismatch. Such instruction can increase attention to initial letters or initial sounds as attack skills and might teach that whole word shape also conveys information about a word.

However, our expectations are that shape information is used primarily in skilled reading, and it does not play an important role in the early stages of acquisition. Haber and Haber (1981) argue that shape information about a word is picked up in peripheral vision and is used to narrow hypotheses about what the word might be—these hypotheses being specifically confirmed by more detailed analyses as the eye subsequently fixates on the word. We do not expect to find the integration of peripheral and central processing in reading to be very advanced until reading skill is substantial, so the advantage given by the peripheral pickup of shape information will not be evident in young readers.

In conclusion, we have argued that the case for word shape as a source of information in reading can be made and defended. The logical arguments against this theory fail because such arguments consider each source of information in isolation rather than analyze the joint constraints among general sources. We have considered the empirical effects on reading of removing word shape, or of having only word shape, and have shown that the results of such experiments are quite consistent with the usefulness of word shape information. We badly need data on the potential role of word shape in the reading acquisition process. Hopefully, the case made for it in this paper will help promote new work in this promising field.

REFERENCES

CALFEE, R.C. Assessment of independent reading skills: Basic research and practical applications. In A.R. Reber & D.L. Scraborough (Eds.), *Toward a psychology of reading*. Hinsdale, N.J.: Lawrence Erlbaum Associates, 1977.

FISHER, D.F. Reading and visual search. *Memory & Cognition*, 1975, *3*, 188-196.

GROFF, P. The topsy turvy world of sight words. *Reading Teacher*, 1974 *27*, 572-578.

GROFF, P. Research in brief: Shapes as cues to word recognition. *Visible Language*, 1975, *9*, 67-71.

HABER, L.R., & HABER, R.N. Visual processes in reading. In F.J. Pirozzolo & M.C. Wittrock (Eds.), *Neuropsychological and cognitive processes in reading*. New York: Academic Press, 1981.

HABER, R.N., & SCHINDLER, R. Errors in proofreading: Evidence of syntactic control of letter processing? *Journal of Experimental Psychology: Human Perception and Performance*, 1981, *7*.

HARRIS, T.L., CREEKMORE, M., & GREENMAN, M.H. *Green feet*. Oklahoma City: The Economy Co., 1972.

HILLERICH, R.L. Word lists—Getting it all together. *Reading Teacher*, 1974, *27*, 353-360.

KUCERA, H., & FRANCIS, W.N. *Computational analysis of present day American English*. Providence, RI: Brown University, 1967.

MARCHBANKS, G., & LEVIN, H. Cues by which children recognize words. *Journal of Educational Psychology*. 1956, *56*, 57-61.

RAYNER, K. Developmental changes in word recognition strategies. *Journal of Educational Psychology*, 1976, *68*, 323-329.

RAYNER, K. Eyemovements in reading and information processing. *Psychological Bulletin*, 1978, *85*, 618-660.

RAYNER, K., & HAGELBERG, E.M. Word recognition cues for beginning and skilled readers. *Journal of Experimental Child Psychology*, 1975. *20*, 444-455.

SMITH, F. Familiarity of configuration vs. discriminibility of features in the visual identification of words. *Psychonomic Science* 1969, *14*, 261-262.

SPRAGINS, A.B., LEFTON, L.A., & FISHER, D.F. Eyemovements while reading and searching spatially transformed text: A developmental examination. *Memory & Cognition*, 1976, *4*, 36-42.

WILDMAN, D., & KLING, M. Semantic, syntactic and spatial anticipation in reading. *Reading Research Quarterly*, 1978-1979, *14*, 128-164.

WILLIAMS, J.P., BLUMBERG, E.L., & WILLIAM, D.V. Clues used in visual word recognition. *Journal of Educational Psychology*, 1970, *61*, 310-315.

Reference Notes

1. Haber, L.R., Haber, R.N., & Furlin, K.R. *Word length and word shape as sources of information in reading*. Under editorial review, 1981.
2. Hochberg, J., Levin, H., & Frail, C. *Studies in oral reading. VII: How interword spaces affect reading*. Mimeographed, Cornell University, 1966.