SOME EXPERIMENTS ON THE ROMAN AND CYRILLIC ALPHABETS OF SERBO-CROATIAN*

G. Lukatela+ and M. T. Turvey++

CROSS-LANGUAGE COMPARISONS: SERBO-CROATIAN ORTHOGRAPHIES AND THEIR SPECIAL PROPERTIES

Much if not most of current theorizing on the reading process and visual information processing is based on investigations with English language materials. Perhaps such processes vary but little across languages and orthographies and therefore a theory based on one language will suffice for all. However, what variations there are may prove to be revealing. We have been asking whether or not the reading of Serbo-Croatian may make use of different characteristics of the written word or different encoding routines than are used in the reading of English.

A distinction that is often made between logographic writing systems, such as Korean, Chinese, and Japanese kanji, and alphabetic systems, such as English and Serbo-Croatian, is that the former refer to the morphology, while the latter refer to the phonology. The logographic system is said to specify units of meaning, whereas the alphabetic system is said to specify the sounds of the spoken language, although the distinction is not as sharp. Indeed, this interpretation of the alphabet is less than ideal as far as English is concerned, for the correspondence between written and spoken English is opaque: graphemes can be made silent by context and, in general, graphemes take on different phonetic trappings in different graphemic contexts. Looking for regularity in the English orthography, Gibson, Pick, Osser, and Hammond (1962) advanced the idea of a spelling pattern, a cluster of letters that corresponds to a sound. While individual letters in English do not have invariant phonemic interpretations, certain arrangements of letters do, particularly when their locations within words are taken into consideration. Whether or not the notion of spelling pattern is valid, the point is obvious: the cipher relating script to utterance in English is complex. We argue that the cipher in Serbo-Croatian is considerably more transparent; and that for the Serbo-Croatian orthography the claim that it specifies the sounds of speech is potentially closer to the mark. But let us pursue the English orthography a little further.

---

+University of Belgrade.
++Also University of Connecticut.

Acknowledgment. This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade, and in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

203
The opaqueness of the script to utterance relation in English is owing, by and large, to two reasons. First, the pronunciation of the language evolved along different lines from the spelling of the language. Consider the following example cited by Henderson (1977). The English digraph gh as in bough and rough specified a unique guttural utterance until the seventeenth century. After the seventeenth century the pronunciation of gh took two directions: it either became silent as in night or took the phonemic interpretation /f/ as in enough. But the spelling had already become standardized largely owing to the efforts of the fifteenth century English printers, such as Caxton; and, in consequence, gh is handed down to the contemporary reader of English as an orthographic anomaly.

The second reason for the spelling-sound opaqueness is that the English orthography may be as close to the morphology as it is to the phonology. Indeed, in the evolution of the English language, Henderson (1977) has stated that the tendency has been for the orthography to reflect etymology, which is tantamount to saying that it reflects the basic units of meaning. In this vein Chomsky (1970) has argued that the English orthography is near optimal for writing the English language. The orthography preserves the morphology, which would not be the case if the optimality principles were phonemic correspondences. Thus, the spelling preserves the following morphological similarities—tele-graphy, tele-graph-ic, tele-graph-y—in the face of the obvious phonetic variability. Similarly anxious and anxiety by virtue of their visual likeness permit the reader, in principle, to go directly from the appearance of the letter sequence to its meaning. Therefore, the fundamental point made by Chomsky (1970) and also by Venezky (1970) (but for somewhat different reasons) should be noted, namely, that the English orthography is systematic in its own right. It is specific to linguistic structure at a deep level and is not to be understood just as a phonemic transcription. Indeed, on the Chomsky-Venezky view, the script-utterance relation is opaque precisely because the script and utterance are alternative specifications of the same underlying structure (cf. Francis, 1970). However, the tempering conclusion of Gleitman and Rozin's (1977) thorough analysis is that it is not so much that English orthography is optimal for this or that grain-size of linguistic analysis, but rather that English writing is a rich mixture of a number of grains of linguistic representation, together with more than a sprinkling of arbitrary features.

