

VISIBLE LANGUAGE

The quarterly concerned with all that is involved in our being literate

Volume XVIII Number 1 Winter 1984 ISSN 0022-2224

- 5 Reading Research in Metalinguistic Awareness:
Findings, Problems, and Classroom Applications
David B. Yaden, Jr.
- 48 Before Numerals
Denise Schmandt-Besserat
- 61 Effects of Chunking and Line Length on Reading Efficiency
Stacey A. Keenan
- 81 The Visual Editing of Texts
Fernand Baudin
- 87 Hebrew Micrography: One Thousand Years of Art in Script
Leila Avrin
- 96 The Authors

VISIBLE LANGUAGE, Volume XVIII, Number 1, Winter 1984. Published quarterly in January, April, July, and October. Postmaster: send address changes to Visible Language, Box 1972 CMA, Cleveland, OH 44106. Copyright 1984 by Visible Language. Second-class postage paid at Cleveland, Ohio, and at additional mailing offices.

Merald E. Wrolstad, Ph.D., Editor and Publisher
P.O. Box 1972 CMA, Cleveland, OH 44106

Advisory Board

Colin Banks, Banks and Miles, London
Naomi S. Baron, Brown University
Fernand Baudin, Bonlez par Grez-Doiceau, Belgium
George Bauer, University of Southern California
Pieter Brattinga, Form Mediation International, Amsterdam
Gunnlaugur SE Breim, London
Dennis Fisher, Aberdeen Proving Ground, Maryland
I. J. Gelb, Oriental Institute, University of Chicago
James Hartley, University of Keele, England
Albert Kapr, Hochschule für Grafik und Buchkunst, Leipzig
Paul Kolers, University of Toronto
C. L. Lehman, Tigard School District, Oregon
Dominic Massaro, University of California, Santa Cruz
Peter Mosenthal, Syracuse University
Alexander Nesbitt, Newport, Rhode Island
Thomas Ockerse, Rhode Island School of Design
P. David Pearson, University of Minnesota
Sharon H. Poggenpohl, Middletown, Pennsylvania
Marvin A. Powell, Jr., Northern Illinois University
Wayne Shebilske, University of Virginia
Jack W. Stauffacher, The Greenwood Press, San Francisco
William C. Stokoe, Jr., Gallaudet College, Washington
Michael Twyman, University of Reading
Richard Venezky, University of Delaware
Peter Wason, University of London
Dirk Wendt, Christian-Albrechts-Universität, Kiel
Hermann Zapf, Darmstadt, Germany

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by *Visible Language* for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of \$1.00 per article, plus \$0.10 per page is paid directly to CCC, 21 Congress Street, Salem, MA 01970. 0022-2224/84 \$1.00 + .10.

General Information

Visible Language is concerned with research and ideas that help define the unique role and properties of written language. It is a basic premise of the Journal that writing/reading form an autonomous system of language expression which must be defined and developed on its own terms. Published quarterly since 1967, *Visible Language* has no formal organizational affiliation. All communications should be addressed to

Visible Language
Box 1972 CMA
Cleveland, OH 44106 USA

Telephone 216/421-7340

Subscription Rates.

	<i>One Year</i>	<i>Two Years</i>	<i>Three Years</i>
Individual subscription	\$15.00	\$28.00	\$39.00
Institutional subscription	25.00	47.00	66.00

Foreign subscribers: add \$2.00 for postage to each year's subscription.

All orders must be prepaid. To be honored free of charge, claims for missing issues must be made immediately upon the receipt of the next published issue.

Back Copies and Reprints. A limited quantity of all back numbers is available at a per issue cost of \$4.00 to individuals and \$5.00 to institutions. A booklet listing the contents of all past Journal issues is available on request. Individual reprints are not available.

Manuscripts. Manuscripts, inquiries about research articles, and other contributions to the Journal should be addressed to the Editor. A guide for the organization, preparation, and submission of manuscripts is available. Three copies of a manuscript should be accompanied by an abstract typed on a separate sheet of paper.

Letters to the Editor. Comments on articles and letters that have appeared in the Journal are welcome and should be addressed to the Editor. The Editor will also relay to the author questions or comments on any article. Your response — and the author's comment in reply — will not be published without your permission and your approval of any editing.

Advertising. Detailed information is available upon request.

Young Ashley Smith in a moment of metalinguistic frustration at the Emory University Reading Center in Atlanta, Georgia. Photograph by Billy Howard, director of photography at Emory University.

Reading Research in Metalinguistic Awareness: Findings, Problems, and Classroom Applications

David B. Yaden, Jr.

Metalinguistic awareness — defined as the ability to reflect upon and analyze the structure of both spoken and written language — is discussed in view of its relationship to the acquisition of reading in young children. The corpus of existing research literature is grouped and examined under three broad categories which are indicative of major lines of research: (a) Concepts about the nature, purposes, and processes of reading, (b) concepts about spoken language units and instructional terminology, and (c) knowledge of print conventions and mapping principles. Examined in other major sections are issues related to the direction of cause between metalinguistic abilities and reading, disparities in research methodology between studies, and commercial instruments purporting to measure metalinguistic knowledge. A major conclusion reached by the review is that although young children are largely unaware of the overriding structure of both speech and print, experience with written language is the most efficient way to enhance metalinguistic growth. Implications for reading instruction and directions for future research are suggested as well.

During the past decade and a half there has been increasing interest among language and reading researchers in the ability of young children to consciously and deliberately reflect upon and analyze the structure of both oral and written language as opposed to merely reacting to its content. This capacity for what is most commonly known as “metalinguistic awareness” (Gleitman & Gleitman, 1979; Gleitman, Gleitman, & Shipley, 1972; Holden, 1972) or sometimes just “linguistic awareness” (Mattingly, 1972, 1979; Ryan, 1980) is believed to encompass a variety of language behaviors including the ability to comment upon the grammaticality of certain types of utterances (de Villiers & de Villiers, 1974; Gleitman, Gleitman, & Shipley, 1972), to segment the stream of speech into words (Tunmer, Bowey, & Grieve, 1983), syllables and phonemes (Liberman, Shankweiler, Fischer, & Carter, 1974), and to understand the conventions of the written language system as well — the latter capability being more specifically described as “orthographic linguistic awareness” (Day, Day, Spicola, & Griffen, 1981). Surprisingly, however, research examining the above behaviors

has disclosed that many young children first learning to read exhibit a widespread inability to think of *language qua language* and oftentimes seem to misunderstand the very nature and purposes of the reading act itself.

While there exist reviews of the literature and critiques of the research regarding the relationship of metalinguistic awareness to language acquisition and growth (Sinclair, Jarvella, & Levelt, 1978; Tunmer, Pratt, & Herriman, 1984) and to aspects of developing cognition (Hakes, Evans & Tunmer, 1980), there are fewer such comprehensive treatments touching upon the broad range of metalinguistic abilities and their direct application to the acquisition of literacy behaviors (cf. Downing & Valtin, 1984; Henderson, 1981; Henderson & Beers, 1980; Yaden & Templeton, in press). Of the more widely quoted extant summaries, Ryan (1980) has only hypothesized the relationship between certain metalinguistic skills and reading. An earlier review by Ehri (1979), the most detailed critique to have appeared, defined metalinguistic abilities narrowly and omitted most of the research which explores children's perceptions of what reading actually is. Even Downing's (1979) presentation, perhaps the most cogent argument for a conceptual foundation to beginning reading ability, focused his discussion around the data supporting oral language segmentation as prerequisite to reading, a view not shared by others (e.g., Donaldson, 1978; Ehri, 1979). It is most significant to note overall that given the range of variations of problem focus, method of data collection and unit of analysis between examinations of metalinguistic abilities and reading, there is virtually no consensus as to exactly what emerging conceptual abilities, if any, might be crucial in enhancing those first steps in learning how to read.

The primary purpose of the following review, therefore, is to bring together and discuss a broad range of data-gathering studies exploring children's concepts about the reading act, linguistic units, and properties of the written language system under the general rubric of "metalinguistic." The main body of the review is organized into three major sections, each representative of a distinguishable strand of research within the general corpus of the literature: (a) concepts about the nature, purpose and processes of reading, (b) concepts about spoken language units and terms in the "reading instruction register" (Downing, 1976), and (c) knowledge of print conventions and mapping principles. Further divisions within the major sections have been made according to the varying data collection procedures employed. To give the reader some sense of the history and cumulative progress of the research to be discussed, studies within each subsection of the review

are presented in chronological order, based upon their appearance primarily in English-language journals (see Downing, in press, and Valtin, 1984, for a review of studies in non-English-speaking countries). Similarly, the ordering of the three major research strands follows the approximate development of interest in the field, although by the late 60's research was being carried out simultaneously in all three areas delineated. Following the main presentation of research, other major sections will summarize questions and research involving the direction of cause between metalinguistic abilities and reading achievement, the disparities between research methodologies, commercial instruments purporting to measure metalinguistic knowledge, and instructional considerations to be made in the teaching of reading based upon current findings.

It is necessary to distinguish at this point between the focus of the following research and related investigations into "metacognition" and reading (e.g., Brown, 1980, Brown & Palincsar, 1982) and "comprehension monitoring" (e.g., Wagoner, 1983). While it may be accurate that metalinguistic abilities are merely one facet of a general growth in cognition (e.g., Hakes et al., 1980; Ryan, 1980; Tunmer & Bowey, 1984) that allows a person to "think about his/her own thinking" and to engage in other metacognitive acts where conscious examination of the actual processes of *mind* takes place (cf. Flavell, 1976), reading research being tagged as "metalinguistic" generally focuses upon preschoolers', kindergartners', or first graders' developing notions of the purposes and processes of literacy acts and structural properties of either their own speech or the written language system. On the other hand "metacognitive" studies as a rule examine the development of comprehension strategies in both children and adults, being interested in such questions as "What do readers know about what they comprehend and how they comprehend?" (Wagoner, 1983, p. 329). Thus, it can be observed that metalinguistic investigations study behaviors that are developmentally prior to the growth of comprehension processes needed to understand fully the messages in written texts. It is of interest to note as well that there is little overlap, if any, between the reference lists of individual studies in the following body of research and those investigations studying behaviors described as "metacognitive" which have been excluded.

One concluding caveat is perhaps warranted before the review begins. Readers familiar with the more traditional use of the term "metalinguistic" as an adjectival form of "metalanguage" (Burchfield, 1976, p. 909) in the literature of philosophy and logic or as a description of a branch of linguistics which examines "the relation of language to

the rest of the culture" (Pei & Gaynor, 1954, p. 135) will note little similarity to the usages of the term in this paper. In the literature to be cited at least, the term has taken on as its referent varying states of psychological awareness as opposed to merely being a description of types of statements made in a metalanguage about another object-language (Cherry, 1980, p. 82; Edwards, 1972, p. 68).

Interestingly, a close analog still exists, however, between the traditional meaning of metalanguage and Downing's (1976) "reading instruction register" in that the latter refers to terms used to talk "about" properties of language systems. Hence, linguistic descriptors and terms in the reading instruction register such as word, syllable, phoneme, sentence, etc., are in the technical sense truly "meta-linguistic." Perhaps the primary insight that educational research in the 70's and 80's has added to the traditional nuances of the term has been that in order to speak in metalanguage and use metalinguistic vocabulary appropriately, one has to also be able to "think" metalinguistically. And this latter capability as will be shown in the following research develops slowly and exists in varying degrees among the population of young children learning to read.

Concepts about the nature, purposes, and processes of reading

Incongruous as it may sound given the long history of teaching reading, researchers have reported that prior to 1960 relevant literature on children's perceptions of reading act was "virtually nonexistent" (Denny & Weintraub, 1963, p. 363). It is not fair to say, however, that early professionals in reading were unaware of the disparities between children's notions of what the act of reading ought to entail since Betts (1946) devoted an entire chapter to "Basic notions about reading" (although from an adult's point of view) and at least mentions in passing that some children indeed were observed to "entertain some rather weird notions about reading" (p. 281). It can be said, however, that the pervasiveness of these "weird notions" in most children learning to read is a discovery only of systematic research in the last quarter century or so when the children themselves have been asked directly about these matters.

One of the first extensive discussions of children's disparate concepts concerning the functions and processes involved in reading is reported in a dissertation by McConkie (1959). Interviewing 81 five-year-olds from middle and lower class families, McConkie noted six categories of response to the question, "What do you think reading is?" Responses ranged from definitions as "Reading is telling stories" or "Reading is writing" to "Reading is looking at pages and studying them" and

"Reading is when you look into books, then you go home" (pp. 104-105). A "frustrating" aspect of the interviews, according to McConkie, was that only a very few children (11%) could express "that they perceived reading as a means of securing information" (p. 107). In addition, across all categories only about a fourth of the children indicated that reading had anything to do with looking at letters or words. However, among these children, McConkie also included those who thought that they looked at "numbers," "things," and "names" as well.

McConkie also asked children how they would teach someone else to read. Interestingly, only one child out of the entire sample said that he would teach someone by helping them to sound out words and letters. Other categories included responses such as "I'd teach him by making him listen" or "He'd talk about the pictures in the book, that's reading." Perhaps the most interesting response was, "I would have him learn the 'elephant'; I know all of mine" (pp. 128-129). In short, most of the children interviewed thought that others learned to read by retelling stories that they'd heard, talking about pictures, or "guessing" at words. One summary observation by McConkie was that "children have quite different perceptions of what constitutes an ability to read" (p. 115). An interesting finding as well was that children in the upper socio-economic class provided usually more "adult-like" comments in defining reading, even though the lower class children were similar in intelligence. Thus, McConkie's findings imply, as do others (e.g., Ferreiro & Teberosky, 1982), that early experiences with books provide children with insights into more conventional notions of reading such as they can expect to be expressed by teachers.

The most widely quoted early investigation of children's slowly developing notions of what reading is for and how it is to be accomplished is Reid's (1966) study of a dozen five-year-olds beginning school in Scotland. Interviewing the children at the beginning, middle, and end of their first year in school, Reid asked a "kernal" set of questions designed to probe their understanding of "technical vocabulary — the language available to them for talking and thinking about the activity of reading itself" (p. 56). In quizzing the children with such questions as "What is in books?", "How does your mummy [sic] know what bus to take?", and "What are these spaces for?", Reid observed that initially only one child out of twelve said that books contained words; several thought that the pictures, not print, carried the meaning; most of the children in the sample used the term "numbers" to refer both to letters and numerals. Further Reid noted that the children seemed unaware that letters stood for sounds in words and more often used single

letters to refer to whole words such as “h” for “horse” when asked to describe what they had written. Reid concluded that her subjects were “exhibiting certain linguistic and conceptual misunderstandings about the nature of the material they had to organize” (p. 61). Taking a Piagetian view of cognition, she suggested that the “resolution of these uncertainties” lay in an understanding of the relationship between classes and subclasses. “In short, the children had to come to see that language and pictures are two kinds of symbols, that ‘names’ form a subclass in the class of written words, and that capitals form a subclass in the class of ‘letters’” (p. 61).

While from the first to the third interview, the children showed progress in more fully understanding these relationships, progress was slow and success not uniform. Thus Reid suggested that perhaps a “fostering of the understanding of classification, order and regularity” (p. 62) might be simultaneously emphasized with reading instruction in order to help children make the connection between written and spoken language and better understand the relationship between letters and words.

Reporting in the same year, Denny and Weintraub (1966) interviewed over 100 entering first graders of varying ethnicity and socioeconomic class with the following three questions: (a) “Do you want to learn how to read?”, (b) “Why?”, and (c) “What must you do to learn how to read in the first grade?” (p. 444). Placing the responses into categories ranging from “vague, irrelevant, and circular” to ones indicating an expressed purpose for reading, the authors noted that 25% of all the responses fell into the categories indicating “vague and meaningless reasons for wanting to learn to read” (p. 444). For the third question as well, over a third (38%) of the children “offered no meaningful explanation of what one must do to learn to read” (p. 446). The authors pointed out, however, that several confounding variables were present. For example, children with no prior kindergarten experience gave more responses categorized as “vague and irrelevant” while the middle class children in the sample gave the fewest responses in these categories. Despite these limitations, Denny and Weintraub suggested that the need existed “for helping pupils see a reason for learning to read and for gaining some insight into how it is going to be accomplished” (p. 446).

Using the interviewing techniques of previous studies, Mason (1967) asked a sample of preschoolers four basic questions: (a) “Do you like to read?”, (b) “Would you like to be able to read?”, (c) “Does anyone in your family read?”, and (d) “Do you like him/her/them to read?” (p. 130). Surprisingly, Mason discovered that most (90%) of the pre-

schoolers already thought they could read and enjoyed doing it! From this finding, Mason posited that "one of the first steps in actually learning to read is learning that one doesn't already know how" (p. 132). He further suggested that coming to this realization of not knowing how to read was a part of reading readiness which needed to be explored by future research.

Downing (1969, 1970) replicated Reid's (1966) first interview study but included as well some concrete aids to ensure that the misunderstandings that Reid found were not merely an artifact of the verbal interview. He also included an experimental portion probing knowledge of specific language units which will be discussed in the next major section. Downing (1970) generally confirmed Reid's earlier conclusions. In his sample of 15 English five-year-olds, no child mentioned that books contained words and several thought that their parents only looked at the pictures when they read. In addition, some of the children used the term "numbers" to describe both letters and numbers which they had produced and none of the children seemed to be sure of exactly what the numbers on the buses might indicate. From the replicated interview, at least, Downing (1970) corroborated Reid's findings that "young children . . . have only a vague notion of the purpose of the written form of language and in what activities the reading task consists" (p. 109).

Use of the concrete aids, however, produced somewhat different results. Whereas only one of the children in Downing's (1970) sample mentioned in the interview that their parents looked at the symbols when they read, half of them when given a book pointed to the print and described it as being either "the writing," "words," or "letters." In addition, while none of the children in either Downing's (1970) or Reid's (1966) sample mentioned the destinations of the buses when asked how their mothers knew what bus to take, 11 out of 12 of Downing's children pointed to both the number and the destination board when given the toy buses to handle. Even though from further testing, Downing concluded that children have a very poor grasp of spoken and written language units in the abstract, he admitted that they were able to demonstrate more competency in identifying these units when in the presence of functional objects displaying printed forms.

In a sequence of studies Johns (1972, 1974) explored the relationship between concepts of reading and actual reading achievement as measured by a standardized reading test. In the first study, Johns (1972) asked the question "What is reading?" to a sample of 53 fourth graders and recorded their responses into "meaningful" and "non-

meaningful” categories. While the correlations were relatively low, Johns found a slight positive relationship between concepts of reading and the vocabulary ($r = .31$) and comprehension ($r = .27$) subtests of the *Gates-MacGinitie Reading Test* (Gates & MacGinitie, 1965). In the second investigation, Johns (1974) asked the same question to an additional sample of 50 fifth grade children but divided the total sample of fourth and fifth graders into above and below average readers based upon their grade equivalent scores from the comprehension subtest. He then compared the type of reader against meaningful and non-meaningful responses. Not surprisingly, Johns found that good readers gave significantly more meaningful responses, although less than half of the “good” reader group gave them. The author concluded by saying that the question “What is reading?” may be interpreted differently even by good readers and that additional research needed to take into consideration of not only how to more adequately assess concepts of reading but also how to more accurately identify good and poor readers.

Also asking children the question “What is reading?” and others such as “Can you read?” and “What do people do when they read?”, Oliver (1975) found that for a sample of preschool, native American children, most four-year-olds associated reading with behaviors such as “blowing the nose,” “putting on glasses,” and “just looking” (p. 868). In contrast to the three-year-olds in the sample, half of whom said they could read already, the majority of five-year-olds said they couldn’t but would like to learn. Oliver also reported that some of these older children described the activity of reading more precisely by indicating that people looked at words and letters when they read. However, while the five-year-olds demonstrated more knowledge of what reading entailed, Oliver stated that, for the most part, “these preprimary children generally seemed to lack a clear concept of written language as coded speech and generally seem to think of reading and writing as something they will learn to do ‘when they get bigger’” (p. 869).

In testing certain psycholinguistic hypotheses that children should view reading as a silent process aimed at gaining meaning, Tovey (1976) discovered that in response to the question “What do you think you do when you read?”, 29% of a sample of 30 first through sixth graders viewed reading as “spelling,” “talking,” “memorizing,” and “breathing” (p. 537). In addition Tovey noted “confusion, ambivalence, and uncertainty” in responses to the question, “Do you look at every word when you read?”. While most of the children said that they did look at every word, a third of these felt that it wasn’t really necessary.

And of those who said they didn't look at every word, most thought they should be. For his sample, at least, Tovey stated that reading seemed to primarily a "word calling" process rather than one of getting meaning from the printed page.

In by far the largest and most comprehensive of examination of children's perceptions of the reading act, Johns and Ellis (1976) asked 1655 predominantly white students in grades one to eight the following three questions: (a) "What is reading?", (b) "What do you do when you read?", and (c) "If someone didn't know how to read, what would you tell him/her that he/she would need to learn?" (p. 119). In response to the first question, the authors reported that a staggering 69% of the sample gave answers that were classified as either "vague, circular, or irrelevant." In addition, only 5% of the students indicated a concept of reading that included both word recognition and meaning (p. 120). For the second question as well, over half the sample gave "meaningless" responses and only a fifth of the sample indicated that reading involved the getting of meaning. In response to being asked how they would teach someone else to read, Johns and Ellis reported like McConkie (1959) did earlier that a large portion (36%) of students could not relate an intelligible procedure that a nonreader might follow in learning to read. In fact, only 8% of the sample indicated that in learning to read one must focus on both decoding and meaning. Over half of the sample (56%) in response to the third question felt that reading was primarily a process of either decoding or the learning of rules and grammar (p. 124).

In concluding their investigation, Johns and Ellis noted that while older children in grades six through eight seemed to have more conventional notions of what learning to read entails, "the vast majority of children have little or no understanding of the reading process" (p. 127). An important finding as well was that slight sex differences were found in the data favoring girls' greater understanding of the functions and processes of reading. Johns and Ellis suggested then that a potential field for future research in this area involved examining concepts of reading as they are influenced by norms of socialization for boys and girls.

The final study reviewed in this section by Mayfield (1983) replicated very closely the findings of earlier investigations, particularly that of Johns and Ellis. Interviewing 82 kindergarten children equivalent in age, socio-economic status and prior learning with the questions (a) "What do you think reading is?", (b) "Can you read?", (c) "Who do you know who can read?", and (d) "What does reading help you to do?", Mayfield reported that prior to specific instruction on the types of

graphic codes which exist in the environment, 73% of the sample gave only "vague" and "circular" responses to the first question. Even after specific and direct instruction in types of code systems, over half of one experimental group still could not provide a meaningful definition of reading. Mayfield did discover, however, that both direct and incidental instruction as to the existence of various types of written codes in the environment helped children more correctly identify literacy behavior as measured by the *Linguistic Awareness in Reading Readiness (LARR) Test* (Downing, Ayers, & Shaefer, 1982).

In summary, all of the studies reviewed reported that young children, even those with several years of reading instruction, did not view the process of reading as being a meaning-getting activity. Indeed, most studies indicated that the majority of students could not provide an intelligible description at all. Most often it was observed that children viewed reading within the confines of a specific school-related task like learning the alphabet, doing workbook pages or retelling stories to the teacher. It was also common for beginning readers to believe that pictures, not print, are what is to be "read" and to use the terms *letter* and *number* interchangeably.

Interestingly, a trend noted in several studies was that a more conventional, adult-like perception of the process of reading as being a combination of decoding print and getting meaning (not necessarily in that order) was very slow in developing and surprisingly impervious even to direct instruction. In fact, those children who gave more accurate depictions of reading were usually those who had had prior and plentiful independent experiences with books in the home. Finally, it is fair to say that in all of the studies reviewed, there are strong suggestions made to classroom teachers that the provision of a book-rich environment with a variety of activities wherein the children can experiment with and use print is perhaps more important early on than specific rule learning and an exclusive focus upon learning letter/sound correspondences.

