# 3

# *The Magic of Reading: Too Many Influences for Quick and Easy Explanations*

Dominic W. Massaro
Alexandra Jesse
*University of California, Santa Cruz*

A skilled reader cannot help but read even the blandest banners on the information highway and real highways. Like listening, contact with the linguistic signal is all that seems to be necessary. This behavior is easily exposed by the Stroop color word test. You are asked to name the color of the print of each of the words in a list. When the words are the names of other colors (e.g., the word *blue* printed in red), however, you either switch gears into slow motion or name the written words rather than the colors (i.e., in our example, you incorrectly answer "blue" rather than "red"). The written word overrides your intention to name the color, contributing to the impression that reading is clearly magical.

The goal of this chapter is to show that reading of words, though indeed magical, is a magic that has been well examined and basically involves the ability of the reader to exploit multiple sources of information in a (overlapping) series of information-processing stages. Many of these sources and stages were studied by Dick Venezky, which makes this chapter a tribute to his insights into the magic of reading. Our proposal is grounded in the assumption that reading words is fundamentally a pattern recognition process, which involves imputing meaning to an input pattern. As our guide to the understanding of visual word recognition, we use a pattern-recognition model, the fuzzy logical model of perception (FLMP), that has achieved scientific success in reading as well as in several other domains of information processing.

The general assumption of the FLMP is that well-learned patterns, such as written words, are recognized by applying a general algorithm, regardless of the modality or the nature of the pattern (see, e.g., Massaro, 1998). The FLMP assumes three operations: feature evaluation, feature integration, and decision. All three processes are successive but overlapping. Feature evaluation provides the degree to which each feature of the stimulus matches the corresponding feature in each prototype in memory. Prototypes are summary descriptions and contain a conjunction of various ideal properties (features) that a member of this prototype category should have. Fuzzy truth values (Zadeh, 1965) reflect the degree to which a given stimulus matches to the features of a prototype. The fuzzy truth values lie between *completely false* (0) and *completely true* (1). In addition to the multiple bottom-up sources of information, various top-down sources are assumed. These sources in reading are the orthographic, phonological, syntactic, semantic, and pragmatic structure, as well as the sublexical mappings from print to sound. Continuous information is available from each source, and the output of the evaluation of each source is independent of the output of another source (see Fig. 3.1).