Concepts about spoken language units

Within the general body of metalinguistic research as it relates to beginning reading ability, the studies exploring children's awareness of the components of their speech and their "verbalizable" knowledge (cf. Templeton & Spivey, 1980) of the metalinguistic terms word, letter, sound, etc., are by far the most extensive. Underlying the concern with speech segmentation in particular is the belief that unless a child is aware of his speech as being comprised of a temporal succession of sounds (cf. Bradley & Bryant, 1983; Elkonin, 1973), he/she will have

less success in understanding the form of written words as made up of sequences of letters. There is, however, a great debate carried on in this section and in the field in general since some researchers (e.g., Ehri, 1975, 1976) feel that it is rather exposure to the written form of language that provides insight into the fact that one's spoken language includes several kinds of identifiable units. This discussion will be taken up again later in the paper.

Studies of children's knowledge of oral language units (i.e., words, syllables) generally fall into three categories distinguished by the methodology used in assessing beginning readers' ability to isolate or identify these units in the speech stream. The most common strategy used is a "word tapping" task in which the child repeats a sentence and counts each word by tapping on the table with a pencil or similar object. A variation of this task has been to have the child point to wooden blocks or poker chips as each word is spoken. A second task used is to ask the general question "What is a word (letter, sentence)?" These investigations tend to seek out developmental trends in that they not only point out disparities between children's and adults' notions of language units, but also gather information on what exact concepts children possess at different ages. A third strategy adopted by fewer researchers involved selecting *a priori* categories of verbal and nonverbal "sounds" and training the subjects to respond "yes/no" when they thought they heard a single sound (phoneme) or word. A more detailed discussion of investigations in each category directly follows.

Word, syllable, and phoneme segmentation

One of the earliest attempts to observe children's ability to segment speech into words was Karpova's study (1966) in Russia in 1955 with a sample of children ages 5-7. Karpova asked children to repeat sentences and respond to the questions, "How many words are here?" and "Which is the first . . . second . . . third word?" Karpova reported that the youngest children (ages 4-5) did not isolate words but rather semantic units. For example, a child aged 4-6 years indicated that the sentence "Galya and Vova went walking" had two words: "Galya went walking, and Vova went walking" (cited in Smith & Miller, 1966, p. 370). Under repeated questioning, children approaching 7 were beginning to isolate nouns and began to break sentences into subject and predicate. It is reported also that some of the oldest children in the sample isolated all of the words correctly excepting functors as prepositions and conjunctions. Karpova also instituted a training procedure in which children moved plastic counters as they repeated

each word. The procedure apparently was quite successful for the children who initially could not segment any words.

Another early study by Huttenlocher (1964) designed to assess word awareness investigated the ability of 66 children ages 4 1/2 and 5 years to either reverse word pairs of different grammatical and nongrammatical relationships or to say the first word of the pair, await a "tap" from the researcher, then say the second. The sample was randomly divided into two groups with each group performing only one of the tasks. Huttenlocher discovered that a third of the children in each group were unable to reverse or segment any pairs. For the remaining subjects, the most troublesome categories involved reversing or segmenting common grammatical sequences such as "man-runs," "I-do," or "is-it" (p. 264). Huttenlocher then hypothesized that children's confusions as to the identification of a single word might particularly come with words not ordinarily used in isolation such as copulatives and pronouns.

In another frequently quoted article, Holden and MacGinitie (1972) generally confirmed Huttenlocher's suspicions that prepositions and auxillaries were not seen as distinct units by young children. In a tapping task where the child repeated an utterance and simultaneously pointed to individual poker chips to indicate a word, the majority of subjects when presented with the sentence "You have to go home" either combined "to" with "have" or "to" with "go." Similarly, when the verb "to be" was used as an auxillary in the progressive form "Bill is drinking sodas," kindergarteners generally made the combination "isdrinking" and a few chose "Billis." "In general," concluded Holden and MacGinitie, "the greater the proportion of content words in an utterance, the greater the percentage of correct segmentations" (p. 554).

In one of the first attempts to correlate awareness of word boundaries with actual reading achievement, McNinch (1974) found that with pre-established readiness groups (good, average, poor) ability to segment a spoken sentence into words did not significantly differ. However, in a multiple regression with visual word boundary scores, oral segmenting ability was the significant predictor of end of the year reading scores on the *Metropolitan Achievement Test* (Prescott, Balow, Hogan, & Farr, 1971). Similarly, Evans (1975) reported that for a sample of 45 kindergarteners and 45 first graders divided into above and below average groups based upon a segmenting task identical to Karpova's (1955/1966), better readers in December as measured by the *Gates-MacGinitie Primary Reading Tests* (Gates & MacGinitie, 1965) were also the better segmenters a few months earlier.

Like Elkonin (1973), Liberman (1973) and Liberman, Shankweiler, Fischer, and Carter (1974) also have offered evidence of the existence of a relationship between syllable and phoneme segmentation and instruction in reading. Employing a tapping task where 135 preschool, kindergarten and first grade children were to identify phonemes and syllables in spoken words, Liberman et al. found main effects for both task and grade, observing that phoneme segmentation was uniformly more difficult for all groups than syllable segmentation and that first graders performed better than kindergartners who, in turn, performed better than preschoolers. While Liberman et al. did not discount the effects of maturation, they posited that in all probability "analysis of language, even of the most elementary sort, requires instruction" (p. 210).

In contrast to the findings of previous analyses of children's inability to segment spoken sentences, Fox and Routh (1975) claimed that even three-year-olds were able to segment sentences into words, words into syllables, and in a few cases even syllables into individual phonemes. Fox's and Routh's task was to have the children listen to a sentence, word, or syllable spoken by the researcher and then respond to the statement "Say just a little bit of it" (p. 335). This statement was repeated until all the words or sounds were completely analyzed. The results showed that ability to analyze the items steadily increased with age. However, even three-year-olds segmented over half of the sentences into words, approximately a third of the words into syllables, and a fourth of the syllables into individual phonemes. These findings contradict earlier statements by Bruce (1964) that until a mental age of 7, children are unable to competently perform word analysis tasks. Fox and Routh also found significant positive correlations between reading comprehension as measured by the *Peabody Individual Achievement Test* (Dunn & Markwardt, 1970) and ability to segment words into syllables and syllables into phonemes.

Ehri (1975), in addition to a word and syllable segmentation measure, also tested children's ability to analyze a sentence for target words and analyze spoken words for specified syllables. Using a sample of preschool, kindergarten, and first grade children, Ehri found that for most tasks, readers' (first graders) mean performance was higher than prereaders (preschool and kindergarten) while the means for the latter two groups did not differ. As a result of their more frequent exposure to printed language, Ehri stated that "readers, in contrast to prereaders, possess substantial conscious awareness of lexical as well as syllabic constituents of speech" (p. 211). As did other researchers (e.g., Holden & MacGinitie, 1972; Huttenlocher, 1964), Ehri noted that

all of her subjects, particularly the prereading groups, failed to distinguish functors such as "the," "a," "to," and "is" as distinct units of language.

In a similar kind of sentence analysis task designed to tap children's lexical awareness, Holden (1977) tested 26 kindergarten and 24 first grade children on their ability to identify the added word in one sentence of a pair of sentences with homophonous words (e.g., "John leaves after dinner; John rakes leaves after dinner," p. 214). Based upon a previous study (Holden & MacGinitie, 1973) which demonstrated that differences between kindergartners and first graders on this same task was not a result of intelligence, Holden surmised that regular patterns of response should occur if indeed there were developmental stages in young children's evolving awareness of word units. Holden's findings supported this notion since the first graders made almost twice as many correct responses as did the kindergartners. Further, Holden demonstrated that short-term memory recall limitations could hardly have been a factor since for many of the incorrect responses, the children often repeated the entire verbal stimulus without error except in a fraction of the cases. Holden observed, however, that even the better performing first graders still exhibited an "unstabilized ability to perceive language at both phonetic and semantic levels simultaneously" (p. 206) as their most common error was to isolate the homophonous word which had changed meaning in the second sentence. Holden concluded, therefore, that the capacity to analyze language "abstractly" apart from its semantic context shows definite developmental patterns (p. 206). She did not discuss, however, how these patterns might be affected by prior experience with books or direct reading instruction.

Another study employing a tapping task examined under this subcategory of speech segmentation was conducted by Leong and Haines (1978). Testing a total sample of 72 children in grades 1-3, the researchers had children segment words into syllables and syllables into phonemes by tapping a wooden dowel on the table as they distinguished each unit spoken. In addition, there were also tasks of identifying the number and order of sound patterns in words (cf. Lindamood & Lindamood, 1971) and recall of sentences varying in grammatical complexity (i.e., "high" or "low"). Results showed that while there was a significant difference across grade in ability to segment words into syllables, there was no difference between groups in segmenting syllables into sounds. However, in the "auditory conceptualization" task of recognizing the number and order of sounds of words, there was a significant difference between grades 2 and 3 combined and grade 1.

To further investigate the relationship of auditory conceptualization, word and syllable segmentation with reading achievement, a canonical correlation was computed with the experimental tasks as independent variables and two measures of reading achievement as the dependent variables. The analysis showed that auditory conceptualization or the recognizing and ordering of sound sequences in words contributed most to the correlation with reading scores ($R = .777$). This was followed in the weightings by recall of high complexity sentences, syllable segmentation, phoneme segmentation, and recall of low complexity sentences (p. 402). Despite the finding unlike previous studies that phoneme segmentation did not discriminate between grades (cf. Liberman et al., 1974), the authors suggested that for some children "acquisition of verbal skills is facilitated if their understanding is brought to the focal level . . . This contemplation of words and sentences, which can be taught . . . , will go some way towards helping the child in the learning to read process" (p. 405).

The final two studies reviewed in this subsection have provided some necessary controls over the mechanics of the data-collection task and the nature of the stimuli themselves not included in previous investigations. In the first, Treiman and Baron (1981) included a nonsense sound counting task along with syllable and phoneme counting to ensure that children could indeed perform the mechanical task itself. In addition, they had the first and second grade children move checkers rather than tap to identify phoneme and syllable units since the former task seemed less affected by rhythmic responses as noted in other studies (e.g., Holden & MacGinitie, 1972). In addition, nonsense words were used as stimuli in order to aid the child in "thinking about sounds" apart from their meaning. Interestingly, Treiman's and Baron's results closely corroborated the traditional finding that phoneme segmentation is uniformly harder than syllable segmentation for all children. However, with the inclusion of the neutral sound test, the researchers observed that second graders performed better than first graders on simply the ability to count. Therefore, the authors suggested that the older children's apparent increasing awareness of sounds in words may simply be a result of their superior ability to enumerate. Other results reported by Treiman and Baron included the finding that for some words fricatives were more easily isolated than stops and that speech segmentation for nonsense words at least proceeded in order of difficulty from vowels being the hardest to discriminate followed by final consonants then initial consonants (p. 172). The authors pointed out, however, that previous research indicated a similar pattern for real words.

The last study reviewed in this section by Tunmer, Bowey, and Grieve (1983) provided additional control in the nature of the stimuli to be isolated. In five separate experiments where groups of children 4 to 7 years of age were given word strings to first repeat orally then tap out the number of word units, the investigators varied such factors as grammaticality, plurality, form class, and stress pattern. Their results showed that while there is an increasing ability to segment speech proportional to chronological age, the effects of varying syllabic congruence (i.e., more syllables than words in stimulus), plurality (presense of plural nouns), word class (i.e., adjectives, verbs, nouns, etc.), and grammaticality (grammatical vs. ungrammatical strings) within the stimulus items had little differential effect between age groups. However, in the experiments designed to explore the influence of stress pattern, Tunmer et al. observed that young children segmented primarily according to phrase and syllable stress; whereas the older children in the sample (6-7 years) began to focus upon morphemic units (p. 592). Even so, the authors noted that "most five-year-olds and a few six- to seven-year-olds do not segment meaningful syntactic phrases into their constituent words in the present studies" (p. 590). Interestingly, Tunmer et al. observed also that "explicit demonstrations and corrective feedback" did little to enhance the four- to five-year-olds' notion of an "abstract concept of word as a unit of language" (p. 591). However, the authors concluded by saying that future research must take into account the effects of memory, stress, word awareness, and other factors before it can be decided whether or not children's awareness of units of language can be enhanced by specific training and subsequently if "lexical awareness" has any direct bearing upon learning to read.

In short, studies in speech segmentation demonstrated that pre-schoolers as well as first and second grade children have great difficulty isolating linguistic units in their speech, particularly phonemes or "sounds." In addition, contentives are much more easily picked out of the speech stream than functors. It was suggested also that specific training in segmentation may be less productive since approaching the age of 7 children seem to use a variety of stress cues to anticipate divisions in oral language rather than knowledge of discrete language elements. Finally, researchers in general admit a correlation between reading ability and phonological awareness, but the direction of cause is still much under dispute.

Identifying what is a word?

In the next group of studies the qualitative content of children's notion about words as units of language is explored. The methodology of the following investigations normally involves a researcher posing to a child an inquiry such as, "What is a word?" Hence, the child's verbalizable or "reflective" (cf. Templeton & Spivey, 1980) knowledge is used as the unit of data. The importance of these types of studies, as noted earlier, lies in their ability to discover the evolving stages of a child's concept of word, this time from the learner's point of view.

Testing a group of 50 five-year-olds four times over a two year period, Francis (1973) asked, "Can you tell me any letter (word, sentence) you know?" Following this task, she also showed them an example of each element on a card and asked the children to identify the particular units. On the first testing occasion, half of the children either chose examples of words or sentences when asked to identify individual letters. Words continued to be confused with letters until the last testing at age 7. The results of asking for each concept were very similar to the recognition test. Words were frequently confused with numbers or names, and words were given as examples of sentences. Overall, Francis noted a pattern from the first to the last testing that letters were mastered before words and words before sentences. She also noted that children generally learned the last two concepts after gaining some facility in reading. In addition, she found that reading ability was positively correlated with knowledge of technical language terms (i.e., word, letter) even with I.Q. controlled (Kendall $r = .34$). Francis concluded, therefore, "that factors independent of a general ability to deal with abstract concepts were involved in learning technical vocabulary and that these were closely related to the reading process" (p. 22).

In probably the most well known study of reflective word knowledge, Papandropoulou and Sinclair (1974), using a list of commonly known words, identified four levels in development of word consciousness as a result of asking children 4 1/2 to approximately 11 years of age the questions "Is that a word?" and "What is a word, really?" An analysis of the results showed that most of the children under age 5 answered in level one which was characterized by the inability to differentiate between a word and its referent (cf. Markman, 1976) as exemplified by responses such as "Children are words" or "It can be a cupboard or a chair or a book" (p. 244). Level two (5-7 years) was characterized by two functions of words: (a) as labels for things, and (b) to express a "topic-comment" relationship such as "I put the dog in the kennel" in response to the request, "Say a short word." At level three (6-6.8

years), words began to take on the feature of elements which made up wholes but which did not yet have individual meanings, for instance, "a word is a bit of a story" and "a word is something simple, very simple, it's all by itself; it does not tell anything" (p. 246). Papandropoulou and Sinclair noted that during the fourth and final stage words finally become "autonomous" elements, having meaning of their own and play a definite role in grammatical relationships. Responses to inquiries at this stage take the form, for instance, of "letters form words . . . a word is something that means something" (p. 247). Based on their findings, the researchers concluded that the concept of a word "undergoes a long and slow elaboration during the ages studied. Gradually, words become detached from the objects and events they refer to, and it is only late in cognitive development that they are regarded as meaningful elements inside a systematic frame of linguistic representation" (p. 249).

In a series of related studies, Sulzby (1978, 1979) used a different approach to eliciting students' "metalinguage" in a task designed to explore elementary students' thinking about known and unknown words in both oral and written form. In the first of these studies, Sulzby (1978) examined the responses of 30 rural, predominantly black students in grades 2, 4, and 6 to the question "How does your (student's) word go with my (researcher's) word for you?" (p. 52). On the whole, Sulzby found that students in all grades tended to give answers indicating a semantic focus rather than structural (e.g., "They both got letters"), although this tendency increased across grades. In addition, students in all grades gave mostly semantic responses even to words presented in written form. A very interesting finding by Sulzby was that even sixth graders were observed to be using instructional terminology (i.e., "metalinguage") incorrectly when giving the less frequent structural responses. Sulzby noted as well that by fourth grade, students would create "hypothetical contexts" for unknown words more frequently than give structural responses.

Using the same task, but a different population of 28 predominantly white children in grades 1, 3, and 5, Sulzby (1979) found again that all students gave significantly more semantic responses although more so in the oral presentation mode this time. As in her first study Sulzby noted the tendency for children to create meanings for unknown words rather than give a simpler structural response (p. 52). Both of these studies offer from a slightly different angle evidence that the structural aspects of words, even in written form, if not immediately available for reflection, are subordinated to the child's need to create some kind of intelligible meaning.

In an extension and replication of the Papandropoulou and Sinclair study (1974), Templeton and Spivey (1980) asked a sample of 24 children ages 4.0 to 7.8 years of age such questions as, "Is _____ a word?" (from a predetermined list), "Why is/is not _____ a word?", and "What is a word, anyway?" (p. 268). In addition the children were queried in a similar manner about long, short, easy and hard words as well. Templeton and Spivey also grouped the sample according to performance on the Piagetian concept attainment tasks of classification and seriation and, thus, were able to describe responses as being characteristic of children at the preoperational, transitional, and concrete levels of operation. The results indicated that the preoperational children in particular were unable "to talk about language abstractly" (p. 274), most often refusing to respond. Transitional children, on the other hand, began to give answers which reflected a notion of "wordness" as having something to do with spoken language (i.e., "It comes out of your mouth," p. 274) apart from a specific context.

Interestingly, Templeton and Spivey pointed out that even the more sophisticated responses to questions like "What is a word?", characteristic of concrete operational children, most often reflected the influence of exposure to print (i.e., "We have to read them" or "It's something that you write," p. 275). Therefore the authors suggested that while a more frequent referral to the internal structure of words was in general more indicative of a higher level of cognitive functioning, the ability to think "metalinguistically" seemed to be enhanced by mere exposure to the written language itself.

In another study, Sanders (1981) analyzed first grade classroom interactions by video and audio recordings and then interviewed three first grade males as to their understanding of the teacher's use of instructional terms such as "beginning sound" and "word." Sanders discovered that while students seemed to observably understand classroom directives, individual interviews revealed confusion on the child's part. One child indicated in the personal interview that "Dog and God and big and dig begin alike" (p. 269). The researcher also noted that the subject confused the referents of letters and words as well as "a long word" or a "string of words" (cf. Templeton & Spivey, (1980). Interestingly, Sanders also observed that the interviewees thought that while learning letter/sound correspondences and letter names was useful for first grade, the skills had little to do with reading itself (p. 269). Further, all of the subjects, as noted by the researcher, adhered to the formula of "three letters, plus or minus a letter" (cf. Ferreiro & Teberosky, 1982) in deciding whether a written array was a word or not. In general, Sanders observed that while classroom

activities provided isolated focus upon many metalinguistic aspects of learning about print, accurate notions of what it means to be literate are derived primarily from functional and meaningful interaction with written language.

The last study examined in this subsection exploring children's reflective knowledge of word and other language units was conducted with Spanish-English bilingual children grouped into reader groups of fair-to-good and non-to-poor. Matluck and Mace-Matluck (1983) elicited responses from 94 children in grades 1 to 4 over a 3 year period regarding their knowledge of decoding processes and understanding of the metalinguistic terms "word," "sentence," and "story." Concerning decoding, very few first grade children in either language could explain why they knew how to pronounce a word. By second grade, however, a large majority of the better readers were giving responses which demonstrated some facility using metalinguistic terminology (i.e., "by syllables, by letters — I sound each letter," p. 28). By the third and fourth grades, over 80% of the good readers were giving accurate, adult-like explanations of print deciphering processes while only half of the poor readers could do so.

In response to the statement, "Tell me what a word (sentence, story) is," Matluck and Mace-Matluck again observed that only a very few first graders, mostly good readers, could give formal definitions of these terms. By second grade, although more of the entire sample attempted answers, only good readers again gave more accurate definitions. Even by fourth grade, a majority of the formal definitions of these terms were still being given by the good reader group. Interesting, like other studies (cf. Templeton & Spivey, 1980), definitions of the term "word" reflected the influence of increased exposure to print (i.e., "A word is a group of letters joined together to pronounce a word," p. 33). The authors concluded by reiterating the hypothesis of previous studies that "the development of metalinguistic skills appears to coincide with experience with literacy and to be related to exposure to literacy training" (p. 33).

Briefly summarizing this group of studies, it can be noted that young children in kindergarten and first grade have an extremely difficult time verbalizing their notions of the metalinguistic terms used in classroom instruction. And when these notions were tapped, the concepts seemed to be extricably woven to semantic content rather than including structural dimensions. When children did begin to verbalize more adult-like perceptions of such terms as "word," the influence of print was evident. This finding lends support to the notion expressed earlier by Ehri (1975, 1976, 1979) that until children are

exposed to the written language, they have little reason to view their speech as being made up of discrete, isolatable units. Finally, while it has been observed that an expression of more sophisticated concepts about language often accompanied increasing development in cognition, the relationship is still confounded by prior exposure to print and the influence of classroom instruction.

Identifying verbal vs. non-verbal units

The next small group of studies discussed have generally used the same general paradigm to assess children's knowledge of word, syllable, and phoneme units. Initially, Downing (1970) devised a task in which children were presented 25 tape recorded auditory stimuli of five types: nonhuman noises (bell-ringing) and human utterances of a single phoneme, word, phrase, and sentence. Each child was tested twice with the "sounds" of each category and asked first if he/she heard a single word and then if he/she heard a phoneme. Results of the presentation of the stimuli to 13 English five-year-olds showed that 5 children responded "yes" or "no" to all stimuli in all categories, thus evidencing no discrimination even between verbal and non-verbal sounds. In addition, five children responded positively in the word phrase of the experiment to phrases and sentences as well. No child, Downing reported, correctly identified either a single word or phoneme.

Later, Downing and Oliver (1973-74) extended the categories to include nonverbal "abstract" sounds (i.e., dice rattling), isolated syllables and both long (e.g., hippopotamus) and short words. He also specified in the pretraining task that the children respond "yes" to only single words. Results, however, followed the pattern of the first experiments. All children, across ages gave significantly fewer correct responses for both syllables and phonemes than for any other auditory class while none of the children in the youngest age group (4.5-5.5) recognized that phonemes or syllables were not words. In addition, Downing and Oliver noted that children even up to 6.5 years confused non-verbal sounds, phrases and sentences and phonemes as words. Downing and Oliver stated, therefore, "A more generalized implication of these findings would seem to be that it is not safe for reading teachers to assume that their beginning students understand linguistic concepts such as word" (p. 581).

Johns (1977) replicated Downing's and Oliver's study with a larger sample (120) of American children ranging in age from 5.6 years to 9.5 years and generally confirmed the latter's results. In Johns' study almost 40% of the subjects at beginning reading age were unable to

consistently identify a single spoken word. In addition, nearly 90% of the subjects in this age group confused single phonemes with words. Johns surmised that such confusions "may be due, at least in part to the fragmentation that occurs in reading instruction. Concentrating on sounds (phonemes) and word parts may only serve to confuse children who are trying to learn what reading is all about" (p. 256).

In a more restricted version of the task, Ryan, McNamara, and Kenney (1977) presented above and below average readers in first and second grades with a word discrimination task in which they were to identify single phonemes, two-syllable words, and two phrases as either a "word," "not a word," or "two words" (p. 399). Their results showed that above average readers scored significantly higher than below average readers in correctly identifying the stimuli. Ryan et al. then administered the same tasks to third and fourth grade remedial readers divided into above and below average reading groups by placement in basal readers. They again discovered that better readers outperformed their poorer reading counterparts in identifying linguistic units.

Finally, in the most recent replication of the study by Downing and Oliver (1973-74), Horne, Powers, and Mahabub (1983) tested 40 male students ages 6 1/2 to 10 1/2 on their ability to distinguish from a range of non-verbal stimuli to types of linguistic utterances. Reader and non-reader groups were equated by intelligence and also given pretraining tasks to ensure their understanding of the response required. An ANOVA comparing reader levels, age, and stimulus class showed that the sample of Horne et al. performed similarly to students in the previous two investigations (Downing & Oliver, 1973-74; Johns, 1977), in that readers outperformed non-readers in all classes and that there was uniform difficulty among all pupils in identifying phonemes and syllables as opposed to the rest of the stimuli. An important extension of the Horne et al. study, however, was the inclusion of the oldest group (9.5 - 10.5) and the finding that non-readers in this group mastered *none* of the stimulus groups excepting short words (p. 11), thus indicating extensive confusion about linguistic terminology and concepts about language units.

In summary of the major section, regardless of the method of data collection used, most studies indicated that a great number of primary aged children as well as some of those with several years of schooling were not able to analyze their speech into units such as phonemes or words, with some even unable to distinguish between linguistic utterances and infrahuman sounds. Further, a tendency noted by several researchers was for children to overlook functors as distinct

language entities and primarily focus upon the semantic aspects of words. With the glaring exception of Fox and Routh (1975), whose method of data collection has been questioned (see Ehri, 1979), the majority of studies reviewed consistently reported that children's concepts of their oral language as being comprised of distinct linguistic units were not stabilized, and some implied that these nascent concepts may be resistant even to direct instruction. Finally, it has been commonly observed that children who are better readers also demonstrated greater facility at analyzing their speech into distinct components and verbalizing more precise notions about the nature of words, sentences and other language units.

Concepts about printed conventions

Clay's (1967, 1969) weekly observation of 100 children's beginning reading behavior over a year's period in New Zealand has provided the impetus for numerous investigations into children's specific concepts about printed conventions such as left-to-right/top-to-bottom directionality, marks of punctuation, and especially "space" as a boundary for written words. This latter area has absorbed much of the attention of American researchers, although some have examined the child's gradual development of an accurate speech-print match as well. Studies in the following section, then, will be further subdivided into separate discussions of children's knowledge of visual word boundaries, the correspondence between the spoken and written word, and concepts about directionality, punctuation and other printed conventions.

Knowledge of written word boundaries

The first American investigations of children's knowledge of printed conventions almost exclusively focused upon recognizing written word boundaries. Meltzer and Herse (1969) provided the basic algorithm by having children first read the sentence, "Seven cowboys in a wagon saw numerous birds downtown today" (p. 4). The instructions then were to count each word while pointing to it and finally to circle each word. With a sample of 39 beginning first graders, Meltzer and Herse noted a recognizable developmental pattern: (a) letters are words, (b) a word is a unit made up of more than one letter, (c) space is used as a boundary *unless* the words are short, in which case, they are combined; or long, in which case they are divided, (d) only long words continue to be divided, and (e) spaces indicate word boundaries except where there is a "tall" letter in the middle of a word (p. 13). As a result of these findings, the authors stated that "a very cursory sampling of the

kindergarten seemed to indicate almost complete ignorance after three months of school of graphic characteristics which define . . . a letter or word" (p. 11). Meltzer and Herse also made the intriguing suggestion that this knowledge of printed conventions was not directly taught; "Rather the assumption is made either that the child already has this information or that he will discover it independently from the material presented to him" (p. 13).

Subsequent replications of the above study — while supporting the finding that children do not use space consistently as a boundary for written words — have not confirmed the existence of a developmental pattern however. Kingston, Weaver, and Figa (1972) noted that the most common error in their sample of 45 first graders was that of combining two short words, usually when one contained only one letter (e.g., "andI" or "Isaid"). Kingston et al. observed that other combination errors involving longer, multisyllabic words seemed "to be a result of a failure to perceive any word meaning in addition to the fact that the printers' space was not recognized as a word boundary cue" (p. 95). Such errors were recorded as dividing at ascenders, descenders, and of putting together the end of one word with the beginning of the next. Kingston et al. concluded that "recognizing the printer's space as the separator of words is secondary to perceiving that a particular linguistic unit represents a meaningful entity" (p. 95).

McNinch (1974) also used Meltzer's and Herse's (1969) task in conjunction with an aural word boundary task (word segmentation) with a sample of 60 first graders. The primary finding was that while performance on the visual word boundary task discriminated between readiness groups (high, average, low), it did not appear as a significant predictor of spring reading scores in a multiple regression. McNinch did not report any patterns of word division.

Mickish (1974) tested 117 first grade students at the end of the year on their ability to segment the spaceless sentence "Theca-tandthedogplayball" (p. 20) by drawing vertical lines in between the words. Even though it could be "safely assumed," according to Mickish, that the term *word* had been referred to "hundreds of times," 50% of the subjects did not correctly segment the sentence. Mickish observed also that children in higher levels of basal readers performed better than children at lower levels.

Blum, Taylor, and Blum (1979) also attempted to replicate the task and findings of Meltzer and Herse (1969) with a sample of 54 first graders and 47 kindergartners. Using the same test sentence "Seven cowboys in a wagon saw numerous birds downtown today" and having the children count and circle the words, the authors reported as did

Kingston et al. (1972) that the most common error of both grades was combining two words and that the putative developmental pattern identified by Meltzer and Herse (1969) was not evident. Blum et al. echoed Clay's (1967) earlier admonition, however, that "exposure to 'meaningful' print results in clarity about word space. The nature and pace of this clarity depends on the nature of the child and the quality and quantity of print exposure" (p. 38).

In one of the more descriptive investigations of the nature and development of printed word boundaries, Sulzby (1981) gathered writing samples from nine kindergarten children and recorded their rereadings and explanations of their composing processes. Dividing the sample into high, moderate, and low "emergent" reading groups, Sulzby was able to observe alternative ways of segmenting printed strings such as dots between words, separate lines for each word (i.e., a columnar display), and even circles drawn around letters in order to, as one child put it, "keep the parts from getting mixed up" (p. 14). Interestingly, Sulzby noted that children in the lower two groups asked many more questions about the processes of writing and when reading their productions than did the children in the high emergent group who perhaps, as Sulzby surmised, asked these questions at an earlier age. An important point noted by Sulzby was that although many young children do not use space conventionally, it does not mean that they are unaware of the principle of segmentation itself.

Thus, the few studies reviewed indicated that the convention of "space" as separating word units in print, if not easily grasped by young children, is not used to begin with. While there is less evidence for a distinct developmental pattern, all of the studies indicated that better readers or those having more exposure to print more closely approximated the adult notion of segmentation. Since Meltzer and Herse pointed out that there is little specific instruction in this area, it can be surmised that children were quite successful in gleaned from their printed environment alone some of the characteristics of written language, albeit slowly.

The speech-print match

Studies in the following section generally assess the oral/visual correspondence in one of two ways. Some investigations have explored the spoken/written word match from the standpoint of either too many or too few words spoken for the number of written words represented (Clay, 1967; Holden & MacGinitie, 1972). On the other hand, several have focused attention on whether or not children understand that long spoken utterances generally are represented in print by words

with many letters as well. Reviewed are examinations of both types.

Clay (1967) observed that subjects in her sample went through several stages before correctly matching spoken and written utterances. During the initial stage, children only matched their memorized rendition of a written text by locating the appropriate page with no reference, however, to the actual written text. In stage two the child was able to find the appropriate line of print and during the third stage located some memorized words within the line itself. Stage four was characterized by a process which Clay called "reading the spaces" or "voice-pointing" where the child exaggerated the spaces between words by prolonged pauses between utterances in oral reading. Finally, some children moved into a more fluent stage where oral reading errors were characterized by a "movement speech" mismatch where there are either too many or too few spoken words for written ones or a "speech-vision" mismatch in which substitutions for written words were governed by prior language habits.

In a study mentioned earlier, Holden and MacGinitie (1972) tested a sub-group of 57 kindergartners in their original sample on their ability to match written sentences with previously spoken and segmented ones. In the matching task, responses were scored as "congruent" if the child matched the correct number of written clusters with the oral segments he/she had counted and "conventional" if the number of spoken words matched the number of written ones as normally printed. The written sentences contained both mono- and polysyllabic words and many were segmented unconventionally (i.e., "Red and green balloons popped.", p. 555). Even with prior instruction in the principles of printing convention, Holden and MacGinitie found that only 5 children in the sample were able to correctly count the words in the spoken utterances and match them to their written equivalents. While several children were able to choose a "congruent" written match with the segments they had counted, the authors stated that none of the children consistently picked out the standard written form of the spoken sentences (p. 556). Summarily, therefore, the authors warned that "a first grade teacher cannot take for granted that children will understand her when she talks about 'words' and their printed representation" (p. 556).

Rozin, Bressman, and Taft (1974) tested a total of 218 children in kindergarten, first, and second grades on their ability to recognize and explain why pairs of words such as "mow-motorcycle" and "ash-asparagus" represented different lengths of spoken utterances. The authors reported significant differences in percentage between suburban kindergartners who were able to match the spoken and written

forms correctly (43%) and urban kindergarteners who performed less well (11%). While the urban group improved performance in first and second grades, Rozin et al. noted that a fair number of urban second-graders could still not perform the matching task adequately (76% and 40% in two classes, respectively). Rozin et al. did not offer any explanations as to the differences between socioeconomic groups; however, they suggested that "it might be useful for a child to grasp the nature of the writing system before delving into its detailed specifics (letter/phoneme mappings). It appears that partial mastery of the details does not guarantee appreciation of the basic system" (p. 334).

Using the same task with some variations in the nature of the stimulus pairs, Lundberg and Torneus (1978) asked 100, nonreading children, ages 4-7, to match long or short written words with their appropriate oral representation and to explain the reason why. The researchers varied such factors as vowel duration and semantic referent (i.e., long/short written words referring to either large or small objects). While the results showed a steady increase in correct matching due to age, Lundberg and Torneus reported that less than 20% of the entire sample met the criterion of 90% correct responses (p. 410). In addition, only the 7-year-olds were able to give explanations of their choices which indicated an accurate understanding of the relationship between the duration of spoken utterances and number of written letters. Other trends noted were that children in all age groups seemed to adhere to a semantic strategy when deciding on the word length while no groups demonstrated reliance upon vowel duration as a cue. In summary, Lundberg and Torneus stated that even the oldest preschoolers "seemed to have poor concepts of the basic principle of our writing system" [Swedish] (p. 412). Further the authors warned that "conventional beginning reading instruction with phonic emphasis starts well before the children have developed necessary metalinguistic skills, with serious educational consequences" (p. 412).

Finally, Evans, Taylor, and Blum (1979) used the same task of Rozin et al. as a component in the development of their own instrument to measure metalinguistic abilities. Using a sample of 53 first graders, they found that in a multiple regression with reading achievement as the criterion, the "mow-motorcyle" test was a significant predictor of achievement while knowledge of visual word boundaries was not. They suggested that tasks such as "mow-motorcycle" which require the child to focus on aspects of both oral and written language are more useful in helping the child understand print since they enhance "decision-making by the child and an active interaction with his language" (p. 17).

It can be observed, then, that children do not immediately understand the convention of spacing between written words as separating lexical units in print. Nor do they, as reported, realize that longer utterances are usually represented by more letters. Interestingly, as demonstrated in other studies, direct instruction regarding these concepts seemed to have little effect; whereas increasing experience with books and interaction with the printed page led to more adult-like notions of how spoken words are represented in print.

Directionality, punctuation, and other printed conventions

Most of the studies reviewed in this final subsection of concepts about printed conventions have used the few commercially available tests in the area to measure a variety of reading-specific behaviors. Clay's (1972, 1979) *Concepts About Print Test* (CAPT), the *Linguistic Awareness in Reading Readiness* (LARR) *Test* by Downing, Ayers, and Schaefer (1982) and Blum's, Evan's, and Taylor's (1982) *BET: Written Language Awareness Test* (WLA) were all developed to give more accurate insight into the child's direct facility with reading behaviors than was possible with traditional reading readiness tests. Among the tests a range of concepts about written language are measured including knowledge of printed letter and word units, understanding of metalinguistic vocabulary, correct directional movements, the function of punctuation marks, and in some cases, discrimination between different types of script and cognizance of various kinds of environmental message-carriers.

Clay (1969) noted that habits of directionality varied according to the attained reading level of the child. Better readers usually established accurate line movement and return sweep after seven weeks of instruction while children in average and low reading groups took 15-20 weeks to develop accurate movements. Clay observed, however, that some children took as long as six months to establish correct directional habits. In Clay's (1967) view, though, exposure to written forms should not be withheld because a child is judged "immature" (p. 24). She stated that a correct orientation to print is "fostered by contacts with written language. The visual perception of print, the directional constraints on movement, the special types of sentences used in books, and the synchronized matching of spoken word units with written word units will only be learned in contact with written language" (p. 24).

In one of the more recent and extensive analyses of American children's knowledge of printed conventions, Day, Day, and colleagues (1979, 1980, 1981,) tested children three times during their kinder-

garten year and twice during their first grade year with the *Sand—Concepts About Print Test* (CAPT) which attempts to measure not only knowledge of word boundaries, but also directional habits and knowledge of punctuation. From a previous factor analysis, Day and Day (1979) identified four dimensions of printed concepts which seemed to develop sequentially. By the end of the first grade, Day, Day, and Hollingsworth (1981) recorded that 80% or more of their sample of 51 first graders at the year's end had mastered basic book orientation habits of directionality and were able to identify upper and lower case letters as well as single words in print. However, roughly only a third to a half of the sample were able to recognize incorrect letter and word sequences or noticed when whole lines of print were placed out of order (top and bottom reversed). In addition, while three-quarters of the sample could identify a comma, only 16% could explain the function of quotation marks. However, Day and Day (1979) cautioned in a previous discussion that strong evidence did not emerge supporting the notion that concepts of print are prerequisite to actual ability since some children whose scores were relatively low on the test (16 out of 24) were observed to be reading by teachers during the first grade year.

Johns (1980) administered the CAPT to 60 first graders ending their first year of instruction and found that above average readers performed significantly better than below average ones on items assessing knowledge of letter and word units and on tasks where the child was to recognize incorrect letter and word sequences and explain the function of various punctuation marks. However, Johns pointed out that several items on the test may have not adequately directed the child's attention to the print; therefore, the differences found between types of reader in recognizing inverted letter and word sequences may be less qualitative than an artifact of the examination procedure. Despite these limitations, Johns surmised that "data from this study indicate that above average readers have a greater understanding of print-related concepts than below average readers" (p. 547).

In a further attempt to replicate and extend findings of previous administrations of the CAPT, Yaden (1982) tested 118 first graders in the spring with the most recent edition of the CAPT, *Stones* (Clay, 1979), and obtained a measure of intelligence as well. Using the reading subtests of the *Standard Achievement Test* (SAT) (Madden, Gardner, Rudman, Karlsen, & Merwin, 1972) as measures of reading ability, Yaden found that in a multiple regression with print awareness scores as the criterion, the subtest of word reading was a better predictor of knowledge of printed conventions than that of intelligence as measured by the *Otis-Lennon School Ability Test* (Otis

& Lennon, 1979). Further, all of the reading subtests (word reading, reading comprehension, word study skills, and vocabulary) retained significant partial correlations with print awareness even with intelligence controlled (cf. Francis, 1973).

In general, Yaden's study supported the findings of previous research that some beginning readers' concepts of letters, words and marks of punctuation are not stabilized even after one year of reading instruction. Yaden also discovered that above average readers had better performance on items purporting to measure directional habits with normal and irregular print, and items pertaining to the identification of incorrect letter and word sequences and marks of punctuation. The study did not confirm, however, hypothesized "large" effects of reading achievement and intelligence upon knowledge of printed conventions based upon computation of prior power analyses (cf. Cohen, 1977). Neither reading achievement nor intelligence can be said to contribute substantially to the relationship with print awareness independent of the other. In combination, however, measures of reading achievement and intelligence proved to be useful predictors of knowledge of printed conventions contributing together approximately 40% of the total variance of scores on the CAPT.

In summary, despite discrepancies in the observation of a distinct developmental pattern in the growth of knowledge of printed word boundaries, there is a remarkable unanimity in the findings that beginning readers do not possess firm concepts of printed language units as letters, words, or punctuation marks. Nor do they immediately understand current directional movements. As noted in studies dealing with oral language units, superior readers recognize these linguistic elements in their written form better than poorer readers. This observation plus the finding that reading ability was a better predictor of print awareness than intelligence (cf. Yaden, 1982; Francis, 1973) lends support to an earlier contention by Ehri (1979) that practice with written language is the best way to enhance meta-linguistic growth.

Summary/discussion

A review of the extant research on children's concepts of the functions and processes involved in reading and their awareness of the units of spoken and written language revealed that beginning readers are largely unaware of the overriding structure of the writing system as well as their own speech. They have disparate notions as to what behavior comprises the act of reading and the necessary steps that they must take in getting ready to become a reader. Perhaps the most

disturbing thing as pointed out by some is that there is little or no instructional time spent orienting the children to what reading is or what useful functions it may serve. As Meltzer and Herse (1969) noted, the children are expected to intuitively grasp these conceptual or "metalinguistic" aspects of reading as if the actual learning of the visual symbols itself was entirely self-explanatory of the higher processes. What research has divulged, however, is that merely learning the code does not automatically give children insight into how print may be used nor how these "bunches of letters," as one child put it, work together to represent the variety of intelligible messages ubiquitous in everyday surroundings.

However, there are several issues within the research literature itself which must be addressed and resolved before any definitive conclusions can be drawn as to what applied measures might be taken by educators to improve reading instruction in the area of developing accurate concepts about print. While these have been discussed elsewhere in more depth (Yaden, 1982), space necessitates only a brief mention of three primary concerns here.

Causation Between Metalinguistic Awareness and Reading Ability and the Effects of Instruction

As Day et al. (1981) have pointed out, the absence of a definite causal direction from metalinguistic ability to reading achievement should admit caution when deciding what practical steps ought to be taken in enhancing metalinguistic awareness. From their own study Day et al. found, in a path analysis model, that performance on a measure of print awareness at the beginning of first grade was more highly correlated with reading achievement at the end of the year than was performance on a standardized readiness test. However, this finding has not been universal. Both McNinch (1974) and Evans, Taylor, and Blum (1979) have reported in their investigations that knowledge of printed conventions is a poor predictor of future reading achievement whereas knowledge of spoken word boundaries and ability to "track sound" in words has a greater relationship with end of the year scores in reading.

On the other hand, Ehri (1976) has provided evidence that readers learned context-dependent words such as prepositions and auxiliaries better than pre-readers of the same age, thus lending credence to the view that experience with print heightens awareness to words which otherwise go undetected in normal speech due to their elision with other words. For Ehri, "word segmentation is an inevitable product of the learner's attempts to achieve competence with printed language

and that no special instruction delivered prior to encountering print is required to accomplish this" (p. 841). Since, however, the majority of studies have been correlational, (e.g., Allen, 1983; Taylor & Blum, 1983; Tunmer & Fletcher, 1981) which have indicated a relationship between reading ability and metalinguistic awareness and have involved only statistical manipulations, it has been improper to tease out any definite temporal sequence, verified by experimental procedures, between reading achievement and the development of metalinguistic awareness.

Two studies reviewed, however, have experimentally applied a treatment for enhancing metalinguistic awareness. In the first, Ollila, Johnson, and Downing (1974) found that a Russian training procedure (cf. Elkonin, 1973) for increasing awareness of the phoneme as a "concrete entity" led to better performance on the *Wepman Auditory Discrimination Test* (Wepman, 1958) than did instruction in two basal programs even when the groups were equated on readiness measures. In the second and most recent study, Bradley and Bryant (1983) trained four- and five-year-old non-readers in sound categorization (i.e., distinguishing the "odd" word not sharing a common phoneme in a group of words) and compared them to a control group of children equated in age and intelligence on abilities to read and spell over a 4 year period. Bradley and Bryant found that the treatment group trained previously in sound categorization performed significantly better in standardized tests of reading and spelling than did the controls. In addition, one experimental group given additional exposure to plastic letters along with the training procedure surpassed all groups in the ability to spell. Thus the authors concluded that while former studies implied a relationship between phonological awareness and reading, "our study is the first adequate empirical evidence that the link is causal" (p. 421).

While the Bradley and Bryant study is by far the strongest argument for the view that training in spoken language awareness enhances reading ability, some qualifications must be added. For instance, Bradley and Bryant reported that initial sound categorization scores for the children at 4 and 5 years of age accounted for less than 10% and 5%, respectively, of the variance in reading scores later on. In addition, two years of intensive training in sound categorization produced only a 3-5 month gain for the treatment group in standardized tests of reading and spelling. Thus the small correlation between phonological awareness and reading ability plus the minimal payback of 2 years of instruction mitigates in a pragmatic sense even the discovery of an apparent, proveable, temporal sequence.

Finally, Ehri's (1979) contention that exposure to print itself brings

about awareness of language structure cannot be fully disproven until groups of non-readers are given the kind of intensive immersion in print that has been shown to enhance early reading ability (Clark, 1976; Durkin, 1966; Teale, 1978). Since it has been demonstrated (e.g., Yaden, 1983, 1984; Yaden & McGee, in press) that children as young as two years of age spontaneously asked questions about printed language, including about mapping principles, an experiment could be designed to test the effects of early reading program such as Durkin has suggested (1974-75), against the effects of training in phonological awareness (cf. Bradley & Bryant, 1983). It may be that immersion in a print-rich environment with adults available to answer questions has a greater effect even than training at this age. Final conclusions as to causation must await this type of experimental study.

Differences in research methodologies

Purposes and processes of reading. One reason for discrepancies in the findings of research on measures of metalinguistic awareness and knowledge of print conventions is that as a conceptual framework, metalinguistic knowledge has yet to be fully and adequately described, and therefore methodologies for tapping the related constructs differ widely. For instance, investigators interested in assessing children's knowledge about purposes for and strategies during reading have typically used the structured interview as a data-gathering method (e.g., Denny & Weintraub, 1966; Johns, 1972, 1974; Mason, 1967; Myers & Paris, 1978; Oliver, 1975; Reid, 1966; Tovey, 1976). Since Downing (1970), however, discovered children's ability to point out acts of reading correctly and describe reading processes in more adult-like terms when actually presented with a book or other models of real-life message carriers, doubt is cast on the structured interview as a reliable indicator of children's functional knowledge of the processes involved in reading.

Spoken language units. Similarly, researchers studying children's understanding of spoken language units such as phonemes, words, or sentences have used a variety of tasks as well including the structured interview (Matluck & Mace-Matluck, 1983; Papandropoulou & Sinclair, 1974; Templeton & Spivey, 1980), a range of segmentation tasks (Holden & MacGinitie, 1972; Huttenlocher, 1964; Liberman et al., 1974; Treiman & Baron, 1981) and other performance measures in which children chose linguistic elements from a range of stimuli (e.g., Downing & Oliver, 1974; Horne et al., 1983; Ryan et al., 1977).

Unfortunately, as there was no uniformity in methodology, no uniformity was present in the findings of the aforementioned studies.

While Fox and Routh (1975) stated that 3-year-olds successfully segmented most of the sentences presented into individual words, Downing and Oliver (1973-74) reported for their sample confusions between phonemes, syllables and words even with eight-year-olds. This gross disparity in methodology and findings is disturbing and warrants further investigation toward the development of a well-defined, theoretical framework for metalinguistic abilities as well as some systematic research methodologies for tapping these constructs (cf. Nesdale & Tunmer, 1984)

Written language units. While inquires into children's knowledge of written word boundaries show less variation in experimental tasks, nevertheless, results vary from study to study and certain practical difficulties have arisen. Both Yaden (1982) and Johns (1980) have commented upon the inadequacy of certain items on the CAPT to direct the children's attention to the print, thus confounding findings especially between above and below average readers. Further Ehri (1979) pointed out that the children in Mickish's (1974) study may have not known the words in the test sentence and, therefore, with no spaces between the words, it would be impossible for the children to respond competently. Additionally, Meltzer's and Herse's (1969) original finding of a developmental pattern in the development of space as a boundary for written words has not been corroborated by subsequent replications using their original task or any other. Thus, with print conventions as well, no uniform data collection methods have been found which yield consistently repeatable results.