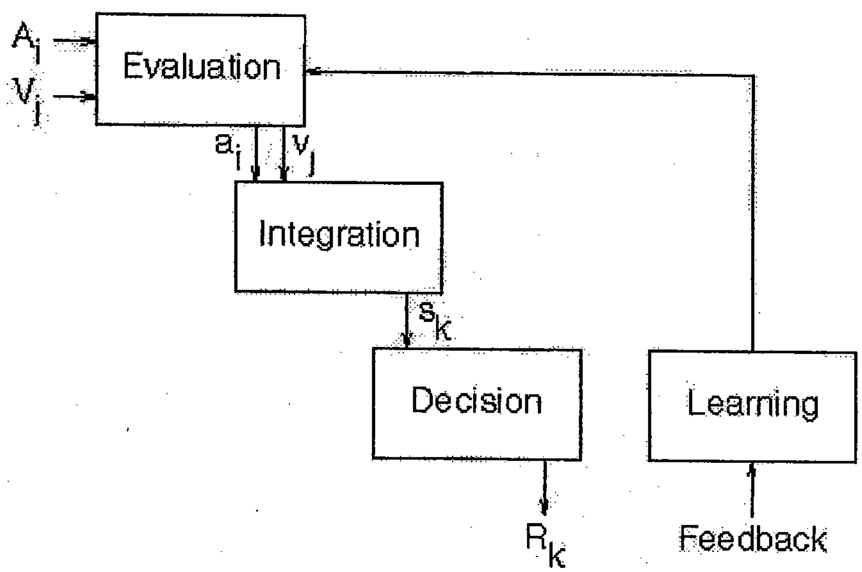


FIG. 3.1. Schematic representation of the FLMP to include learning with feedback. The three recognition processes are shown to proceed left to right in time to illustrate their necessarily successive but overlapping processing. These processes make use of prototypes stored in long-term memory. The sources of information are represented by uppercase letters. Auditory information is represented by $A_i$ and visual information by $V_j$. The evaluation process transforms these sources of information into psychological values (indicated by lowercase letters $a_i$ and $v_j$). These sources are then integrated to give an overall degree of support, $s_k$, for each alternative k. The decision operation maps the outputs of integration into some response alternative, $R_k$. The response can take the form of a discrete decision or a rating of the degree to which the alternative is likely. The feedback is assumed to tune the prototypical values of the features used by the evaluation process.

Feature integration combines all degrees of matches from each source of information for each prototype. The outcome of this process is the total degree to which each prototype matches the stimulus. The third process in the model makes a decision based on a relative goodness rule (Massaro & Friedman, 1990), the relative support of one alternative compared to the support for all other alternatives. The model predicts that one feature has its greatest effect when a second feature is the most ambiguous. Through this assumption, the model predicts that the time for decision increases with the ambiguity of the information available to the decision stage (Massaro, 1987).

Consider the elaboration of the FLMP, depicted in Fig. 3.2, as a description of how the many different sources of information can influence letter and word processing in reading. The presentation of a letter pattern initiates a sequence of processing stages. Visual features are evaluated, and this information has several consequences. First, complete or even partial information from the features can activate letter patterns in long-term memory. Needless to say, the more visual information available, the more easily letter and word recognition can take place. Second, recognition of letters can be supplemented by the reader's knowledge of how letter patterns occur in the language. We call the form of this knowledge ortho-

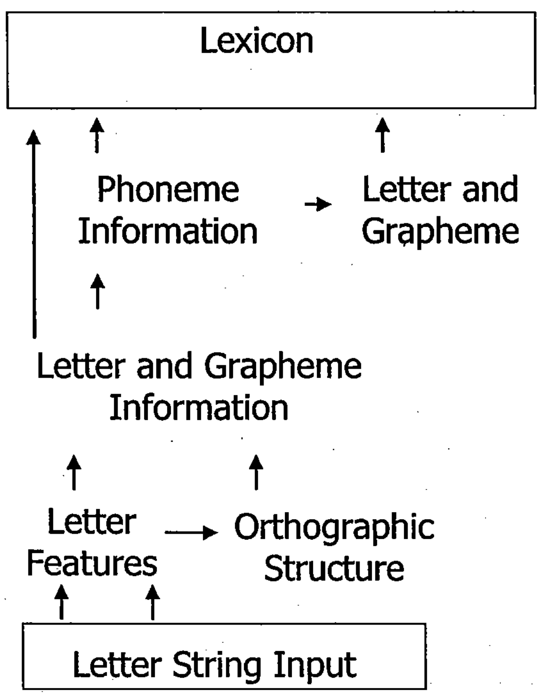


FIG. 3.2. The different processes between presentation of a letter string and access to the lexicon as described by an elaboration of the fuzzy logical model of perception, which shows the processing streams of the many different sources of information that can influence letter and word processing in reading.

graphic structure. Letters that occur together more often should be easier to recognize than those in an infrequent or unlawful arrangement because of the contribution orthographic structure.

Letter information activates words and spoken language representations, which we call phoneme information. Because readers also know the relationship between sounds and spellings, the activation of phonemes in turn activates a set of spelling patterns. Like the information about the association of letters to phonemes, the activated spelling patterns associated with phoneme information also feed forward to the lexical level and can aid or hinder word activation. A phoneme pattern limited in the number of ways it can be spelled would facilitate lexical access because only these spellings would activate the lexicon. When a phoneme pattern can be spelled in many different ways, it would hinder lexical access because a larger set of different possible spellings would be activating the lexicon. The information passed from this sound-to-spelling source (sound-to-spelling fluency) does not affect evaluation or integration but can influence the time needed for decision making (Massaro, 1987).