Tests of Metalinguistic Awareness

Finally, in an effort to substantiate the relationship of metalinguistic awareness to actual reading achievement and to establish some stable, systematic measures of the conceptual nature of reading and awareness of print conventions, a few researchers have developed various formalized tests of linguistic awareness covering a wide array of abilities. While all of the tests measures written conventions including concepts of letter, word and sentence units, marks of punctuation and directional movements, some also include measures of oral language segmentation (e.g., Blum, Evans, & Taylor, 1982) and items to assess knowledge of the functions and types of printed messages found in everyday life (e.g., Downing, Ayers, & Schaefer, 1982). The tests in general explore reading-specific and book-handling behaviors not included in traditional readiness tests and have been often found to correlate more highly with reading achievement measures than either

readiness tests themselves (e.g., Day et al., 1981) or intelligence (e.g., Yaden, 1982). In addition, versions of the LARR (Downing, et al., 1982) have been found valid for usage with non-white populations (e.g., Downing, Ollila, & Oliver, 1975) and to successfully assess the effects of early school instruction and home environment on the acquisition of literacy behaviors (e.g., Downing, Ollila, & Oliver, 1977).

Despite, however, the additional insights into the relationship of metalinguistic awareness and reading achievement gained by the use of these tests, several uncertainties remain to be resolved. Day et al. (1981) have suggested that the high positive correlations between knowledge of print conventions and reading achievement may simply be due to a redundancy across test items. Further, they noted that statistical control of other aspects of linguistic awareness may reveal that knowledge of printed conventions contributes little to actual reading ability. Another difficulty pointed out by Evans et al. (1979) and Johns (1980) is that studies using relatively small sample sizes (e.g., $N \leq 60$), but considering several variables in the analysis generally suffer from a loss of power and generalizability. American investigations using the CAPT (e.g., Day & Day, 1980; Johns, 1980) have generally been of this size, although Yaden (1982) has recently confirmed many of the previous findings with a sample nearly twice as large.

In sum, although the CAPT (Clay, 1972, 1979) has been reviewed favorably (e.g., Goodman, 1981) neither it nor the other instruments have been used widely, particularly in America. While McDonell and Osburn (1978) have reported that the CAPT can be a useful measure of readiness skills in a classroom, few other discussions exist other than by the authors themselves (e.g., Ayers & Downing, 1982; Clay, 1979a; Downing, Ayers, & Schaefer, 1978; Taylor & Blum, 1983) as to the efficacy of these tests in measuring early reading behaviors. More extensive standardization procedures and investigations of applied usages are definitely needed.

Conclusion

In closing, Donaldson (1984) has recently cautioned literacy researchers against the "fashionable" trend of emphasizing "what children *can* [Donaldson's emphasis] do rather than what they cannot do" (p. 174), and ignoring the real differences in the ease of learning to speak as opposed to learning to read and write. She goes on to say that "some things take longer than others to learn and are achieved later or with less universal success. We do no good to children, or to science, by trying to deny it" (p. 174). Thus, that over two decades of research into

metalinguistic abilities, even despite varying methods of data collection and analysis, has turned up overwhelmingly negative results in documenting children's awareness of spoken and written language units and the language of reading instruction speaks strongly toward further investigations into refining existing instruments and the development of experimental research procedures which can yield consistent, replicatable results. It has been repeatedly shown that beginning readers are often unaware of the practical applications of written language in everyday events and have tenuous notions as to the nature, functions, and constraints of the elements constituting written language. That traditional tests of readiness have overlooked these more global aspects of literacy is a fault. To continue to overlook them in the face of mounting evidence for their existence will be inexcusable.

Vygotsky (1978), in discussing the history of written language, has said that "children should be taught written language, not just the writing of letters" (p. 119). The implication here is that to view written language as merely the reproduction of certain isolated, graphic shapes is to miss the importance of the printed code altogether. More recently, Ferreiro and Teberosky (1982) have powerfully reiterated this view.

It has traditionally been thought that to learn to read children must possess good language (or a sufficient level of oral language development) evaluated in terms of vocabulary, diction, and grammatical complexity. If we believe that we must consider language awareness, the perspective changes. Rather than being concerned with whether children know how to speak, we should help them become conscious of what they already know how to do, help them move from "knowing how" to "knowing about," a conceptual knowing. (p. 298)

Thus, while the author of the present paper does not claim to have presented an exhaustive discussion of the literature, it is hoped that enough has been examined that a reevaluation of current reading approaches in terms of ensuring children's "conceptual knowing" might be soon in coming.

The author would like to thank Peter Mosenthal and Merald E. Wrolstad, the general editor, for their many helpful comments and suggestions on earlier drafts of this manuscript. The author takes full responsibility, however, for the final product as it stands and the viewpoints expressed therein.

References

- Allan, K. K. (1982). The development of young children's metalinguistic understanding of the word. *Journal of Educational Research*, 76, 89-93.
- Ayers, D., & Downing, J. (1982). Testing children's concepts of reading. *Educational Research*, 24, 277-283.
- Betts, E. A. (1946). *Foundations of reading instruction*. New York: American Book Company.
- Blum, I. H., Evans, M., & Taylor, N. E. (1982). *BET: Written Language Awareness Test*. Baltimore: University Park Press.
- Blum, I. H., Taylor, N. E., & Blum, R. A. (1979). Methodological considerations and developmental trends in children's awareness of word boundaries. In M. L. Kamil & A. J. Moe (Eds.), *Twenty-eighth Yearbook of the National Reading Conference*, (pp. 33-38). Clemson: National Reading Conference.
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read — a causal connection. *Nature*, 301, 419-421.
- Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453-479). Hillsdale: Lawrence Earlbaum Associates.
- Brown, A. L., & Palincsar, A. S. (1982). Inducing strategic learning from texts by means of informed, self-control training. *Topics in Learning and Learning Disabilities*, 2(1), 1-17.
- Bruce, D. J. (1964). The analysis of word sounds by young children. *British Journal of Educational Psychology*, 34, 158-170.
- Burchfield, R. W. (1976). *A supplement to the Oxford English Dictionary: Vol. 2 (H-N)*. Oxford: Clarendon Press.
- Cherry, C. (1980). *On human communication: A review, a survey, and a criticism* (3rd ed.). Cambridge: MIT Press.
- Clark, M. M. (1976). *Young fluent readers*. London: Heinemann Educational Books.
- Clay, M. M. (1967). The reading behavior of five-year-old children: A research report. *New Zealand Journal of Educational Studies*, 2, 11-31.
- Clay, M. M. (1969). Reading errors and self-correction behavior. *British Journal of Educational Psychology*, 39, 47-56.
- Clay, M. M. (1972). *Sand—Concepts About Print Test*. Auckland: Heinemann Educational Books.
- Clay, M. M. (1979a). *Reading: The patterning of complex behavior* (2nd ed.). Auckland: Heinemann Educational Books.
- Clay, M. M. (1979b). *Stones—Concepts About Print Test*. Auckland: Heinemann Educational Books.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.
- Day, H. D., & Day, K. C. (1979). *Item and factor analysis of the Concepts About Print Test: Patterns in the Sand*. Unpublished manuscript, Texas Woman's University, Denton, TX.

- Day, H. D., & Day, K. C. (1980). *The reliability and validity of the Concepts About Print and Record of Oral Language*. Denton, TX: Texas Woman's University. (ERIC Document Reproduction Service No. ED 179 932)
- Day, H. D., Day, K. C., & Hollingsworth, S. (1981, October). *Gender differences in the relationship between preschool knowledge of print conventions and other cognitive abilities*. Paper presented at the 12th Annual Meeting of the Rocky Mountain Educational Research Association.
- Day, H. D., Day, K. C., Hollingsworth, S., & McClelland, D. K. (1980). Sex differences in orthographic linguistic awareness. *Journal of the Illinois Reading Association*, 8, 21-28.
- Day, K. C., & Day, H. D. (1979). Development of kindergarten children's understanding of concepts about printed and oral language. In M. L. Kamil, & A. H. Moe (Eds.), *Twenty-eighth Yearbook of the National Reading Conference* (pp. 19-22). Clemson: National Reading Conference.
- Day, K. C., & Day, H. D. (in press). Tests of metalinguistic awareness. In Yaden & Templeton (Eds.).
- Day, K. C., Day, H. D., Spicola, R., & Griffen, M. (1981). The development of orthographic linguistic awareness in kindergarten children and the relationship of this awareness to later reading achievement. *Reading Psychology*, 2, 76-87.
- Denny, T. P., & Weintraub, S. (1963). Exploring first graders' concepts of reading. *The Reading Teacher*, 16, 363-365.
- Denny, T. P., & Weintraub, S. (1966). First graders' responses to three questions about reading. *Elementary School Journal*, 66, 441-448.
- de Villiers, J. G., & de Villiers, P. A. (1974). Competence in child judgement: Are children really competent to judge? *Journal of Child Language*, 1, 11-22.
- Donaldson, M. (1978). *Children's minds*. New York: W. W. Norton & Company.
- Donaldson, M. (1984). Speech and writing and modes of learning. In H. Goelman, A. A. Oberg, & F. Smith (Eds.), *Awakening to Literacy* (pp. 174-184). Exeter, NH: Heinemann Educational Books.
- Downing, J. (1969). How children think about reading. *Reading Improvement*, 23, 217-230.
- Downing, J. (1970). Children's concept of language in learning to read. *Educational Research*, 12, 106-112.
- Downing, J. (1971-72). Children's developing concepts of spoken and written language. *Journal of Reading Behavior*, 4, 1-19.
- Downing, J. (1973). A summary of evidence related to the Cognitive Clarity theory of reading. In P. O. Nack (Ed.), *Twenty-second Yearbook of the National Reading Conference: Vol. 1* (pp. 178-184). Boone, NC: National Reading Conference.
- Downing, J. (1976). The reading instruction register. *Language Arts*, 53, 762-766.
- Downing, J. (1979). *Reading and reasoning*. New York: Springer-Verlag.
- Downing, J. (in press). Cognitive clarity: A unifying and cross-cultural theory for language awareness phenomena in reading. In Yaden & Templeton (Eds.).

- Downing, J., Ayers, D., & Schaefer (1978). Conceptual and perceptual factors in learning to read. *Educational Research*, 21, 11-17.
- Downing, J., Ayers, D., & Schaefer, B. (1982). *Linguistic Awareness in Reading Readiness (LARR) Test*. Slough, England: NFER-Nelson Publishing Company.
- Downing, J., & Oliver, P. (1973-74). The child's conception of a word. *Reading Research Quarterly*, 9, 568-582.
- Downing, J., Ollila, L., & Oliver, P. (1975). Cultural differences in children's concepts of reading and writing. *British Journal of Educational Psychology*, 45, 312-316.
- Downing, J., Ollila, L., & Oliver, P. (1977). Concepts of language in children from differing socioeconomic backgrounds. *Journal of Educational Research*, 70, 277-281.
- Downing, J., & Valtin, R. (Eds.). (1984). *Language awareness and learning to read*. New York: Springer-Verlag.
- Dunn, L. M., & Markwardt, Jr., F. C. (1970). *Peabody Individual Achievement Test*. Circle Pines, MN: American Guidance Service.
- Durkin, D. (1966). *Children who read early*. New York: Teachers College Press.
- Ehri, L. C. (1975). Word consciousness in readers and prereaders. *Journal of Educational Psychology*, 67, 204-212.
- Ehri, L. C. (1976). Word learning in beginning readers and prereaders: Effects of form class and defining contexts. *Journal of Educational Psychology*, 68, 832-842.
- Elkonin, D. B. (1973). Methods of teaching reading. In J. Downing (Ed.), *Comparative Reading* (pp. 551-578). New York: MacMillan.
- Evanechko, P., Ollila, J., Downing, J., & Braun, C. (1973). Investigation of the reading readiness domain. *Research in Teaching of English*, 7, 61-78.
- Evans, M. (1975). Children's ability to segment sentences into individual words. In G. McNinch & W. D. Miller (Eds.), *Twenty-fourth Yearbook of the National Reading Conference* (pp. 177-180). Clemson, SC: National Reading Conference.
- Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. Exeter: NH: Heinemann Educational Books.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-236). Hillsdale: Lawrence Earlbaum Associates.
- Francis, H. (1973). Children's experience of reading and notions of units in language. *British Journal of Educational Psychology*, 43, 17-23.
- Gates, A. I., & MacGinitie, W. H. (1965). *Gates-MacGinitie Reading Tests: Primary I*. New York: Teacher College Press.
- Gleitman, H., & Gleitman, L. (1979). Language use and language judgement. In C. J. Fillmore, D. Kempler, & W. S-Y Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 103-126). New York: Academic Press.
- Gleitman, L. R., Gleitman, H., & Shipley, E. F. (1972). The emergence of the child as grammarian. *Cognition*, 1, 137-163.

- Goodman, Y. M. (1981). Test review: Concepts About Print Test. *The Reading Teacher*, 34, 445-448.
- Hakes, D. T., Evans, J. S., & Tunmer, W. E. (1980). *The development of metalinguistic abilities in children*. New York: Springer-Verlag.
- Henderson, E. H. (1981). *Learning to read and spell: The child's knowledge of words*. DeKalb: Northern Illinois University Press.
- Henderson, E. H., & Beers, J. W. (Eds.). (1980). *Developmental and cognitive aspects of learning to spell*. Newark, DE: International Reading Association.
- Holden, M. H. (1972). Metalinguistic performance and cognitive development in children five to seven. *Dissertation Abstracts International*, 33, 2791B-2792B. (University Microfilms No. 72-31, 214)
- Holden, M. H. (1977). Word awareness, reading and development, *Perceptual and Motor Skills*, 44, 203-206.
- Holden, M. H., & MacGinitie, W. H. (1972). Children's conceptions of word boundaries in speech and print. *Journal of Educational Psychology*, 63, 551-557.
- Holden, M. H., & MacGinitie, W. H. (1973). *Metalinguistic ability and cognitive performance in children from five to seven*. Paper presented at the meeting of the American Education Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 078 436)
- Horne, M. D., Powers, J. E., & Mahabub, P. (1983). Reader and non-reader's conception of the spoken word. *Contemporary Educational Psychology*, 8, 403-418.
- Huttenlocher, J. (1964). Children's language: Word-phrase relationship. *Science*, 143, 264-265.
- Johns, J. (1972). Children's concepts of reading and their reading achievements. *Journal of Reading Behavior*, 4, 56-57.
- Johns, J. (1974). Concepts of reading among good and poor readers. *Education*, 95, 58-60.
- Johns, J. (1977). Children's conceptions of a spoken word: A developmental study. *Reading World*, 16, 248-257.
- Johns, J. (1980). First graders' concepts about print. *Reading Research Quarterly*, 15, 529-549.
- Johns, J. L., & Ellis, D. W. (1976). Reading: Children tell it like it is. *Reading World*, 16, 115-128.
- Karpova, S. N. (1966). The preschooler's realization of the lexical structure of speech. In F. Smith, & G. A. Miller (Eds.), *The Genesis of language: A psycholinguistic approach* (pp. 370-371). Cambridge: MIT Press. (Original work published 1955)
- Kingston, A. J., Weaver, W. W., & Figa, L. E. (1972). Experiments in children's perceptions of word and word boundaries. In F. P. Green (Ed.), *Twenty-first Yearbook of the National Reading Conference: Vol. 2* (pp. 91-99). Milwaukee: National Reading Conference.
- Leong, C. K., & Haines, C. F. (1978). Beginning readers' analysis of words and sentences. *Journal of Reading Behavior*, 10, 393-407.

- Liberman, I. Y. (1973). Segmentation of the spoken word. *Bulletin of the Orton Society*, 23, 65-77.
- Liberman, I. Y., Shankweiler, D., Fisher, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 18, 201-212.
- Lindamood, C. H., & Lindamood, P. C. (1971). *Lindamood Auditory Conceptualization Test*. Boston: Teaching Resources.
- Lundberg, I., & Torneus, M. (1978). Nonreaders' awareness of the basic relationship between spoken and written words. *Journal of Experimental Child Psychology*, 25, 404-412.
- Madden, R., Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. (1972). *Standard Achievement Test: Primary Level I*. New York: Harcourt Brace Jovanovich.
- Markman, E. M. (1976). Children's difficulty with word referent differentiation. *Child Development*, 47, 742-749.
- Mason, G. E. (1967). Preschoolers' concepts of reading. *The Reading Teacher*, 21, 130-132.
- Mattingly, I. G., (1972). Reading, the linguistic process and linguistic awareness. In J. F. Kavanaugh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading* (pp. 133-148). Cambridge: MIT Press.
- Mattingly, I. G. (1979). The psycholinguistic basis for linguistic awareness. In M. L. Kamil, & A. J. Moe (Eds.), *Twenty-eighth Yearbook of the National Reading Conference* (pp. 274-278). Clemson: National Reading Conference.
- Matluck, J. H., & Mace-Matluck, B. J. (1983, May). Metalinguistic skills and reading achievement. In T. D. Horn (Chair), *The development of school-related language and reading skills by Spanish-English bilingual children*. Symposium conducted at the Convention of the International Reading Association, Anaheim, CA.
- Mayfield, M. I. (1983). Code systems instruction and kindergarten children's perceptions of the nature and purpose of reading. *Journal of Educational Research*, 76, 161-168.
- McConkie, G. W. (1959). *The perceptions of a selected group of kindergarten children concerning reading*. Unpublished doctoral dissertation, Teacher's College, Columbia University.
- McDonnell, G. M., & Osburn, E. B. (1978). New thoughts about reading readiness. *Language Arts*, 55, 26-29.
- McNinch, G. (1974). Awareness of aural and visual word boundary within a sample of first graders. *Perceptual and Motor Skills*, 38, 1127-1134.
- Meltzer, N. S., & Herse, R. (1969). The boundaries of written words as seen by first graders. *Journal of Reading Behavior*, 1, 3-13.
- Mickish, V. (1974). Children's perceptions of written word boundaries. *Journal of Reading Behavior*, 6, 19-22.
- Myers, M., & Paris, S. G. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology*, 70, 680-690.

- Nesdale, A. R., & Tunmer, W. E. (1984). The development of metalinguistic awareness: A methodological overview. In Tunmer, Pratt, & Herriman (Eds.).
- Oliver, M. E. (1975). The development of language concepts of preprimary Indian children. *Language Arts*, 52, 865-869.
- Ollila, L., Johnson, T., & Downing, J. (1974). Adapting Russian methods for auditory discrimination training in English. *Elementary English*, 51, 1134-1141, 1145.
- Otis, A. S., & Lennon, R. T. (1979). *Otis-Lennon School Ability Test: Primary I*. New York: Harcourt Brace Jovanovich.
- Papandropoulou, I., & Sinclair, H. (1974). What is a word? Experimental study of children's ideas on grammar. *Human Development*, 17, 241-258.
- Pei, M. A., & Gaynor, F. (1954). *A dictionary of linguistics*. New York: Philosophical Library.
- Prescott, G. A., Balow, I. H., Hogan, T. P., & Farr, R. C. (1971). *Metropolitan Achievement Tests*. New York: Harcourt Brace Jovanovich.
- Reid, J. F. (1966). Learning to think about reading. *Educational Research*, 9, 56-62.
- Rozin, P., Bressman, B., & Taft, M. (1974). Do children understand the basic relationship between speech and writing? The mow-motorcycle test. *Journal of Reading Behavior*, 6, 327-334.
- Ryan, E. B., McNamara, S. R., & Kenney, M. (1977). Linguistic awareness and reading performance among beginning readers. *Journal of Reading Behavior*, 9, 399-400.
- Ryan, E. B. (1980). Metalinguistic development and reading. In F. B. Murray (Ed.), *Language awareness and reading* (pp. 38-59). Newark, DE: International Reading Association.
- Sinclair, A., Jarvella, R. J., Levelt, W. J. M. (Eds.). (1978). *The child's conception of language*. Berlin: Springer-Verlag.
- Sanders, T. S. (1981). Three first graders' concept of word and concepts about the language of literacy instruction. In M. L. Kamil & A. J. Moe (Eds.), *Thirtieth Yearbook of the National Reading Conference* (pp. 266-272). Clemson: National Reading Conference.
- Sulzby, E. (1978). Children's explanations of word similarities in relation to word knowness. In P. D. Pearson, & J. Hansen (Eds.), *Twenty-seventh Yearbook of the National Reading Conference* (pp. 51-55). Clemson: National Reading Conference.
- Sulzby, E. (1979). Semantic salience in relations to word knowness. In M. L. Kamil, & A. J. Moe (Eds.) *Twenty-eighth Yearbook of the National Reading Conference* (pp. 49-54). Clemson: National Reading Conference.
- Sulzby, E. (1981, December). *Kindergartners deal with word boundaries*. Paper presented at the annual meeting of the National Reading Conference, Dallas.
- Taylor, N. E., & Blum, I. H. (1983). *The effects of written language awareness on first grade reading achievement*. Unpublished manuscript, Catholic University of America, Washington, D.C.

- Taylor, N. E., Blum, I. H., Logsdon, D. M., & Moeller, G. D. (1982). *The development of written language awareness: Environmental aspects and program characteristics*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Teale, W. H. (1978). Positive environments for learning to read: What studies of early readers tell us. *Language Arts*, 55, 922-932.
- Templeton, W. S., & Spivey, E. M. (1980). The concept of word in young children as a function of level of cognitive development. *Research in the Teaching of English*, 14, 265-278.
- Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation of reading ability. In G. E. McKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice: Vol. 3* (pp. 159-198). New York: Academic Press.
- Tunmer, W. E., & Bowey, J. A. (1984). Metalinguistic awareness and reading acquisition. In Tunmer, Pratt, & Herriman (Eds.).
- Tunmer, W. E., Bowey, J. A., & Grieve, R. (1983). The development of young children's awareness of the word as a unit of spoken language. *Journal of Psycholinguistic Research*, 12, 567-594.
- Tunmer, W. E., & Fletcher, C. M. (1981). The relationship between conceptual tempo, phonological awareness and word recognition in beginning readers. *Journal of Reading Behavior*, 13, 173-185.
- Tunmer, W. E., Pratt, C., & Herriman, M. L. (Eds.). (1984). *Metalinguistic awareness in children*. New York: Springer-Verlag.
- Valtin, R. (1984). Awareness of features and functions of language. In J. Downing, & R. Valtin (Eds.).
- Vygotsky, L. S. (1978). *Mind in Society* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge: Harvard University Press.
- Wagoner, S. A. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly*, 18, 328-346.
- Wepman, J. P. (1958). *Wepman Auditory Discrimination Test*. Chicago: Language Research Associates.
- Yaden, D. B. (1982). A multivariate analysis of first graders' print awareness as related to reading achievement, intelligence and gender (Doctoral dissertation, University of Oklahoma, 1982). *Dissertation Abstracts International*, 43, 1912A.
- Yaden, D. B. (1983). *A categorization of two children's questions about print as they learn to read: A case study*. Paper presented at the annual meeting of the Oklahoma Reading Council, Lawton. (ERIC Document Reproduction Service No. ED 227 472)
- Yaden, D. B. (1984). *Developing a sense of story: Why children reread their favorite books*. Paper presented at the National Council of Teachers of English Annual Spring Conference, Columbus, OH.
- Yaden, D. B., & McGee, L. M. (in press). *Reading as a meaning-seeking activity: What children's questions reveal*. National Reading Conference Yearbook.
- Yaden, D. B., & Templeton, W. S. (Eds.). (in press). *Metalinguistic awareness and beginning literacy: Conceptualizing what it means to read and write*. Exeter, NH: Heinemann Educational Books.

Before Numerals

Denise Schmandt-Besserat

The paper deals with the development of counting devices in the ancient Middle East between 10,000-3000 B.C. and, in particular, bone tallies, clay tokens, and numerical notations on clay tablets. These technologies handled plurality in increasingly abstract terms. Data is tested against a model for the development of abstract numbers proposed by the historian of mathematics Tobias Danzig.

Introduction

In previous years I have studied the role played by prehistoric counters in the origin of writing.¹ I am presently studying the role played by the same counters in the origin of counting and, in particular, in the origin of abstract numbers.²

I first define some terms used in the discussion. *Numerals* are symbols to represent abstract numbers. Abstract means removed from the concrete reality. *Abstract counting* refers to using number concepts abstracted from any particular concrete entity. Our numbers 1, 2, 3, etc. . . . are expressing the concepts of oneness, twoness, threeness as abstract entities divorced from any particular collection. As a result 1, 2, 3 . . . are universally applicable. Concrete counting, on the other hand, does not abstract numbers from the things counted. As a result, in *concrete counting* the number words that express the concepts "one," "two," "three," etc., differ according to whether, for instance, men, canoes, or trees are being counted. These different sets of number words, which change according to the category of item counted, are called *concrete numbers*. Such examples as twins, triplets, and quadruplets to count children of a same birth is the closest analogy to concrete numbers in our own society. It is well understood, however, that in our society such special numerical terms which refer to particular groups are not really used for counting whereas concrete numbers were. *Counting in one-to-one correspondence* consists of matching the items to be counted by an identical number of counters. For instance, matching each sheep of a flock with a pebble. This method of counting does not require any concept of numbers.

The Hypothesis

Danzig,³ Smith,⁴ Kramer,⁵ and Flegg,⁶ to name only a few, are among the historians of mathematics who have postulated that there were three main steps in the evolution of counting: (1) one-to-one correspondence, (2) concrete counting, and (3) abstract counting.

1 *One-to-one correspondence*

The historians of mathematics quoted above hypothesize that, ages ago, counting consisted only in the repeated addition of one unit with no idea of cumulative amounts. Tribes such as the Vedda of Ceylon never reached much beyond this level in historical times. They counted coconuts, for instance, by matching each coconut with a stick. For each stick added they counted "and one more" until the collection of coconuts was exhausted. Then, they merely pointed to the resulting pile of sticks saying "that many."⁷ At this stage, in other words, people lacked concepts for numbers. Collections were conceived, therefore, as series of individual disconnected entities rather than as coherent wholes.

2 *Special numerations*

At this second stage the notion of sets is suggested to have been acquired. It would have fused, however, the concepts of number and of the objects counted. As a result, different things would have been counted with different numerical expressions or concrete numbers. This is inferred from languages where the words for numbers change according to the things counted. Menninger cites as an example the Fiji Islanders who call ten boats "bola" and ten coconuts "boro."⁸ One of the most quoted example of concrete counting is that the British Columbia tribes studied by Franz Boas.⁹ The concrete numbers they used to count men, canoes, long objects, flat objects, round objects or time; measures or other items are illustrated in Table I. Diakonoff recently published an article dealing with concrete numbers in which he gives the example of Gilyak, a language spoken on the River Amur, which had no fewer than twenty-four classes of numbers. For instance, the word used to express the number "2" was *mex* when referring to spears and oars; *mik* for arrows, bullets, berries, teeth, fists; *meqr* for islands, mountains, houses, pillows; *merax* for eyes, hands, buckets, footprints; *min* for boots; *met* for boards and planks; *mir* for sledges, etc. . . .¹⁰ There are numerous remnants of such usage among Paleo-European,¹¹ Paleo-Asiatic, Micronesian, and Afrasian languages¹² suggesting that a stage of special numerations for individual classes of

Number	Men	Canoes	Long Objects	Flat Objects	Round Objects	Measures	Counting
1	k'al	k'amaet	k'awutskan	gak	g'erel	k'al	gyak
2	t'epqadal	g'alpeeltk	gaopskan	t'epqat	goupel	gulbel	t'epqat
3	gulal	galtskantk	galtskan	guant	gutle	guleont	guant
4	tqalpqdal	tqalpqsk	tqaapskan	tqalpq	tqalpq	tqalpqalont	tqalpq
5	kcenecal	kctoonsk	k'etoentskan	kctonc	kctonc	kctonsilont	kctonc
6	k'aldal	k'altk	k'aoltskan	k'alt	k'alt	k'aldelont	k'alt
7	t'epqaldal	t'epqaltk	t'epqaltskan	t'epqalt	t'epqalt	t'epqaldelont	t'epqalt
8	yuktleadal	yuktaltk	ek'tlaedskan	yuktalt	yuktalt	yuktaldelont	guandalt
9	kctemacal	kctemack	kctemaetskan	kctemac	kctemac	kctemasilont	kctemac
10	kpal	gy'apsk	kpeetskan	gy'ap	kpeel	kpeont	gy'ap

Table I. The Tsimshians of British Columbia used these various number words according to whether they were counting men, canoes, long objects, flat objects, round objects or time; measures and any other item. The use of different numeration systems to count different items is called "concrete counting."

items may have preceded the acquisition of abstract numbers in several parts of the world. In the case of Gilyak where the different sets of numerals seem not totally unrelated, but constitute perhaps only modifications of the same root forms, it could be argued that the language inflected the numerical expressions according to the semantic categories they modified. The use of numerical classifiers in Japanese,¹³ in Aztec and Maya languages¹⁴ can be viewed, probably, also as relics of such concrete counting practice.

Certain English numerical expressions to express "two" and "many" are comparable to concrete numbers, for example, "a couple of days," "twins," "a brace of pheasants," "a pair of shoes," "a school of fish," "a flock of sheep," "a herd of cows," and "a pride of lions." These different words to express quantities in specific situations may suggest that in our own society there was a time when not only concrete counting was common¹⁵ but when counting was restricted to concrete numerations limited to "one," "two," and "many."

What is our present knowledge on the evolution of counting in the ancient Middle East and, in particular, is there any evidence for the use of concrete counting? Diakonoff postulates that there is. The Soviet sumerologist and linguist argues that the many different numerical signs to express quantities, capacity, area measures, etc. . . . point toward an ancient tradition of concrete counting in proto- or pre-historic Mesopotamia.¹⁶ As will be discussed below, the archaeological material supports Diakonoff's hypothesis.

3 *Abstract numbers*

At this third and final stage, the concepts of numbers would have been abstracted from the items counted, giving rise to abstract numbers which could be applied universally, like our own concepts of one, two, three, etc. Smith remarked that in a number of societies, the words to express abstract numbers derived from a concrete numeration of particularly frequent use. He cites, for instance, the Niues of the Southern Pacific who counted with abstract numbers that meant literally "one fruit, two fruits, three fruits," whereas in other cases the words corresponding to our "one, two, three" were expressed by such words as "one grain, two grains, three grains," or "one stone, two stones, three stones."¹⁷

In sum, according to the hypothesis presented above, counting would have evolved in spurts followed by plateaus over an exceedingly long time. As Russell wrote, "It must have required many ages to discover that a brace of pheasants and a couple of days were both instances of the number 2."¹⁸ The study on cognitive development in children

carried out by Piaget recognizes comparable stages leading to the mastery of counting.¹⁹ According to Piaget's analysis, children start by matching collections in a one-to-one correspondence at an early age, but associate quantities with numbers relatively late.²⁰ At present, no one can explain how this evolution takes place in the child except by a process of maturation.²¹ We know equally little about the mind of early humans and, at present, do not understand the cognitive processes that led to the development of counting and its timing. It is to this issue that ancient reckoning devices and their proper interpretation may contribute new insights.

The evolution of reckoning technologies in the Ancient Middle East

Here we deal with artifacts found in excavations in the Middle East which have been identified as counting devices. These objects include tally sticks, tokens, and notations on clay envelopes and tablets. I analyze the way each device may have handled plurality and suggest that the archaeological data substantiates the hypothesis presented above in each of its three successive steps, as follows:

1 Tallied bones used to count in a one-to-one correspondence

Animal bones and antlers bearing series of notches found in Mesolithic sites of the Middle East about 10,000 B.C.^{22,23} are the earliest artifacts interpreted by scholars as reckoning devices.²⁴ It is not surprising to find tallied bones as the earliest evidence for the art of counting in the ancient Middle East because notched sticks are among the most primitive reckoning devices that are attested from all parts of the world.

Whatever the Mesolithic notations represented, they seem to have functioned in a one-to-one correspondence. The markings appear to be case specific. That is to say, the same kind of notch would have stood, according to the occasion, either for a bison or a reindeer. Only the person, or persons, keeping tally could have known, therefore, what was being recorded.

Such notations would involve abstraction in the sense that one concrete object seems to be represented by one abstract notation. This would have had the effect of bringing together for scrutiny the repeated occurrences of the objects counted; however, there is nothing in the tallies that indicates any notion of sets. The notches are arranged in series of units which are apparently never articulated into quantified collections. The tallies seem to illustrate, therefore, the first level of counting, in a one-to-one correspondence.

2 *Tokens used for concrete counting*

I have presented evidence pointing to the use of tokens for counting between 8000-3100 B.C.²⁵ This token system seems to reflect the conceptual level at which only units of the same kind could be counted together.

The singularity of the token system lies in the multiplicity of shapes of the counters. Whereas the mesolithic tallies had employed a series of identical notches incised on a bone, tokens were now modeled in clay into specific, systematically repeatable shapes, easy to identify. While the series of notches could only be understood by those who had initiated them, a group of tokens could be identified at all times with units of a specific product. In other publications, I have argued that each shape stood for a unit of precise commodity.²⁶ For instance, a sphere seems to have equaled a large measure of grain and a cone a small measure of grain, whereas an ovoid probably represented a jar of oil (Figures 1 and 2).

I would like to emphasize here my assumption that the units of products expressed by the tokens should be understood as traditional containers in which the goods were dealt with in daily life. They would correspond to such measures of common usage such as "a pitcher of beer," "a carafe of wine," and "a mug of coffee." These units, in other words, should be considered as only casually standardized and entirely non-mathematical entities. Grain, for example, might have been handled in baskets of various usual sizes in which case the cone could stand for "a basket of grain" and the sphere for "a large basket of grain." The "basket" and the "large basket" would be used in different circumstances requiring different quantities of grain but the "large basket" would be in no way considered as a direct multiple of the "basket."

I have also argued that the tokens were used in a one-to-one correspondence. In other words, one jar of oil would have been represented by one ovoid, two jars by two ovoids, and so on.²⁷

In spite of this one-to-one correspondence that characterizes the token system, it entails certain elements of abstraction. First of all, the units of real goods, such as quantities of grain and oil, are replaced by clay symbols, and this itself is the result of a process of abstraction. Secondly, the tokens abstracted the data from its context, thus allowing the accounts to abstractly manipulate goods. For example, the Sumerian accountant dealing with the administration of temple flocks using tokens did not need to take into account the actual whereabouts of the sheep involved.

On the other hand, the tokens remained concrete in several ways:

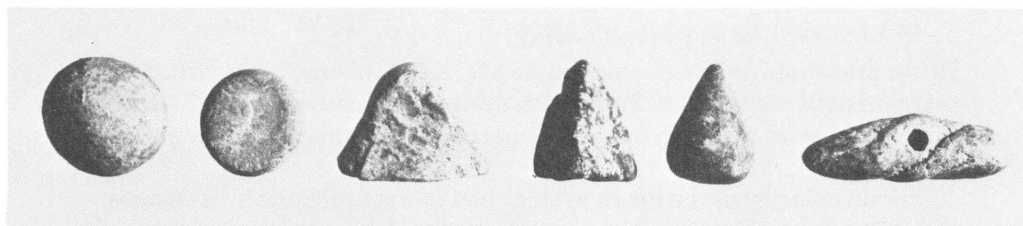
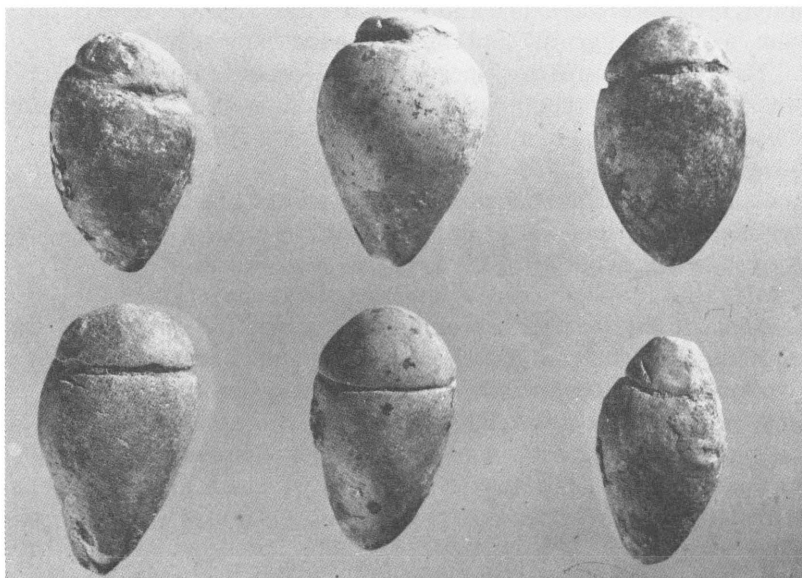


Figure 1. Tokens are small artifacts made of clay and shaped into various specific forms such as the sphere, disk, tetrahedrons, cone, and cylinder illustrated here. The tokens served as counters to keep track of goods in the ancient Middle East when writing had not yet been invented. Each token was a symbol representing one unit of a particular commodity. The cones, spheres, and disks, for example, probably stood for different measures of grain, whereas the tetrahedrons may have represented units of service and the twisted cylinder a bundle of rope.

Figure 2. Counting with tokens was performed in a one-to-one correspondence. Six jars of oil, for example, were represented by six ovoid tokens, as illustrated here, each ovoid standing for "one jar of oil." Each token merged together in a same symbol, therefore, the concept of the item counted and the concept of the number "one." This method of counting is known as "concrete counting." It is fundamentally different from "abstract counting" which expresses the concepts of oneness, twoness, threeness independently from the items counted. There were no abstract numerals such as 1, 2, 3, etc. . . . in the token system.



1 The counters were three-dimensional, tangible, and could be manipulated with the hand.

2 The token system fused together the notion of quality and two concepts of quantity. The ovoid which probably stands for “one jar of oil” merges together, for example, the concepts of “oil”; of quantity (how much) “jar”; and of number (how many) “one.” This way of translating the visual image of the item is somewhat analogous to our concepts of “a keg of beer” and “a bottle of Chianti.”

3 Most importantly, the tokens represented plurality as it is in nature: in one-to-one correspondence. Three jars of oil were expressed by one ovoid + one ovoid + one ovoid which translated what three jars of oil are in reality: one jar of oil + one jar of oil + one jar of oil.

In sum, like the numerical expressions of concrete counting, the tokens did not abstract numbers. Each token merged the notions of (a) nature/quantity of a product; (b) the number *one*. This is why, like number words used in concrete counting, each token shape was specific to one item counted. Ovoids could only count jars of oil and jars of oil could only be counted with ovoids. Likewise, the cones could only count small measures of grain and large measures of grain could only be counted by spheres. Should we imagine what counting device would suit concrete counting, we would have to think in terms of a system, like that of the tokens, with different counters for expressing the different concrete numerations. It is conceivable, therefore, that the token system could reflect or derive from the practice of concrete counting. Like we count “one, two, three . . .” with the help of the beads of an abacus, the various types of tokens would have suggested the appropriate numeration to be used.

3 *From concrete to abstract counting*

When tokens were replaced by their image impressed on the surface of a clay tablet, a sphere, for instance, was replaced by a circular impressed mark and a cone by a conical impressed mark (Figure 3). These impressed marks — ideograms — could no longer be grasped in the hand and manipulated, like the tokens had been. In this sense, the impressed ideograms were removed one step further from the actual real good they represented. Semantically, however, the impressed marks were identical to the tokens. Each ideogram still fused together the concepts of (1) nature/quantity (i.e., measure of grain) and (2) the number *one*. At this stage plurality was still expressed in a one-to-one correspondence. For example, two measures of grain were shown by *two* conical marks. The impressed tablets, therefore, do not reflect any change in the counting practice.

At the next stage in the evolution of writing, which is characterized by the technique of tracing ideograms with the *sharp* end of a stylus rather than impressing them with the *blunt* end of a stylus, plurality is no longer expressed by one-to-one correspondence. The incised pictographs representing units of goods, such as “jar of oil,” are no longer repeated according to the number of units in question. Three jars of oil, for example, are never indicated by repeating the unit “jar of oil” three times. Instead, the sign “jar of oil” is preceded by numerals — symbols expressing an abstract number. Notations expressing abstract numbers are first present, therefore, on the pictographic tablets of Uruk IVa, ca. 3100 B.C. This does not say that 3100 B.C. is the time when abstract numbers were conceived. It says that 3100 B.C. is the time when we observe a change in the record keeping technique.

Figure 3. The tablet displays two kinds of information. First, the all-over pattern showing various kinds of jars is the impression of a seal which identified an office or an individual. Second, the circular and conical marks are notations impressed with a stylus. These marks replaced the tokens after 3200 B.C. They still perpetuated the form of the tokens and stood for the same units of goods. For example, the circular and conical marks shown on this tablet stood for measures of grain. The marks, like the tokens, were used in a one-to-one correspondence. There were still no numerals to express abstract numbers.



It appears that the conical and circular impressed marks — which continued to be impressed while the other signs were being traced with the sharp end of a stylus — could be read, according to the context, either as units of grain or “one” and “six.”²⁸ For example, on the tablet shown in Figure 4, the three circular marks and the three conical marks associated with the sign “jar of [oil/beer?]” are probably to be read as “18 jars of [oil/beer?].” It was a leap from the original concrete reading: a measure of grain, to a secondary abstract meaning: an abstract number. The choice of metrological units of grain for a more general use appears logical, first, because grain was the commodity most widely exchanged in the ancient Middle East. It played the role of currency and must have been, therefore, the most familiar accounting system. Second, the grain accounting system provided a

Figure 4. With the introduction of *incised* pictographs — rather than *impressed* signs — plurality is no longer represented by a one-to-one correspondence. The incised pictographs representing units of goods such as “a jar of oil” are no longer repeated according to the number of units in question. Instead, the sign “jar of oil” is preceded by numerals — symbols expressing an abstract number. Each conical mark is to be read as the numeral “1” and each circular mark as the numeral “6.” The tablet thus records a total of “21 jars of oil.” The cone and sphere which represented the most basic units of grain led, ultimately, therefore, to the development of numerals standing for the abstract numbers “1” and “6” in the Sumerian numerical system.



unique gamut of units which could be easily converted into a sequence of numerical units of growing magnitude. Further studies will be necessary to show how the various measures of grain became standardized to become multiples of one another, leading to such equation as 6 units (ban) equal 1 large unit (bariga). Thus, it appears that the cone and the sphere which probably represented the most basic quantities of grain handled in daily life ultimately led to the development of numerals standing for 1 and 6 in the Sumerian numerical system.

The system of notation was not fully abstract, however, and small numbers were still indicated by 2, 3, 4, 5 impressed conical marks. A new element of abstraction in notation was the use of the impressed circular mark for the number six. This created an economy of notation since eighteen could be represented as three circular and three conical marks.²⁹

Conclusion

The archaeological evidence suggests an evolution from concrete to abstract counting in the ancient Middle East hence supporting the linguistic evidence. As is typical in concrete counting, the notions of the nature of the commodity and quantity (how many) were inseparable in the tokens used for counting between 8000-3100 B.C. Writing, which appears about 3100 B.C. first provided two parallel systems of notations which split the notions of quality and quantity (how many). The first system of notations were numerals (impressed marks) expressing abstract numbers and the second (incised ideograms) expressed the things counted. The new technology for record keeping appears to reflect, therefore, a radically new method of data processing with the use of abstract numerals. This is also supported by an abrupt reduction in the number of shapes of tokens about 3100 B.C.³⁰ It is assumed that the few remaining shapes, namely plain spheres and disks, were henceforth used as counters to calculate numerical amounts. The tokens would have no longer expressed concrete numbers.

Acknowledgements

I owe very special thanks to Marvin A. Powell who edited the paper and helped me correct and improve the original manuscript presented to *Visible Language*. I am most grateful to Carol F. Justus for reading and discussing with me the many drafts of this paper.

Figures 1, 2, 3, featuring respectively tokens from Susa, from Tello, and an impressed tablet from Susa has been provided by the Musée du Louvre, Département des Antiquités Orientales. Figure 4, showing a tablet from Godin Tepe (Gd.73.295) was obtained by courtesy of T. Cuyler Young, Jr., Royal Ontario Museum, Toronto. I want to express my appreciation to William W. Hallo who is currently engaged in a full study of the tablets of Godin Tepe for his permission to use this tablet. Table I illustrates the seven numerations used by the Tsimshians of British Columbia according to L. L. Conant, *The Number Concept*, MacMillan, New York, 1896.

1. Denise Schmandt-Besserat, "How writing came about," *Zeitschrift für Papyrologie und Epigraphik*, 47, 1982, 1-5.
2. Denise Schmandt-Besserat, "Tokens and counting," *Biblical Archaeologist*, 46, 1983, 117-120.
3. Tobias Danzig, *Number: the Language of Science* (fourth edition), MacMillan Company, New York, 1959, p. 6.
4. David E. Smith, *History of Mathematics*, vol. 1, Ginn and Company, Boston, 1951, p. 6-8.
5. Edna E. Kramer, *The Nature and Growth of Modern Mathematics*, Hawthorn Books, New York, 1970, p. 4-5.
6. Graham Flegg, *Numbers, Their History and Meaning*, Schocken Books, New York, 1983, p. 8-14.
7. Karl Menninger, *Number Words and Number Symbols* (revised edition), M.I.T. Press, Cambridge, 1977, p. 33.
8. Menninger, p. 11.
9. Franz Boas, "Fifth report on the northwestern tribes of Canada," *Proceedings of the British Association for the Advancement of Science*, 1889, p. 881.
10. I.M. Diakonoff, "Some reflections on numerals in Sumerian towards a history of mathematical speculations," *Journal of the American Oriental Society*, 103: 1, 1983, 88.
11. Menninger, p. 11 and 21.
12. Diakonoff, p. 87, 88, 93.
13. Mary Sanches, "Language acquisition and language change: Japanese numeral classifiers," in Ben Blount and Mary Sanches, eds., *Sociocultural Dimensions of Language Change*, Academic Press, New York, 1977, p. 51.
14. Brent Berlin, *Tzeltal Numeral Classifiers*, Mouton, Paris, 1968, p. 20.
15. David E. Smith, *Number and Numerals*, Bureau of Publications, Teachers College, Columbia University, New York, 1937, p. 2.
16. Diakonoff, p. 88.

17. Smith, p. 8.
18. Bertrand Russell, *Introduction to Mathematical Philosophy* (tenth impression), George Allen and Unwin, London, 1960, p. 3.
19. Jean Piaget, *The Child's Conception of Number* (fourth impression), Routledge and Kegan Paul, London, 1969.
20. Howard E. Gruber and J. Jacques Voneche, *The Essential Piaget*, Basic Books, New York, 1977, p. 298.
21. Jean Piaget and Barbel Inhelder, *La Genèse des Structures Logiques Élémentaires*, Delachaux et Niestle, Neuchatel, 1959, p. 13.
22. O. Bar-Yosef and N. Goren, "Natufians remains in Hayonim cave," *Paleorient*, 1, 1973, fig. 8: 16-17.
23. Jean Perrot, "Le Gisement Natoufien de Mallaha (Eynan), Israel," *L'Anthropologie*, 70: 5-6, 1966, fig. 22:26.
24. Alexander Marshack, *The Roots of Civilization*, McGraw-Hill, New York, 1972.
25. Schmandt-Besserat (1983).
26. Denise Schmandt-Besserat, "The envelopes that bear the first writing," *Technology and Culture*, 21: 3, 1980, 371-375. The analysis of the groups of tokens found enclosed in envelopes supports the fact that until the last quarter of the fourth millennium B.C. tokens always represented units of goods — never numbers. As I have explained elsewhere, the tokens could not stand for numbers, because, if they had, the messages, or record, carried by the envelopes were useless. For instance, if the spheres stood for ten, as some have erroneously suggested, an envelope yielding three spheres would read: "18 of an unspecified product" or $18x$. Such imprecision is fully inconsistent with the idea of the token system which was obviously developed as a conscious effort to differentiate between the goods counted. It is fully inconsistent also with the bookkeeping practice in the Sumerian bureaucracy. From the lists of goods featured on the tablets, we know that the accountants entered each product with painstaking precision, differentiating, for instance, between multiple kinds of breads, beers, or various breeds of sheep. The reading for three spheres as "three units of grain" is therefore logical. It is also impossible that some tokens represented goods and other numbers, as I had suggested earlier. The case of an envelope from Uruk makes it obvious. It contained, among others, 3 units of oil, 9 units of service, and 24 spheres. The only intelligible reading of these tokens is: 3 jars of oil, 9 units of service, and 24 bushels of grain. Otherwise it is impossible to know how many sixs referred to each item.
27. Schmandt-Besserat, (1980), p. 371.
28. Joran Friberg, "Numbers and measures in the earliest written records," *Scientific American*, 250: 2, 1984, 116. Friberg argues that the circular sign can be read either as 6 or 10.
29. Robert K. Logan, "Cross cognitive impacts of the notations for writing and numbers" (working paper) 1984, p. 30.
30. Denise Schmandt-Besserat and Diane Jacob-Rost, "Tokens from the sanctuary of Eanna at Uruk," *Forschungen und Berichte*, forthcoming.