Using this model as a framework, we discuss three of the sources potentially involved in the word recognition process in detail. The first source is visual influences, such as the features of letters and the overall shape of a word. Second, we describe research indicating that knowledge about the orthographic structure of a word might help its recognition. Finally, we discuss evidence that the two-way association between orthographic and phonological information influences the word recognition process.

## INFLUENCES IN WRITTEN WORD RECOGNITION

In our model, letters and words are recognized via the visual features that make them up. Features can be elemental or relatively global depending on how much of a letter they describe. Elemental features of uppercase *E* include three horizontal lines and one vertical. A global feature of lowercase *c, e,* and *o* is a circular envelope that distinguishes them from other letters, such as *f, h,* or *j*. Discovering the functional features in reading is a challenging empirical endeavor (for reviews, see also Massaro & Sanocki, 1993). Our goal here is simply to provide the reader with the flavor of what is already known and recent studies addressing this problem.

Reading research began as an active area of psychological inquiry at the turn of the century (see Huey, 1968; Woodworth, 1938). For the last three decades, after a period of relative inactivity during the heyday of behaviorism, the process of reading written words has been intensely studied. One finding that led to this renewed interest was the demonstration that a letter could be better recognized when presented in the context of a word than when presented in a random letter string or even when presented alone. This advantage, called the word advantage or word superiority effect, was shown to exist even if the possibilities of postperceptual guessing and memory loss were eliminated (Reicher, 1969).

What was it about words that contributed to this word advantage? A natural interpretation of the word superiority effect is that words are recognized as wholes without intermediate processing of the features of letters that make them up. This little paragraph has circulated cyberspace in the last quarter of 2003, with the implication that words are read as wholes:

> Aoccdrnig to a rscheeahcr at an Elingsh uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer are in the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae we do not raed ervey lteter by itslef but the wrod as a wlohe.

Are you impressed that you were able to read this passage? Maybe you shouldn't be because you read much more slowly and laboriously than normal. Reading aloud would have also revealed the added difficulty created by scrambling the internal letters. Holistic word recognition is an old idea in reading research. Like John Updike, we are not fans of holism: "Next to the indeterminacy principle, I have learned in recent years to loathe most the term 'holistic,' a meaningless signifier empowering the muddle of all the useful distinctions human thought has labored at for two thousand years" (Roger Lambert, in John Updike's *Roger's Version*, p. 171).

Some researchers and educators (Haber, Haber, & Furlin, 1983; Johnson, 1975) proposed that words are recognized as patterns of unique shapes rather than as unique sequences of letters. We call these properties global supraletter features because they supposedly are composed of multiletter patterns and even whole word patterns. The earlier paragraph shows convincingly that we can read scrambled words, even if they are misspelled or incomplete (like *rscheeahcr* or *iprmoetnt*). But are we actually reading words as a whole? And do we need the first and last letter to stay in their original position?

A little thought reveals that global features cannot be sufficient for even the expert reader. One of the strongest arguments against the idea of supraletter features is the small potential contribution of supraletter features to reading. Overall word shape, for example, does not sufficiently differentiate among the words of a language. In a classic study, Groff (1975) examined the shapes of high-frequency words taken from schoolbooks. The shape was defined by drawing a contour around the letters.

Only 20% of the 283 words was represented by a unique shape. Groff rightly concludes that the small number of words that can be represented by a unique shape precludes the use of this cue for accurate word recognition. Using a much larger sample of words, Paap, Newsome, and Noel (1984) also showed that there is not sufficient uniqueness of word shapes that could be used to mediate word recognition.