Effects of Chunking and Line Length on Reading Efficiency

Stacey A. Keenan

This study investigated whether text with one phrase or "chunk" on each line aids reading in comprehension and proofreading tasks. The chunked format was produced by a computer program that uses syntactic rules to determine chunk boundaries. Contrary to expectations, the chunked format was read significantly more slowly than the standard format in all tasks, at two difficulty levels, and with both screen and paper presentation. A new explanatory variable, line-length variability, could account for these results and also reconcile conflicting findings from previous research. From a literature review, we can infer that both chunking and shorter line lengths have positive effects on reading efficiency, but high line-length variability has a stronger negative effect. Because chunking increases line-length variability, it can interfere with reading.

How can we design text formats to aid reading? This question has generated many hypotheses both from text designers and behavioral scientists. One strategy is to use the findings of psychological research to design text that supports cognitive processes.

One ubiquitous cognitive process is the segmentation and grouping of related information. People segment streams of words into meaningful phrases and sentences as they listen to speech (see Carrithers & Bever, in press, for references) and as they read. Readers pause at the ends of sentences and phrase boundaries (Just & Carpenter, 1980; Mitchell & Green, 1978). These researchers suggest that these pauses correspond to a cognitive process of interclause integration, most likely to link the phrase or sentence to earlier material.

If people need to pause at phrase boundaries for integrative comprehension processing, perhaps reading would be eased if the pause boundaries were visible. One way to reach the goal of designing text to support cognitive processing, then, might be to show phrases typographically. Henceforth, I will call these meaningful phrases "chunks" (Simon, 1974).

Many researchers have investigated whether printing text to show chunk boundaries aids reading. Three methods have been used to show

chunks: (a) printing only one chunk on each line, (b) printing extra space between chunks (spaced-unit method), but putting as many words and chunks per line as fit, and (c) using two lines for each chunk and printing extra space between chunks horizontally (square-span method).

Several researchers have shown that unskilled readers (children or low-ability adults) comprehend more and read faster with texts that are formatted to have one chunk per line (Cromer, 1970; Grist, 1982; Mason & Kendall, 1979). Other studies have shown that segmented formats improve comprehension and increase reading speed for normal adult readers (Frase & Schwartz, 1979; North & Jenkins, 1951; Dean & Schwartz, 1982). However, Dean & Schwartz found no difference in recall between passages with one phrase per line and passages with short lines (not chunked).

But there is conflicting evidence about whether chunked text aids comprehension for adult readers. Several studies found no advantage for adults reading chunked text (Carver, 1970; Klare, Nichols, & Shuford, 1957). Aaronson & Scarborough (1976), although not testing a chunked format, found that subjects who were reading for comprehension (i.e., to answer a yes-no question) did not pause at phrase boundaries, although subjects who were reading for verbatim recall did. Aaronson and Ferres (1983) concluded that "chunking" is only important when the task is at a syntactic level (e.g., a recall or memorization task), not when the task is mostly at a semantic level (e.g., a comprehension task). Other research has found chunking to help in memory tasks (Anglin & Miller, 1968).

Thus, it seems that chunked formats help unskilled readers and help normal, skilled readers if the task involves memorization or syntactic processing. However, there is some conflict about whether chunked formats can help skilled readers to comprehend more efficiently. Differences in reading materials, tasks, and procedures may account in part for the differences in findings about skilled readers. Frase & Schwartz (1979) used more difficult and complex text than any of the other studies, and their subjects performed a verification task, whereby they read a sentence and then tried to confirm it by reading the text. Because the subjects knew what they were looking for, they might have been able to skip some lines of a chunked format. Other researchers used a multiple-choice task and simpler reading materials.

Other uses for chunked text

People have suggested uses for chunked text aside from reading for comprehension or recall. Some editors say that it is easier to edit text

in a chunked format. Many typists type one phrase per line when using word processing equipment, because it makes later search and change operations easier. Because proofreading tasks involve syntactic processing, they probably involve mental segmentation, and hence, chunked formats may be an advantage. As yet, there are no studies testing the effect of a chunked format on proofreading.

Methods of producing chunked formats

With the exception of Grist (1982), all the studies mentioned above relied on human judgment to form meaningful segments or chunks. Boundary judgments are reliable (Frase, Macdonald, & Keenan, in press; Dean & Schwartz, 1982; Johnson, 1970), and Carver (1970) reported that an immediate constituent analysis showed that chunks usually correspond to syntactically significant substrings of sentences. However, some people include "minor breaks" that others exclude (Aaronson & Scarborough, 1976; Klare, et al., 1957). Boundary judgments made by a computer program are consistent (and faster than human judgments). We have such a program at AT&T Bell Laboratories, called *chunk*, which uses syntactic rules to print a text with one phrase per line (Keenan, 1980; Frase, et al., in press). Grist (1982) used this *chunk* program in his study. The segmentation algorithm is based on an analysis of the boundaries marked by a person, and the *chunk* program segments text into chunks that generally agree with those formed by a person. (See the Method section for details.)

Design of the present study

In summary, studies show that at least three factors contribute to reading efficiency: (a) chunking, (b) text difficulty, and (c) task demands. The present study was designed to try to resolve the conflicts among findings of past research about chunked formats, by systematically exploring the relationship among these three factors. To do this, I tested whether segmentation by the *chunk* program aids reading in various tasks with text of two difficulty levels. The tasks ranged from reading for comprehension (semantic processing) to proofreading (syntactic processing).

Because there is evidence that a format with short lines aids reading as much as a chunked format (Dean & Schwartz, 1982), I matched the mean line lengths of the chunked and standard formats. That is, each passage in the standard format had the same mean length as it had in the chunked format. In this way any difference between formats would be attributable to the effect of chunking and not to the effect of mean line length.

If a chunked format does help reading, it would be easy to implement in a screen presentation, but impractical to implement on paper, because it wastes space. This is not a concern for computer-stored text, where any display format is easily and quickly produced from the stored representation. Although video display terminals (VDTs) are becoming widely used, many people still find it unpleasant to read from them. Therefore, I included presentation mode (paper vs. screen) as a factor in this study.

Hypotheses

I hypothesized an interaction between task and format. I expected subjects to perform the proofreading tasks faster when the text was in chunked format than when the text was in standard format. The syntactic editing task, in particular, required that subjects process phrases to get the correct answer. Therefore, I expected subjects to be aided by the chunked format most in that task. In comprehension tasks, readers go beyond the surface or syntactic level, and concentrate on the meaning of the text. Therefore, a syntactically chunked text is less relevant to their purpose. For comprehension tasks, therefore, performance would not differ according to format. However, sentence verification, while involving comprehension, may also involve some syntactic processing. When the reader knows in advance what to look for while reading, he or she may use syntactic cues to choose sections to be read carefully. Therefore, I expected some advantage to the sentence verification task with a chunked format, but not so great an advantage as for the proofreading tasks.

Method

Subjects

Subjects were 24 clerical staff (23 female) from AT&T Bell Laboratories, who volunteered to participate. All had completed high school; none had had more than two years of college. Their mean age was 34.

Reading Materials

Each text passage was about one paragraph long and had about 130 words. The easy passages were from *Adventures in Living Plants*, a sixth grade text by Edwin B. Kurtz, Jr., and Chris Allen, published in 1965 by the University of Arizona Press. Portions of the book were reproduced with permission from the publisher. Although the paragraphs from this book were easy to read, they comprised detailed facts on botany. The reading grade level for all easy paragraphs combined

was 4.8, as measured by the Kincaid readability formula (Coke, 1978). Coke (unpublished paper, 1978) showed that although the Kincaid formula tends to underestimate the reading grade level of easy text, it is more accurate than the Flesch formula, the Automated Reading Index, and the Coleman-Liau formula are on easy text. She also found the Kincaid formula to be the most accurate for measuring difficult text.

The difficult passages were from *The Machinery of the Brain*, a college level book by Dean E. Wooldridge, published in 1963 by McGraw-Hill. Portions of the book were reproduced with permission of the publisher. The paragraphs from this book described the physiology of the brain and some animal behavior that is controlled by the brain. The reading grade level for these paragraphs combined was 16.8, as measured by the Kincaid readability formula. Both books presented much detailed information that was not likely to be known by the subjects. Indeed, after the experiment, subjects reported no special training in or knowledge of the material presented.

The practice passages were selected randomly from a set of reading materials, of varied subject matter and medium difficulty, which have been used in other reading studies. Most of these paragraphs were factual and all were suitable for questioning. The Kincaid reading grade level of these passages taken together was 9.9.

Presentation

The passages were presented in equivalent formats either on a Hewlett-Packard 2645A VDT (dark background, light blue letters) or printed with a daisy wheel printer on white paper. Both displays had the same number of lines on each page or screen, and the same number of words on each line.

The chunked format was produced by the *chunk* program. The standard format was produced by a text formatting program with the line length for each passage set to be the same as the average length of the lines in the chunked format for that passage. However, the right margin was unjustified so that not every line was the same length. Examples of the chunked and standard formats are shown in Figures 1 and 2.

Chunking algorithm

Development of the algorithm for the *chunk* program was based on an analysis of how people mark phrases in text. The program breaks text after certain punctuation marks, and before or after certain words which have been identified as break words. Most break words are

prepositions and conjunctions. These breakpoints are arranged in levels so that only the highest-level or strongest will cause a break when two or more breakpoints are close together.

The *chunk* program was written several years ago at AT&T Bell Laboratories. Its performance was tested in a small experiment with five human judges. There was 90% agreement among the five judges that the chunk boundaries marked by a person demarcated meaningful units of information. In addition, we found high agreement between the chunk boundaries marked by that person and those marked by the *chunk* program. The *chunk* program chose 73% of the same chunk boundaries that a person chose for 60 passages of easy, medium, and difficult text. Another measurement method, used by Johnson (1970), considers each word to be a decision point (i.e., Does

Figure 1. Example of the chunked format.

Things happened so fast I forgot
to tell you about the special chemicals
that make the cell wall stretch like bubble-gum.
The chemicals are called auxins.
Auxins are made by cells
that are rapidly dividing.
Then the auxins move down the stem
to where cells are ready to enlarge.
And then they cause the cells
to grow longer.
Auxins are very potent chemicals in plants.
A small ball of auxins the size
of a pinhead
is enough chemical to start millions
and billions of cells grower.
Three men discovered auxins —
two chemists and a botanist.
The botanist, Dr. Fritz Went,
discovered a way
of measuring the effect of auxins
on plants.
The chemists, Dr. Kogl and Dr. Haagen-Smit,
identified the first auxin.

END

this word end the chunk?). Calculated in this way, the *chunk* program made 89% of the same word-by-word boundary decisions as a person did for the 60 test passages. A more balanced measure, the geometric mean of the word-by-word agreement proportions, shows 83% agreement between the *chunk* program and a person's judgments.

Tasks

Four reading tasks were chosen to represent a variety of reading skills, from high-level comprehension to lower-level proofreading. The tasks were: recognition, sentence verification, syntactic editing, and search for misspellings. The first two tasks can be classified as comprehension tasks, as the subject was required to answer a question based on what she or he read in a passage. The latter two tasks can be classified as

Figure 2. Example of the standard format.

Things happened so fast I forgot to tell you about the special chemicals that make the cell wall stretch like bubble-gum. The chemicals are called auxins. Auxins are made by cells that are rapidly dividing. Then the auxins move down the stem to where cells are ready to enlarge. And then they cause the cells to grow longer. Auxins are very potent chemicals in plants. A small ball of auxins the size of a pinhead is enough chemical to start millions and billions of cells grower. Three men discovered auxins — two chemists and a botanist. The botanist, Dr. Fritz Went, discovered a way of measuring the effect of auxins on plants. The chemists, Dr. Kogl and Dr. Haagen-Smit, identified the first auxin.

END

proofreading tasks, as they required the subject to find a single word error in each passage. These tasks are described below.

1 *Recognition.* Subjects read a passage, and then answered a multiple-choice question about the passage. Subjects had to understand at least the gist of the passage in order to answer correctly. Simply recognizing that phrases occurred in the passage was not enough, because phrases from the passages were used in both correct and incorrect choices.

2 *Sentence verification.* Subjects read a sentence, which was either true or false; then they read a passage to verify the sentence. Verification required information from several sentences to be integrated. Subjects were instructed to read the entire passage and then to re-read the true-false sentence, before giving an answer, even if they thought they knew the correct answer. In this way, each subject read the same number of words and reading rate could be measured. This task was simpler than the recognition task, because subjects knew what information they were looking for as they read. They did not need to understand the entire passage in as great depth as they did in the recognition task. Half of the sentences at each difficulty level in this task were true, half were false.

3 *Syntactic editing.* Subjects read a passage looking for syntactic errors. The errors were single words with the ending changed to make the word the wrong part of speech, the wrong tense, or the wrong number (i.e., plural vs. singular). However, the incorrect word was still a real word. For example, in the following sentence the word "formation" should be the word "formed."

They are formation near the tip of the root.

The passage in Figures 1 and 2 also contains a syntactic error; the word "grower" should be the word "growing." Four of every five passages contained such a word; one passage at each difficulty level had no error. Since these errors were easy to miss (for example, one might read the above example as "They are *a* formation near the tip of the root"), passages without errors were included to discourage subjects from reading passages twice if they did not see the error on the first reading. Hence, subjects did not know if they answered incorrectly. This made the task equivalent to the recognition and sentence verification tasks in that respect.

The subjects were informed that about one out of every five passages would not have a mistake. They were told to read carefully and not to read any passage twice. Only the four scores for passages with errors

were included in the data analysis. Thus there were four scores for each difficulty level in each of the four tasks.

Although this is a proofreading task, subjects could not find the error by reading words in isolation. Since each error was still a real word, the errors could only be found when read in the sentence or phrase context. So it was a lower-level task than either of the comprehension tasks, in that it required little comprehension. However, it was a higher-level task than the next one, search for misspellings, because it required syntactic processing of whole sentences or phrases.

4 Search for misspellings. Subjects were asked to read a passage, looking for misspelled words. As in the syntactic editing task, four of every five passages contained an error (misspelling) and one passage at each difficulty level had no error. The subjects were informed that about one out of every five passages would not have a mistake. Only the four scores for passages with errors were included in the data analysis.

In all tasks, subjects were given instructions for the task they were to perform before they began reading. Subjects were given four practice trials for each task type, except for the syntactic editing task. A pilot test showed that subjects needed more practice with the syntactic editing task, so there were eight practice trials for that task.

Design

A $2 \times 4 \times 2 \times 2$ (Format $\times$ Task $\times$ Text Difficulty $\times$ Presentation Mode) factorial design was used, with repeated measures on task and text difficulty. In other words, each subject performed all four tasks, at both difficulty levels. Format and presentation mode were between-subjects measures. Subjects were randomly assigned to one of four conditions: paper-chunked, paper-standard, screen-chunked, and screen-standard, with six subjects in each group. Subjects read eight passages for each of the two comprehension tasks, and 10 passages for each of the two proofreading tasks. The sequences of tasks and passages were randomized. Within each task, half of the passages were easy, and half were difficult.

Scoring

In tasks such as the ones used in this experiment, subjects can usually sacrifice accuracy for a faster performance or vice versa (Wickelgren, 1977). To avoid the possible confounding of a speed-accuracy trade-off, reading speed was the only dependent measure, while the accuracy was held relatively constant. Subjects were instructed to read carefully

because it was more important for them to get the correct answer than to read very quickly.

Reading rate for the recognition and sentence verification tasks was calculated by dividing the time spent reading a passage by the number of characters (including spaces) in the passage. For the two proofreading tasks, reading rate was calculated by dividing the reading time by the number of characters preceding the answer. I recorded reading times with a digital stopwatch.

Answers were coded in one of three categories: right, wrong, or "wrong, but." The third category was created for answers which were wrong, but did not indicate that subjects were not reading carefully. For five questions (out of 32) the same wrong answer was given by nearly all subjects. This indicates that these questions were too hard for the subjects or were poorly worded, but not that the subjects were not reading carefully. The wrong answer given by most subjects to these questions received the "wrong, but" categorization. In addition, in the syntactic editing and search for misspellings tasks, occasionally subjects would find a word preceding the target that they thought was used or spelled incorrectly. This type of mistake was also scored as "wrong, but." Right answers were scored as 1.0, wrong answers were scored as 0.0, and "wrong, but" answers were scored as 0.5. Thus, the means for errors can be interpreted as the proportion of answers which were correct.

Results

Accuracy. Although accuracy was not the dependent measure, it was analyzed to see if subjects were trading reading speed for accuracy, or vice versa. The mean proportion of correct answers was .81, meaning that 81% of all the answers given were correct. A $2 \times 2 \times 4 \times 2$ (Mode $\times$ Format $\times$ Task $\times$ Difficulty) split-plot analysis of variance was conducted, with subjects nested in mode-format groups and subjects crossed with task and difficulty. Subjects were treated as a random variable. Two main effects were significant: task, $F(3,60) = 2.90, p < .05$, and difficulty, $F(1,20) = 20.86, p < .01$. The main proportion of correct answers was .86 for the easy text and .76 for the difficult text. Subjects performed most accurately on the search for misspellings and recognition tasks (proportion correct = .86 and .83, respectively), and somewhat less accurately on the syntactic editing and sentence verification tasks (proportion correct = .77 and .78, respectively). Because subjects in different mode-format groups performed with equal accuracy, we may use reading rate as the dependent measure in the

rest of the analysis. Subjects in different mode and format groups did not trade accuracy for speed in different ways.

Reading Rate. All subjects varied greatly in reading rate, but one subject in the screen-chunked group read much faster than any other subject (up to 50 characters per second for some passages) and answered only half of the questions correctly. Her scores were eliminated and replaced by the means of the other five subjects in the screen-chunked group.

The same split-plot analysis of variance design described above for accuracy analysis was used to analyze reading rate. Reading rate was measured in characters per second. The overall mean was 17.32 characters per second (and 2.97 words per second, or 178 words per minute). Reading times, like reaction times, are often positively skewed when the measured times are short. In this case for example, no subject read less than 5 characters per second. Therefore, there was not much variation in reading rates for subjects who were slower than average, but faster subjects ranged up to 40 characters a second for some passages. I performed a logarithmic transformation on the data, as is recommended with skewed measures (Myers, 1966).

Contrary to expectations, the chunked format was read at a significantly slower rate than the standard format, $F(1,20) = 5.78$, $p < .05$. Subjects read the chunked format at a rate of 15.8 characters per second and they read the standard format at a rate of 18.85 characters per second. In other words, the chunked format caused a 16% reduction in reading speed. Mode had no significant effect on reading rate ($F(1,20) < 1.0$), and no interactions involving mode or format were significant.

The main effects of difficulty and task were significant, as expected. The easy text was read faster than the difficult text, $F(1,20) = 21.83$, $p < .01$. Subjects read the easy text at a rate of 17.92 characters per second and the difficult text at a rate of 16.72 characters per second. Task also had a significant effect on reading rate, $F(3,60) = 57.26$, $p < .01$. The comprehension tasks were read faster than the proofreading tasks. The reading rates for each task in characters per second (cps) were: sentence verification = 20.5 cps; recognition = 19.1 cps; search for misspellings = 16.0 cps; and syntactic editing = 13.7 cps.

Tests of the strength of association for the significant effects showed that format accounts for 8% of the variance, task accounts for 32% of the variance, and difficulty accounts for 2% of the variance. Thus, the significant effects account for a total of 42% of the variance in reading rates. The great deal of variability between subjects probably accounts for most of the rest.

Discussion

The results are quite different from those I hypothesized. Not only did the chunked format cause slower reading, but task did not interact with format. That is, the chunked format was detrimental to the performance of all tasks, whereas I had expected the chunked format to ease significantly performance of only some tasks. Even past studies that found no beneficial effect of chunking did not find chunking to be so detrimental to reading efficiency.

To find an explanation for these contradictory results, we must look at how this study differs from past studies. The formats used in this study differed from the formats used in others in two important ways. First, I used a chunked format that was produced by a computer program, rather than by human judges. Second, I matched the average line lengths of passages in both formats, so that the standard format had shorter lines than the standard used in other studies. I will now consider the possibility that computer-chunked text or matched line lengths produced these surprising results.

Were the chunks unacceptable?

One possibility is that although chunking may help readers, *chunk's* chunks may not be good enough. Frase & Schwartz (1979) showed that a misapplied segmentation strategy caused slower reading than standard text. If the chunks produced by the *chunk* program represent a poor segmentation strategy, then one would expect the chunked format to cause slower reading. Although the chunked format did produce slower reading, there is some evidence that *chunk's* chunks are like those people produce (see the section on the chunking algorithm in the Method section). However, without a more stringent test of *chunk's* acceptability, we cannot conclude whether the *chunk* program's chunks were so poor as to produce the slow reading. However, it seems unlikely that this is the case.

Was the chunked format too variable in line length?

Let us now consider the possibility that the line lengths, and not the chunks, caused the unexpected results. When one looks at the two formats (see Figures 1 and 2), the striking difference between them is that the chunked format has such a jagged right margin. Although the mean line length was held constant, the line-length variability necessarily was not. The overall mean line length for the chunked text was 30 characters per line, and was 32 characters per line for the standard text. However, the standard deviations of the line lengths of the standard passages averaged 2.6 characters, whereas the standard

deviations of the line lengths of the chunked passages averaged 10.2 characters.

Effect of line-length variability on reading

Extreme variability in line length may slow reading by disrupting the rhythm of eye movements, in particular, of return sweeps. A return sweep is the movement of the eye from the end of one line to the beginning of the next. They are usually somewhat inaccurate; "under-shoots" often occur, which require the eye to make a corrective regressive movement to get to the beginning of the line (O'Regan & Levy-Schoen, 1979). Return sweeps take from twice as long to five times as long to make as ordinary saccadic movements. Leisman (1978) hypothesizes that when lines of text are both right- and left-justified, the reader builds an internal map with the coordinates of the beginning and ending of lines. Then the brain constructs a "program" to execute the return sweep automatically. If this explanation is true, the extreme unpredictability of the distance from the right margin to the left margin in the chunked format would make such an automatic program impossible for a reader to set up. Even if Leisman's hypothesis about automatic return sweeps is not true, variable lines still might cause return sweeps to be inaccurate more often. If so, reading speed must decrease because corrective saccades take as long to make as normal saccades do.

Effect of mean line length on reading

The present study differed from most previous studies (except Carver, 1970) by holding mean line length constant across formats. Subjects in both format groups read text with medium-length lines (about 3 inches wide, averaging 5 words per line). Tinker (1963) cites evidence that people prefer to read lines of about 14-31 picas (2-5 inches) in width and that they also read such widths more quickly. (Tinker's width recommendations vary with point size and leading. These are for standard 10-point type with 2-point leading.) Because most previous studies did not control for line length, we cannot tell if, when subjects read more efficiently with a chunked format, they were responding to the presence of visible chunks, or to the shorter lines.

Besides being preferred by readers, medium-length lines may help reading eye movements. For example, return sweeps may be more accurate when lines are of medium length, because the eye can see the beginning of the next line in peripheral vision, when it is at the end of a line. Tinker (1965) reports that regressive eye movements increased by over 50% when text had long lines (43 picas). He attributes this

increase to readers making inaccurate return sweeps, and the subsequent need to make a corrective regressive movement. A revised theory of how to improve text design needs to consider eye movement limitations in addition to aids to cognitive processing. That is, a potential aid to comprehension, such as placing one chunk per line, should not be allowed to disrupt eye movements, or the cognitive advantage may be offset by the disruption.

Hence, the unexpected results of this study suggest an explanation for previously conflicting experimental results. Below, I consider three critical variables in classifying text formats: chunking, mean line length, and line-length variability. These three together might allow us to resolve contradictory findings of other research. Presumably, a chunked format aids reading, and a moderate line length with little variability aids eye movements.

Reinterpretation of studies with chunked formats

Most studies, including this one, have ignored a potentially important characteristic of text format — that of line-length variability. Most past studies with chunked formats have also ignored line length as a variable. Another look at the results of past studies shows that if these variables are considered, most conflicts might be resolved. Table I shows a reinterpretation of the results of this and three previous studies on the effect of chunked formats on the reading efficiency of skilled readers. The new interpretation considers three format variables, chunking, line length, and line-length variability — whether the experimenters specifically included all three variables or not. Chunking and medium mean line length are considered to be advantages; high variability in line lengths is considered a disadvantage to efficient reading. The first column of Table I names the study and the second column names the formats that are being compared (in the original experimenter's terms). The third column names which variables differ between the formats being compared, whether they were explicitly studied or not. The fourth column shows which format was read faster. The fifth column shows what we can conclude from the results of the comparison in light of a three-variable theory.

Frase & Schwartz (1979) compared five formats. Their meaningfully-segmented, meaningfully-indented format (MI) is, in my terminology, a chunked format. They found indentation to be only a weak aid to reading; therefore, I will include it as a variable in Table I, but I will not dwell on it. Their other three non-standard formats were permutations of chunking and indenting with misapplied segmenting and indenting strategies. All four of these experimental formats had

Table I. Reinterpretation of Chunked Format Studies Using Three-variable Theory

Study	Comparison	Variables which differ	Result	Conclusion
Frase & Schwartz (1979)	S vs. MI	chunking, line length, ll variability, indentation	MI 18% faster	chunk + indent > variab. + med. lines
	S vs. MNI	chunking, ll variability, line length	MNI 12% faster	chunk > variab. + med. lines
	S vs. NMI	line length, ll variability, indentation	S = NMI	med. lines = variab. + indent
	S vs. NMNI	line length, ll variability	S 16% faster	med. lines < variab.
	(MI + MNI) vs. (NMI + NMNI)	chunking	M's > NM's	chunking = advantage
North & Jenkins (1951)	S vs. spaced-unit	chunking	spaced-unit faster	chunking = advantage
Carver (1970)	chunking vs. "newspaper"	chunking, ll variability in-line variab.	chunk = newspaper	chunk = variab. - in-line variab.
Keenan (1984)	S vs. chunking	chunking,	S 19% faster	chunk < variab.

S = standard format

MI = meaningfully-segmented, indented format

MNI = meaningfully-segmented, nonindented format

NMI = nonmeaningfully-segmented, indented format

NMNI = nonmeaningfully-segmented, nonindented format

ll variability = line length variability

medium lines that varied in length as compared to the standard format (S), which had a right margin at 66 characters. The first row in Table I shows the comparison of the chunked format (MI) and the standard format (S). The MI format was chunked, had medium-length lines, was meaningfully indented, and had high line-length variability. The format characteristics of the S format were opposite to those of the MI format. Frase & Schwartz found that subjects read text in the MI format 18% faster than they read text in the S format. From this we can conclude that the combination of the advantages to reading of chunking, indentation, and medium-length lines outweigh the disadvantage of extremely variable line lengths. Because chunking, indentation, and line length are confounded, we cannot conclude simply that chunking aids reading. The second row in Table I shows that the combination of the advantages of chunking and medium-length lines (without the advantage of indentation) are enough to outweigh the disadvantage of variable line lengths. In the third row we see that without the advantage of chunking, medium-length lines and indentation do not outweigh line-length variability, but equal it.

The fourth row of Table I shows the comparison of the improperly segmented, medium-length line format (NMNI) and the standard format (S). These two formats differ only in length (the NMNI format had shorter lines) and in line-length variability (the NMNI format had variable line length). Because subjects read the standard format 16% faster, we can conclude that the single advantage of having shorter lines is outweighed by the disadvantage of having variable line lengths.

Frase & Schwartz did one comparison which did not confound line length and line-length variability with chunking. The fifth row in Table I shows the comparison of the two chunked formats with the two non-meaningfully segmented formats. All of these formats had variable, medium-length lines. Because the chunked formats were read significantly faster than the others, we conclude that a chunked format does aid reading, when all other things are equal.

The spaced-unit format used by North & Jenkins (1951) did not introduce line-length variability or shorter lines, because it did not place one chunk on each line. Instead, the spaced-unit format resembles the standard format in every way, except that there are extra spaces placed at chunk boundaries. Hence, their study really did show that chunking is an advantage in reading for comprehension.

Carver (1970) found no difference in reading speed or comprehension between a chunked format and a "newspaper" format. Carver's newspaper format was similar to the standard format in the present

study, in that the lines were as short as chunked lines, but his format was right-justified. Although a right-justified format eliminates line-length variability, it also introduces within-line variability, because extra spaces are inserted between words to force the last word on the line to end at the right margin. Campbell, Marchetti, & Mewhort (1981) show that unpredictable spaces between words (produced by a fixed-space, right-justification technique like the one Carver used) reduce reading speed about as much as unpredictable right margins do. From Carver's study we may conclude that the advantage of a chunked format with uniform in-line spacing over a non-chunked format with within-line variability, is offset by the disadvantage of variable line lengths in the chunked format.

As shown in the last line of Table I, the present study confounded the effects of chunking with line-length variability. That subjects read the standard format 19% faster than the chunked format shows that the advantage of chunking is outweighed by the disadvantage of line-length variability.

In summary, this analysis presents evidence for three findings:

- 1 Chunking appears to be an advantage when line length and variability are held constant.
- 2 Medium-length lines are an advantage in reading when chunking and line-length variability are held constant.
- 3 Low line-length variability (i.e., a regular margin) may be an advantage in reading.

In addition, line-length variability seems to have the strongest influence on reading efficiency. As Table I shows, neither chunking nor shorter lines alone provides an advantage great enough to override the disadvantage of having variable line lengths. Eye movement disruptions seem to be more detrimental to reading speed than chunking is helpful to comprehension.

Practical implications

If eye movement disruptions are the reason for the chunk disadvantage, then chunked formats, as they have been implemented to date, do not provide any advantage to reading over shorter line standard formats, such as formats found in newspapers. However, future chunk algorithms might try for less variability. A format between chunked and medium-length line standard might be better than either. Such an algorithm would always start a sentence on a new line, as *chunk* does now. In addition, it would need to consider whole sentences at a time in deciding where to break lines. Most sentences, especially when the text

is difficult, have more possible chunk boundaries than are used. A smarter program would look ahead to see which breaks would result in the least variation in length. Plass & Knuth (1982) describe a line-breaking algorithm that considers a whole paragraph at a time rather than a single line at a time in choosing breaks, and minimizes "badness" of breaks across the paragraph, by assigning penalties to breaks that are considered undesirable for one reason or another. A new chunking algorithm could work similarly, assigning a penalty of zero to all possible chunk boundaries and assigning high penalties to all other possible boundaries. Minor chunk boundaries might be assigned low penalties, rather than zero penalties. The highest penalties would go to spaces between words which should not be split up. Demerits would be assigned to lines that are outside variability limits. The new *chunk* program would then create chunks to minimize penalties, and also to minimize line-length variability.

Conclusion

The unexpected results of this study clarify the contribution of line-length variability to reading efficiency. When variability is considered, the conflicts among results from studies of chunked formats are resolved. A chunked format does ease reading, and so does a format with medium-length lines. But these advantages are not so strong as the advantage of little variation in line length. High variability in line lengths disrupts eye movements enough to outweigh the benefits of a chunked format with medium-length lines.

Further studies are needed to confirm these conclusions. A study similar to the present study, but which includes line-length variability as an independent variable, would show the individual and combined effects of chunking, mean line length, and line-length variability on reading. Studies of eye movements when the reader is reading text with lines of varying lengths would show whether inaccurate return sweeps are the cause of slower reading in such formats. Finally, further studies of the acceptability of the chunks produced by the *chunk* program (and a new *chunk* program, which minimizes variability), are needed before entirely acceptable chunked formats can be produced.

References

- Aaronson, D., & Ferres, S. (1983). A model for coding lexical categories during reading. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 700-725.
- Aaronson, D., & Scarborough, H. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 56-70.
- Anglin, J.M., & Miller, G.A. (1968). The role of phrase structure in recall of meaningful verbal material. *Psychonomic Science*, 10, 343-344.
- Campbell, A.J., Marchetti, F.M., & Mewhort, D.J.K. (1981). Reading speed and text production: A note on right-justification techniques. *Ergonomics*, 24, 633-640.
- Carrithers, C., & Bever, T.G. (in press). Eye-fixation patterns during reading confirm theories of language comprehension. *Cognitive Science*.
- Carver, R.P. (1970). Effect of a chunked typography on reading rate and comprehension. *Journal of Applied Psychology*, 54, 288-296.
- Coleman, E.B., & Hahn, S.C. (1966). Failure to improve readability with a vertical typography. *Journal of Applied Psychology*, 50, 434-436.
- Cromer, W. (1970). The difference model: A new explanation for some reading difficulties. *Journal of Educational Psychology*, 61, 471-483.
- Dean, R.S., & Schwartz, N.H. (1982). Inducing comprehension in the inconsistently lateralized reader. Manuscript submitted for publication.
- Frase, L.T., Macdonald, N.H., & Keenan, S.A. (in press). Intuitions, algorithms, and a science of text design. In T. Duffy & R. Waller (Eds.), *Designing usable text*. New York: Academic Press.
- Frase, L.T., & Schwartz, B.J. (1979). Typographical cues that facilitate comprehension. *Journal of Educational Psychology*, 71, 197-206.
- Grist, S. (1982). *Effect of discourse segmentation on comprehension and reading rate of eighth-grade children*. Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.
- Johnson, R.E. (1970). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior*, 12-20.
- Just, M.A., & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Keenan, S.A. (1980, April). Computer projections of the cognitive effects of text changes. In L.T. Frase (Chair) *Computer aids for writing and text design*. Symposium conducted at the Annual Meeting of the American Educational Research Association, Boston.
- Kirk, R.E. (1982). *Experimental design: procedures for the behavioral sciences* (second edition). Monterey, CA: Brooks/Cole Publishing Company.
- Klare, G.R., Nichols, W.H., & Shuford, E.H. (1957). The relationship of typographic arrangement to the learning of technical training material. *Journal of Applied Psychology*, 41, 41-45.
- Kurtz, E.B., & Allen, C. (1965). *Adventures in living plants*. Tucson: University of Arizona Press.

- Leisman, G. (1978). Ocular-motor system control of position anticipation and expectations: Implications for the reading process. In J.W. Senders, D.F. Fisher, & R.A. Monty (Eds.), *Eye movements and the higher psychological functions*, (pp. 195-207). Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Mason, J.M., & Kendall, J.R. (1979). Facilitating reading comprehension through text structure manipulation. *The Alberta Journal of Educational Research*, 25, 68-76.
- Mitchell, D.C., & Green, D.W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology*, 30, 609-636.
- Myers, J. L. (1966). *Fundamentals of Experimental Design*. Boston: Allyn and Bacon.
- North, A.J., & Jenkins, L.B. (1951). Reading speed and comprehension as a function of typography. *Journal of Applied Psychology*, 35, 225-228.
- O'Regan, K., & Levy-Schoen, A. (1979). The control of eye movements in reading (tutorial paper). In P.A. Kolars, M.E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 1, pp. 7-36). New York: Plenum Press.
- Plass, M.F., & Knuth, D.E. (1982). Choosing better line breaks. In J. Nievergelt, G. Coray, J.D. Nicoud, & A.C. Shaw (Eds.), *Document preparation systems* (pp. 221-242). New York: North-Holland Publishing Company.
- Simon, H.A. (1974). How big is a chunk? *Science*, 183, 482-488.
- Tinker, M.A. (1963). *Legibility of print*. Ames: Iowa State University Press.
- Tinker, M.A. (1965). *Bases for effective reading*. Minneapolis: University of Minnesota Press.
- Wickelgren, W.A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67-85.
- Wooldridge, D.E. (1963). *The machinery of the brain*. New York: McGraw Hill.

Author Notes. This paper is based on a master's thesis prepared for New York University, "Effects of Display on Reading Tasks." Special thanks go to my colleague and thesis advisor, Larry Frase, who contributed much advice, discussion, and encouragement.

The Visual Editing of Texts

Fernand Baudin

Words mesmerize more than they inform. Any piece of writing is an image as well as a message. Hence, whether you want to mesmerize or to inform you must acquire a mastery over the proper letterforms and how to arrange them on any surface. This is more than can be taught in primary schools where all teaching begins as handwriting. Typography, as an extension of handwriting, can no longer be considered a preserve of specialized craftsmen. Therefore all teachers at any level in any branch of learning should be able to analyze, to study, and to describe any text as a *constellation* of alphabets and a *configuration* of columns and lines — that is, to perceive what is on the page and to practice the visual editing of the case one wants to defend. Adapted from a lecture given at Stanford University, May 1983.

The idea I want to develop here can be expressed briefly as the argument for analyzing, studying, and describing text pages as *configurations* of columns and lines and as *constellations* of alphabets. But before I begin developing the idea as an extension of the teaching of handwriting, I want first to put my subject into perspective.

The fundamental importance of handwriting and all its technological extensions as typewritten, printed, and computerized matter in the various worlds of learning, in the several systems of education; in the arts, in the sciences, in politics; in business as well as in the business of everyday life need not be stressed. Of importance here is the fact that Stanford University is the very first “to perceive the need for a new program in the study of digital typography where computer scientists and artists may be taught the fundamentals of typographic understanding.” Given the circumstances this may prove a decisive step towards meeting, if not solving, the innumerable challenges of the Computer Age. If only because the Stanford example may induce more universities to resume their leading role in pressing advanced technology into the service of the written word — which is the technology of technologies — just as the printed word and alphanumerals have been pressed into the service of NASA to send homo americanus physically to the moon and back. This is not only in keeping with all due respect for the poetics and the aesthetics of learned tradition, but also in conformity with the ever pressing and timeless urge to establish more

and more connections in an ever closer network of human interchanges.

This is about the right time to be mindful of the historical fact that for better and for worse advanced technologies and writing have always been directly linked with the seats of power and learning. In princely and clerical chanceries, in libraries and universities, it was a saying and a fact that the pen was mightier than the sword. This is as much as to say that the uses of the pen were not exclusively calligraphic. It is not less relevant to remember that the introduction and development of the art of “writing without pen on the press” (to use and translate the contemporary description) was eagerly adopted by the universities, the humanists, and the reformers. And that all this eventually resulted in the one new format which was invented after the passage of scroll to codex, namely the newspaper format. Clearly the press is now mightier than the pen. It even proved mightier than a President of the United States. This is not to say that the press — or the screen or the chip of whatever — is mightier than the alphabet and the written word.

Given the circumstances it is altogether relevant to reconsider what used to be called the teaching of writing, meaning handwriting. This is generally discussed in terms of style, method, model, or tool: pen, nib, ball-point, or whatever. What strikes me is that in spite of all the cultural differences of the older continent as opposed to the traditions of the New Continent, there is everywhere an obvious malaise concerning the teaching of handwriting in particular as well as concerning teaching in general. Fair enough. It is pointless to inflict on you citations from *Visible Language* let alone recent issues of *Time*, *U.S. News*, *Reader's Digest*, et al. They are all too familiar. As early as the thirties a Belgian teacher examined 30 contemporary Belgian methods and concluded that teachers are more important than models or systems or methods. I agree.

Today some radicals want to do away with all the handwriting humbug — suggesting “Writing is something of the past.” I disagree most emphatically, as a matter of course, and shall content myself with recalling that the U.S. Postmaster General a few years ago passed the message that “Handwriting is a root of democracy.” Good for him. Good for you. And, incidentally, good for me; because his commemorative postage stamp coincided with a modest contribution I had made myself to the same effect in *Visible Language* at about the same time. This must suffice to suggest all the implications of depriving anyone of the essentials of social, political, and individual freedom and self-expression — namely: reading and HANDwriting.

Recent work of French historians helps looking in a different direction and for a different approach. In the two volumes of *Lire et Écrire, l'Alphabetisation des Français de Calvin à Jules Ferry* (Editions de Minuit, 1977) François Furet and Jacques Ouzouf have analyzed and worked out 16,000 answers to a questionnaire which was designed in 1877 by M. Maggiolo, a retired teacher. He wanted to know something about the spread of literacy in four periods: 1686-90, 1786-90, 1816-20, 1872-76. Sixteen thousand teachers were willing in 1877 to look up in the archives and find out how many people managed to sign their names in their marriage certificates during these four periods of time. The result is a unique monument of historic information, remarkably summarized by the authors as follows: "Literacy is *not* a product of the school system. Literacy as a product of the school system is a mistaken view shared in equal measure by the most irreconcilable enemies. The republicans believe that the French Revolution introduced the primary school. While the monarchists believe that the French Revolution made an end of it. As a matter of fact, the French masses went on learning to read and write from Calvin's day until the end of the nineteenth century — and 1789 was in no sense a landmark in the process. Both practices have been encouraged, organized, and financed by the families and by the communities, in other words by society itself. They were felt as equally indispensable: for salvation in the first place, when Reformation had dislodged the clergy from their monopoly of Holy Writ. Later on literacy was equated with modernity because there can be no market-place of any description without written contracts between free people. That is why the history of literacy is so dependent on the various kinds of inertia in a social fabric which is considered as a cultural model to be handed down by the 'elite' to the 'popular' classes."

My second French source is also a lasting monument: the *Histoire Mondiale de l'Éducation* (4 indexed volumes, a total of 1700 pages by 40 contributors) under the editorship of two professors in French universities and specialized in the subject. The title *World History of Education* is self explanatory. This is the kind of work that even the editors prefer not to summarize in a few sentences. I have read a lot of it and I intend to read much more; I cannot say that I read the lot. Therefore I confess that I am going to use it shamelessly to my purposes. I am not even going to try to give you a broad outline of its contents.

In my view this work has two essential merits. First, it tries to assess a complicated subject in the widest possible context. In doing so it brings together a large amount of valuable information scattered

among countless specialized sources (to be found in the bibliographies of the individual contributions). Second, it contains far more direct and oblique criticism than the individual contributors would accept from any outsider. One author even admits that "The USSR and Eastern Europe, the Countries of the Warsaw pact, dispensed with many educational experiments, utopian or uncontrolled." Probably because they found other ways to spend the taxpayers' money.

In the last paragraph of his chapter on the techniques of elementary training during the nineteenth century, M. Vial is quite explicit: "Contempt is now the general attitude for this menial and manual discipline." The chaos resulting from such an attitude over a period of one hundred years is hardly surprising, however alarming it may be. To try and put the blame on any one person or political party or any particular system would be a waste of time.

How this attitude developed and prevailed in the face of the glaring fact of the servability of handwriting can be explained very briefly. Learning and for that matter teaching handwriting has never been a pleasant occupation. Not everybody has a calligraphic penchant. Not everybody is a born teacher. For most people the serious business of writing was copying, that is to say: drudgery, pure and simple. By the end of the nineteenth century writing masters had all but died out — at least in the West. Literary and commercial hands were all alike and fast degenerating. Printers were no longer humanists. Teachers and school inspectors were coming into their own under the compulsory school system for educating the masses. By that time the graphologists had developed a method which associated calligraphy exclusively with the copperplate hand, then in its more degraded state and equated with a total lack of personality — as well as plain stupidity. Physicians, hygienists, and psychologists resumed the battle over vertical against sloping letterforms, which according to Javal (France, 1905) had divided the writing masters for over a century. In the United States Thomas Edison took sides in favor of the vertical style. It was also at that time that longhand gradually gave way to be finally and totally replaced by the typewriter in all commercial and administrative offices.

Not surprisingly, handwriting is as unpopular as ever — or even less popular.

Typography fares no better. In one generation photocomposition and word processing have disrupted the professional training of printers and composers, such as it was and such as they were. I have a graphic illustration of the resulting situation — at least in French speaking countries. Right now, on my desk in Belgium, I have five manuscripts

by five authors, all of them academics and working daily on books published in France during the sixteenth century — the best book typography there ever was. As a matter of course, they are daily consulting the bibliographies of their subject. Even so when it came to put pen on paper, to type their own bibliography, they were at a loss. The publisher of the volume is an antiquarian book dealer and a newcomer in publishing. So is the printer in this particular branch of typography. They came to me not only as a book designer but also as a teacher because they had never been given any information about book production — so that they cannot even learn from the books they are using, consulting, studying daily. Of course, that is good for me! But the total situation is a serious matter indeed. So much so that in France, Charles Peignot in 1980 went straight to Georges Bonnin, directeur de l'Imprimerie Nationale asking him to join forces in order to try and restore the typographic tradition in France.

The Imprimerie Nationale was founded by Richelieu in 1640 and is to the French typographic tradition what the Académie Française (also and significantly founded by Richelieu, 1636) is to French literature. The antecedents of the Typefoundry DeBerny-Peignot can be traced back to Honoré de Balzac in the nineteenth century. Since the Art Nouveau period, the Peignots have been leaders in the “typographie a la française.”

The Imprimerie Nationale accepted the challenge and for the last three years an informal group of twenty people have met monthly in the Imprimerie Nationale, with the late Charles Peignot as the chairman of the we call “le CERT,” Centre d'Etudes & de Recherches Typographiques. The first decision made was to publish a book to celebrate the Tradition Française (what else?) under the poetic title, *De Plomb, d'Encre, et de Lumière*, which to the French mind suggests the evolution from hot metal to cold type. This book was followed by meetings with an ad hoc commission interministérielle which will very soon make practical decisions — at least I hope so.

I contributed a chapter to the book, “Constellations et Configurations d'Écritures,” which I already described as a method for analyzing and describing text pages (as distinct from title pages, in the manner of DeVinne). By the configurations of text matter I mean the lines, columns, notes (footnotes, headnotes, endnotes, sidenotes), cut ins, page numbers, signatures, running heads, etc. By the constellation I mean the several letterforms, written or printed, in various styles which are eventually combined on one text page. All this is illustrated with examples taken from the homework of a schoolboy, a typewritten page of copy, a commercial letter, a few novels, dictionaries, and a

daily newspaper. Every single example — of growing complexity and diversity — is fully commented on. These examples suggest the kind of exercises which would be done any time, anywhere, with whatever piece of written or printed matter happens to be around. The object of the exercise is not to help boys and girls forget about handwriting. Quite to the contrary. This is intended to help the boys and girls as well as the scientists all through their several studies to become aware of the form as well as the content of what they are reading or writing.

Teachers and students alike can only benefit from being made more perceptive of what I would describe as the *visual editing* of any piece of written matter that they may be handling as reader and as writer. Everybody benefits by being more alert to the fact that the visual editing is not part of any medium but must be made part of any text. Calligraphy in the 1980's is clearly an art form in its own right and should be taught as such. Visual editing as an extension of handwriting should be part of any course of studies. In such a way and to the extent that even a hopeless scribbler ought to know how any piece of writing worth distributing, should be adequately edited — visually as well as grammatically — in order to be efficient as well as acceptable to the addressees. Whatever the system used, sobriety, clarity and coherence (i.e., style) can be given. This is a matter of culture.