There is also experimental evidence against the idea of word recognition based on supraletter features. Adams (1979) asked whether disrupting word shape (mixing upper- and lowercase and type fonts of letters) eliminates the identification advantage of words over nonword letter strings. If the word shape is contributing to the word advantage, because it is used to access the lexicon, then the advantage should diminish when the shape of words is altered and can therefore no longer be used to access the mental lexicon. The word advantage did not change when the global word shape was eliminated (see also Thompson & Massaro, 1973).

One would think that the word shape idea was sufficiently demolished but Paap and his colleagues (1984) tested whether the number of words that share a certain word shape could still influence word recognition. When a shape matches a small set of words (e.g., *cellar*), then the shape feature restricts the lexical search to this small set of candidates, and therefore all words of this small set should be processed faster or more accurately than words in a larger set (e.g., *recall*). When the shape is shared by a large set of words, a response cannot be given until letter identification is almost completed. Contrary to this expectation, Paap et al. (1984) actually found that words with rare shapes are not accessed faster than words with common shapes, falsifying the word shape hypothesis.

Although three decades of empirical evidence indicate that words are not read as a whole, the first and last letters may be more important than the medial ones. The paragraph of scrambled words that was sent so actively over the Internet could have been inspired by the research of Jordan and colleagues (Jordan, Thomas, Patching, & Scott-Brown, 2003). Jordan et al.'s study goal was to show that exterior letters (i.e., the first and the last letter of a word) are special in reading. Indeed, there is some truth to the hypothesis that first and last letters have an advantage over their embedded letter cohort. This advantage occurs because neighboring letters are not always kind to one another. Lateral masking refers to the interference that a letter has on its neighbor(s). An embedded letter in a word has two interfering neighbors, whereas the first and last letters have only one. Accordingly, a letter will necessarily be (ceteris paribus) more visible at the first and last position than in the middle of a word. Jordan et al.'s results could be simply evidence of this lateral masking rather than implication of a special functional unit of exterior letters used to access the mental lexicon.

If the first and last letters were responsible for word recognition, then we would also expect that words would be uniquely defined by their first and last letters in analogous fashion to what we expected from word shape. A quick look at the 1,000 most frequent words in English reveals that there are many words that share their first and last letters, even when word length is controlled:

| | | |
|---|---|---|
| *wish wash* | *while whole* | *that test* |
| *short shoot* | *whose where* | *step stop* |
| *share shape* | *week weak* | *shake share* |
| *wide wife* | *tree true* | *scale scene* |

In the spirit of finding a magical solution, we thought that it would be valuable to combine the whole word shape and first-last letters solutions and determine if these two factors in combination provide sufficient information for reading words. We found that only 9% of the 1,000 most frequent words was uniquely defined by their exterior letters. Adding word length as a defining feature increased this percentage to 40%. In comparison, only 24% of the words has a unique word shape. When exterior letters, interior word shape, and length were considered as features, 75% of the thousand most frequent words was uniquely described. At first glance, the reader might believe that three out of four times is not bad. However, this requires the reader to recognize the first and last letters, the length of the word, and the word shape of the interior letters. This is not a trivial amount of processing to bypass a strategy simply of processing the letters of the word.

Although we have rejected minimalist hypotheses about reading words, we have not yet accounted for the magic of word recognition. What is it about words that make them so easy to recognize by the expert reader? To better appreciate how words are read, it is important to understand that readers can operate reasonably well with partial information but sometimes must falter. This is a common outcome in pattern recognition more generally. We recognize our friend in a crowd and then discover it was not our friend. Another friend who shaved his beard goes unnoticed. All of us have experienced misunderstanding a sentence because we recognized a word incorrectly. This shows that we do not usually require complete unambiguous information before making a decision in word recognition. Second, we use multiple sources of information in pattern recognition. Many sources of nonvisual information supplement the featural information from the letters. In our infamous paragraph, syntactic and semantic constraints facilitated its reading. A colleague's skilled fourth grade reader had trouble with the paragraph, ostensibly because she had less knowledge that was critical to reading its visually degenerate