If this is as simple as all that, why ignore it? If it is difficult, this is only one more reason why visual editing should be taught generally. If only as the one technique which is instrumental during a lifetime in all technologies and arts. Also because everybody can really *do* something about that at least. And however modest, it is something meaningful to oneself and to the community. It should no longer be considered as a matter of professional, specialized training. Today it is a matter of general literacy.

Hebrew Micrography: One Thousand Years of Art in Script

Leila Avrin

The art of Hebrew micrography — minute writing in the shape of objects or designs — began in the late ninth century in Eretz Israel. From there it spread to Egypt, Yemen, and Europe, where it reached its height from the thirteenth to fifteenth centuries. While nearly all medieval micrography was in Bibles, after the Renaissance it appears in marriage contracts and on other sheets of parchment or paper hung on the walls of synagogues and homes. This longest lasting of Jewish art forms continues to be widely practiced today by scribes and calligraphers.

In all literate cultures, writing is used artistically. The subtle abstractions of Chinese and Japanese calligraphy, the elegant distortions of Arabic inscriptions, and the elaborate zoomorphic and historiated initial letters of Medieval Latin manuscripts are all examples of script as decoration. In the ninth century Jewish scribes in Eretz Israel invented Hebrew micrography — a uniquely Jewish art form — and this tradition has been sustained until the present day.

Micrography (as shown in the accompanying illustrations) is minute script written into abstract patterns or formed into the shape of objects, animals, or human figures. Most people are familiar with some form of micrographic writing: the Passover Haggadah on a goose egg, Psalms on a grain of wheat, a depiction of the Western Wall on a sheet of now-graying paper. There are two basic types of Hebrew micrography. The less common is the calligram, where lines of unequal length are written in parallel rows so that the picture is composed entirely of text (Figure 1). More typical, however, is the microgram, in which the minute Hebrew text delineates the outline of the subject. (Figures 2-4).

Byzantine and Medieval European poets often used the calligram form: Publilius Optatianus Porfyrius in fourth-century Byzantium, Venantius Fortunatus in sixth-century Europe, and Hrabanus Maurus in the ninth-century Carolingian Empire revived calligram art, which began with Aratus and other Classical Greek poets. While these literary calligrams constitute one of several European traditions of decorative writing, they were not written in the diminutive script we find in Hebrew micrography. Jewish scribes may, at times, have been



Figure 1. *Flying Camel*, symbol of the Levant Trade Fair, Tel Aviv, 1933.



Figure 2. *Solomon and the Queen of Sheba*, Frankfurt am Main, lithograph, 1899.

Figure 3. Detail of Figure 2,
Solomon and the Queen of Sheba.

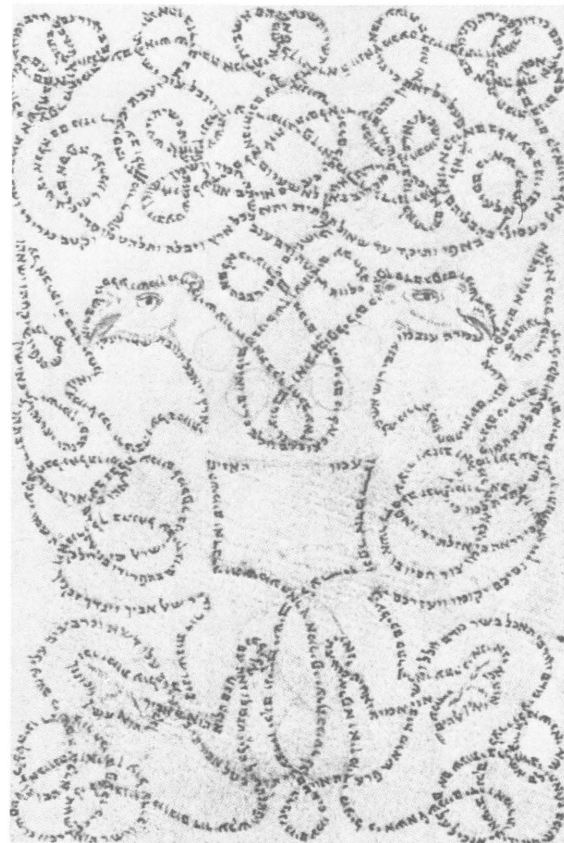


Figure 4. Detail of the title page of *Circumcision Book* by scribe Mordecai (Marcus) Donath of Nitra (Czechoslovakia), Hungary, 1825. Text is the Song of Moses (Deuteronomy 32).

influenced by literary calligrams, but the actual source of Hebrew micrography should not be sought in the Classical Greek, Byzantine, or Carolingian calligrams.

Hebrew micrography was the creation of the masorah scribes of Tiberias in Eretz Israel. The *masorah* ("tradition") is the system of marginal biblical notes which counted and listed each word in the Hebrew Bible, how many times and where it appeared in exactly the same form. The *soferim*, ritual scribes, adept at writing tiny mezuzot ("doorpost" scrolls), which had to be written in a disciplined, minute hand, were already accustomed to minuscule script. Figuring the text into designs was an outlet for their creative talents while occupied with the drudgery of copying out the masorah. The earliest dated Medieval Hebrew manuscript, the Moshe Ben-Asher Codex from 895 CE, already shows micrographic masorah in the two forms it assumed throughout the period, as marginal decoration accompanying the Bible text ("internal micrography"), and as carpet pages surrounding the Bible text ("external micrography"). Members of the Ben-Asher family were considered master *masorettes*, from the late eighth to the early tenth centuries.

In the earliest Bible codices of Eretz Israel and Egypt, the decoration was usually geometric and abstract, in keeping with the iconoclastic nature of contemporaneous Islamic art, although architectural and vegetal motifs are also found. Letters themselves, as in the name of the scribe or the patron, were also occasionally written out in calligram micrography. The textual subject matter of external micrography was not always the masorah; longer biblical passages are sometimes found; and several *ketubot* ("marriage contracts") from the Cairo Genizah show that micrography was not limited to Bible manuscripts alone.

The uniqueness of micrography as a Jewish art form lies not only in its origins, but also in its continued existence. Handed down from one scribe to another, generation after generation, it spread from Eretz Israel and Egypt southward to Yemen and northward to Europe. In Yemen, Hebrew micrography reached its zenith in the fifteenth century. Marginal masorah there was simple and geometric, and closely knit parallel lines, zigzags, and diagonals were popular designs. Identical facing carpet pages, usually with a central rosette, distinguish its external micrography from the decorative end pages of the Tiberian and Egyptian codices (Figure 5). Fish forms are the earliest Middle Eastern use of animals in this context, but by the fifteenth century micrographic birds and beasts abounded in European manuscripts (Figure 6). The textual material of Yemenite carpet pages was biblical, with Psalms as the favorite.

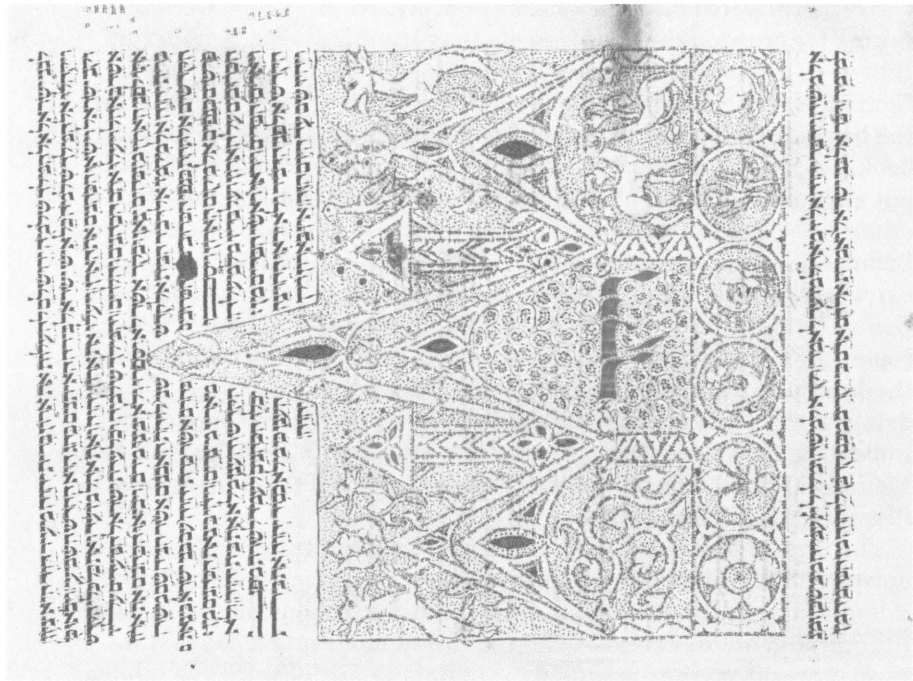


Figure 6. *Pentateuch*, beginning of Leviticus, Bavaria, 1290.

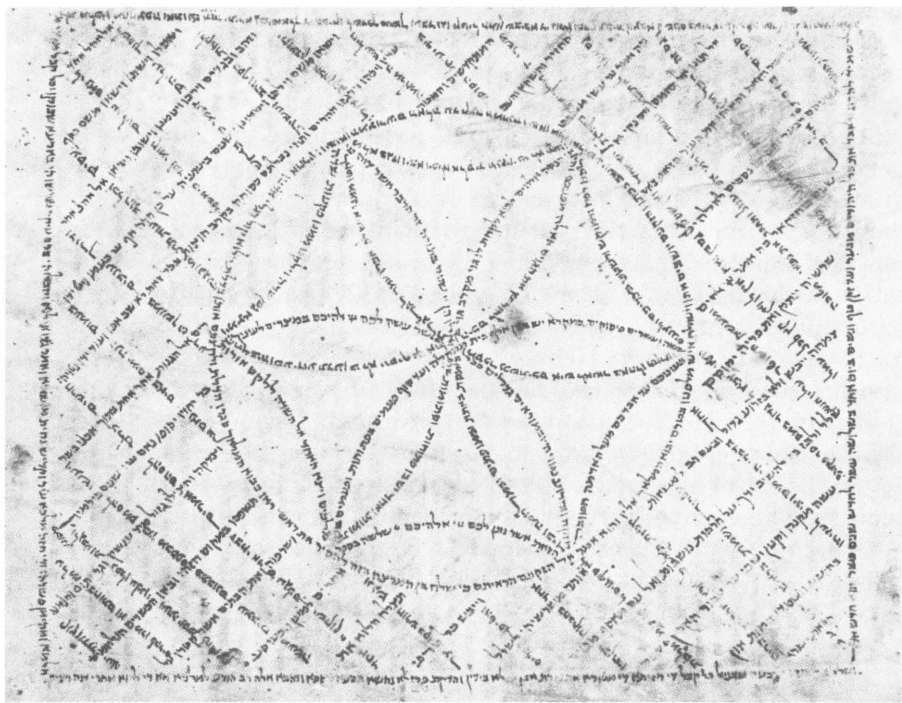


Figure 5. *Pentateuch*, Yemen, 1508.

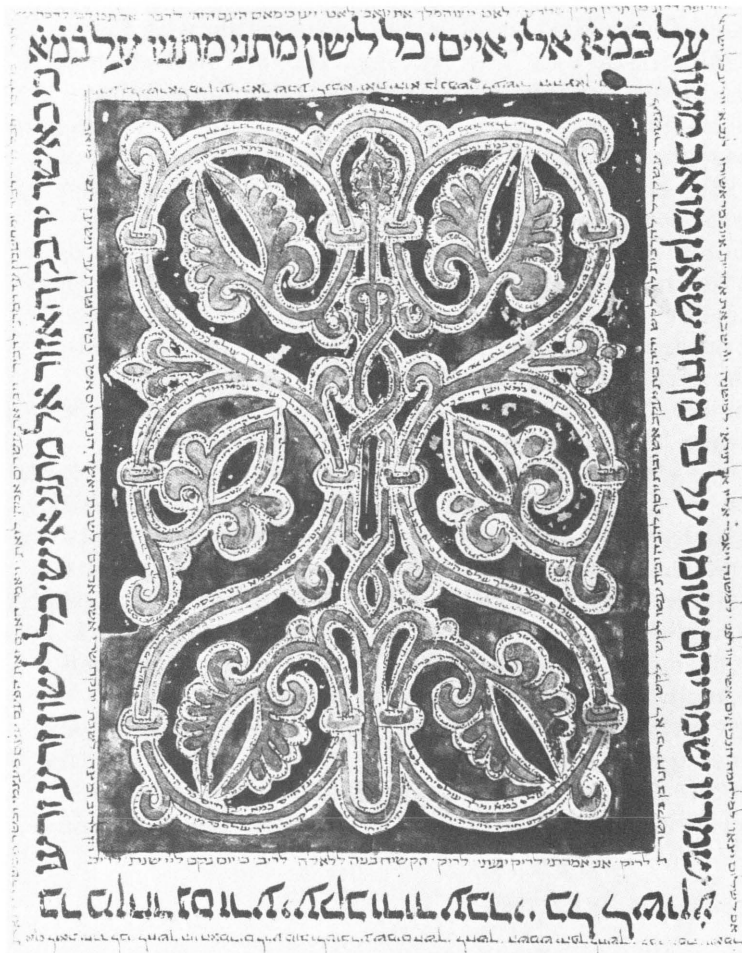
By the thirteenth century the Jewish scribes of Europe were already taking for granted the tradition of micrographic masorah in luxury Bibles. In Spain the motifs were usually, but not exclusively, non-figurative. One or several leaves of geometric carpet pages decorated the beginning and end of the codex, or were placed between biblical books (Figure 7). Full-page scenes drawn in micrography were rare, but examples do exist. Multiple-page renderings of the menorah and other vessels of the Tabernacle, a very popular iconographic theme in Sefardi illumination, are also found. Complex interlacing forms are extremely sophisticated and represent the climax of Jewish art in Spain. In marginal masorah, the scribe now drew upon an extensive repertory of geometric, vegetal, abstract, and representational forms, though the latter were not as common as in northern Europe. *Candelabra*, the "tree-of-life," and *Magen David* frequently appeared, indicating that these were regarded as specifically Jewish symbols in Medieval Spain. Occasionally the subject of the biblical text would be illustrated in micrography.

The interlacing micrograms associated with Sefardi tradition also appeared in Ashkenazi manuscripts, but there they were frequently inhabited by animals and grotesques — common marginalia in Gothic illuminated manuscripts. Although full-micrographed pages were rare, more often, the panels surrounding the initial word of a biblical book were rendered with the text of the masorah (Figure 6). At times, up to a third of the page was decorated in this fashion. Marginal masorah was also woven into a variety of animate and inanimate forms: lions, elephants, ducks, goats, horses, deer, bears, camels, keys, flags, masks, dragons, unicorns, and jousting knights. Occasionally, scribes left their names in micrography as well.

The heyday of both Ashkenazi and Sefardi micrography was the thirteenth to fifteenth centuries. There were two reasons for its eventual decline. With the invention of printing in the mid-fifteenth century, manuscript art weakened. Printers began seeking new decorative modes that could be printed with a press: first woodcuts, later copper-plate engraving, and, much later, lithography and steel engraving (nineteenth century). Hebrew metal types simply could not be set into designs that the pen could render. Secondly, in micrography's golden age, the motifs and subject matter were in harmony with the skills of the micrographer, and miniscule interlaces and grotesques suited the nib of his pen. But later, especially in the nineteenth century, when scribes turned to pictorial Bibles for inspiration, their style became as insipid as their models; few of these late micrograms can be regarded as works of art. The vitality was gone.

Nevertheless, Jewish scribes never abandoned micrography as a form of artistic expression. By the Baroque period, Italy had become the major land to demonstrate a continuous tradition. It was in *ketubot* that the art now flourished. Scribes chose various Psalms, Proverbs, and passages from Ruth and the Song of Songs, as well as blessings for the good fortune of the bride and groom based on biblical passages, as their micrographic texts to be drawn into geometric and architectural forms, flowers, family crests, and even the nude forms of Adam and Eve. Other types of micrographic illustrations on parchment or fine paper were made in Italy. *Omer* calendars used in counting the days between Passover and the beginning of Shavuot [Pentecost], *sukkah* decorations, Purim pieces, and full-page biblical scenes were intended to be displayed rather than hidden away in books.

Figure 7. *Damascus Keter*, Bible, Burgos, Spain, 1260.



Micrographic art then spread from Italy to Eastern Europe; by the eighteenth century Jewish scribes and artists in Austria and Hungary were also demonstrating their talent for this uniquely Jewish art form. By the nineteenth century there were Hebrew micrographers in England, France, Holland, Russia, and Poland, and toward the end of the century the art was brought to America and North Africa and reintroduced into Eretz Israel. At that time, too, scribes and printers began realizing the potential of the lithographic press for inexpensive reproduction and dissemination of micrography. Jerusalem's holy sites, biblical scenes, heroes and heroines, and portraits of rabbis, early Zionists, and Jewish poets, were printed in quantity and sold, sometimes on behalf of charitable institutions (Figure 2). Copyright meant little, and printers would frequently remove the name of the original artist, substituting their own in the reprint.

Today, Jewish micrography is alive and flourishing. Good examples can be found in contemporary *ketubot* in Israel and the United States; and passengers flying El Al Israel Airlines are familiar with Fred Pauker's *Tefilat Ha-Derech*, "The Traveler's Prayer," in the form of micrographic doves. Tel Aviv Pop micrographer Abraham Haba, besides writing biblical verses on sea-shells, and *megillot* and *haggadot* on eggs, has made hundreds of portraits of famous persons with texts relating to their accomplishments: he has also micrographed charming illustrations in contemporary naive style (Figure 8). Israeli artist Jacob El-Hanani has created his own variation on calligram micrography in his Constructivist style, using minute cursive Hebrew script to form a textured carpet of writing. Although the result differs from the traditional micrograph and calligram, his perseverance proves El-Hanani to be a true descendant of the Medieval masorah scribe.

Hebrew micrography's survival over eleven centuries can be understood on several levels. The biblical and rabbinic subjects of the illustrations, the nature of the text, and the endurance of its venerable tradition satisfy the aesthetic needs of even the most iconophobic Jew. But micrography can also be looked upon as a perpetual game between scribe and reader. Each micrograph challenges the reader to seek the beginning and end of the text, to identify the source of the passage, and to seek missing words or mistakes. The scribe strains the reader's vision with his microscopic letters, written in an ink that will eventually fade, melting the images as they are deciphered. But micrography must be regarded as more than a mere game; script was, and is, a true expression of Jewish culture. The alphabet and writing have long maintained a symbolic and mystical existence in Jewish folklore and philosophy. Jewish creativity has been embodied in and

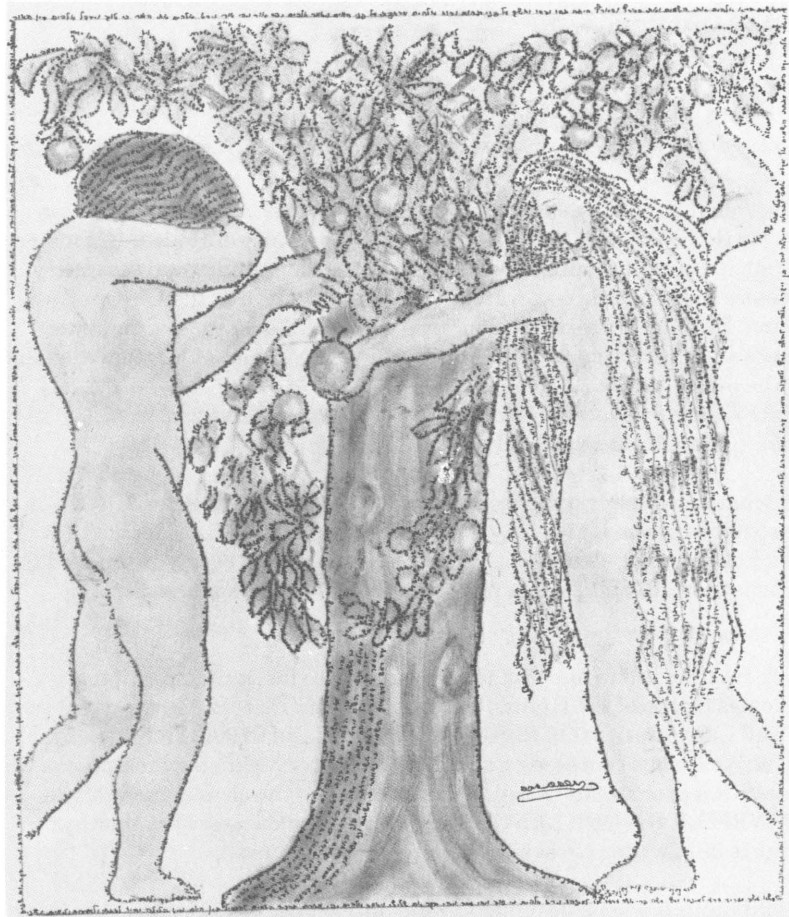


Figure 8. *Adam and Eve* by Abraham Haba, Tel Aviv, 1977.

transmitted by the oral and written word, and it is only natural that Judaism's indigenous and longest lasting art form should be bound up with writing itself.

Reprinted with kind permission from *LSA*, a publication of the College of Literature, Science, and the Arts, The University of Michigan, Ann Arbor, MI 48109. The essay originally appeared in the catalog for an exhibition at the Israel Museum in Jerusalem in 1981, for which Dr. Avrin served as guest curator. Photographs courtesy the Israel Museum. Collections: Israel Museum, Figures 1-3; Jewish National and University Library, Figures 5 & 7; Abraham Haba, Figure 8; Moldovan Family, New York, Figure 4; Kongelige Bibliotek, Copenhagen, Figure 6.

The Authors

David B. Yaden, Jr., is assistant professor in the Department of Curriculum Instruction in the College of Education at the University of Houston (Houston, TX 77004). He taught reading and English at the elementary and secondary levels before finishing his doctorate in reading education at the University of Oklahoma. Dr. Yaden's research concerns young children's metalinguistic awareness and their spontaneous inquiries into the nature of written language during the preschool years. He is co-editor of *Metalinguistic Awareness and Beginning Literacy: Conceptualizing What it Means to Read and Write*, forthcoming from Heinemann Educational Books.

Denise Schmandt-Besserat is associate professor at the Center for Middle Eastern Studies at the University of Texas (Austin, TX 78712). Her earlier research has been on paleolithic art and the earliest uses of clay in the Middle East. She is presently studying an archaic recording system based on clay tokens.

Stacey A. Keenan is a member of technical staff in the Documentation Technologies Group of the Human Performance Engineering Department at AT&T Bell Laboratories (190 River Road, Summit, NJ 07901). Her work over the last six years has centered on applying computer technology to improve the documentation process. Ms. Keenan participated in the development of the UNIX™ WRITER'S WORKBENCH™ software, which is a set of computer programs to help writers to evaluate and revise their text.

Fernand Baudin is a book designer (64 rue du Village, 5983 Bonlez, Belgium) and a recently retired teacher at NHIBS, a college of design in Antwerp. He has restyled *LeSoir* (a national daily newspaper), designed the first Stanley Morison exhibition ever (1966) at the Royal Library, Brussels, for which he also designs most catalogues, and together with John Dreyfus edited and designed *Dossier Mise en Page* and *Dossier A-Z* (Andenne 1972 & 1973). This year he published *La Typographie au tableau noir* (Paris) which won a Prix exceptionnel de la Société des Gens de Lettres de France.

Leila Avrin is a lecturer in the Graduate School of Library and Archive Studies at The Hebrew University of Jerusalem (Givat-Ram Jerusalem 91904, Israel); Dr. Avrin teaches summers at the School of Library Science at the University of Michigan. She is chairman of Israel Bibliophiles. Her research focuses on illuminated Hebrew manuscripts, particularly micrography, and medieval bindings on Hebrew books.