Let us now turn to Serbo-Croatian, Yugoslavia's major language. Serbo-Croatian, unlike English, is pronounced as it is written; that is, individual letters have phonemic interpretations that remain consistent throughout changes in the context in which they are imbedded. All written letters are pronounced; hence, in Serbo-Croatian there are no silent letters and no double letters.

This state of affairs—a straightforward regularity between script and utterance—is by virtue of a historical development that sharply contrasts the evolution of the Serbo-Croatian orthography with that of the English orthography. The modern Serbo-Croatian orthography was constructed at the beginning of the nineteenth century by Karadžić on the basis of a simple rule: "Write as you speak and read as it is written!" In Serbo-Croatian, therefore, constraints on sound sequences are the sole sources of constraints on letter sequences. This contrasts with English in which restrictions on letter
sequences derive not only from phonological constraints but also from a desire to preserve the etymology and graphemic conventions. That is, from a "...1400-year accumulation of scribal practices, printing conventions, lexicographers' selections, and occasional accident which somehow became codified as part of the present orthographic system" (Venezky & Massaro, 1979, p. 25). In English, illegal phonological sequences (such as /wh/) can be orthographically regular spellings (such as wh) but no such peculiarity is permitted in Serbo-Croatian.

Karadžić (1814) selected the speech spoken in mid-Yugoslavia as the ideal and to each phonemic segment of the speech he assigned a letter character or, in a few cases, a combination of letters. Karadžić took the majority of letters from the alphabet existing at the time but since the number of letters available was less than the number of phonemes needed, he borrowed or modified several letters from other alphabets. In fact, two alphabets were constructed: a Roman alphabet and a Cyrillic alphabet. In modern Yugoslavia, Eastern Serbo-Croatian uses primarily the Cyrillic script whereas Western Serbo-Croatian uses primarily the Roman. In some regions (e.g., Bosnia, Herzegovinia), however, both scripts are used about equally.

The Serbo-Croatian language has 30 phonemes. In the Cyrillic alphabet there is one letter for each phoneme; in the Roman, 27 phonemes are represented by single letters and three phonemes by pairs of letters: LJ, NJ, DJ. Figure 1 compares the Roman and Cyrillic alphabets in uppercase and in Table 1 the letters (both uppercase and cursive) of the alphabets are given their corresponding letter-names in the International Phonetic Alphabet (IPA) transcription.

An important fact about the Roman and Cyrillic alphabets is that they map onto the same set of phones but still comprise two sets of letters that are, with certain exceptions, mutually exclusive. Of the total set of letters comprising the two alphabets the majority are unique to one or the other alphabet (see Figure 1). A number of letters, however, are shared by the two alphabets. Of these shared letters, some receive the same phonemic interpretation whether read as Roman or Cyrillic (referred to as common letters) and some receive two phonemic interpretations, one in the Roman reading and one in the Cyrillic reading (referred to as ambiguous letters). Therefore, one may recognize instances in which letters are different in shape but pronounced the same way, e.g., the Cyrillic Ɔ and the Roman I are both pronounced like the ea in seat; instances in which letters are the same in shape and pronunciation; and instances in which the letters are of the same shape but pronounced differently, e.g., the Cyrillic Ɔ is pronounced like the n in wine, the Roman H like the ch in the Scottish rendering of loch.

Three examples underscore the unusualness of Serbo-Croatian bi-alphabetism. The sentence, This is my mother, translated into Serbo-Croatian is spelled: TO JE MOJA MAJKA. In IPA it is rendered as: [to je moja majka]. There is no way to tell whether this particular sentence is written in Roman or Cyrillic, since only the common letters have been used. The sentence, The deer climbs, translated into Serbo-Croatian is spelled in Cyrillic as: CPHA CE BEPE. In IPA it is rendered as: [arna se vere]. However, if CPHA CE BEPE were read as Roman, it would be uttered as: [tspka tse bepe], which is a meaningless utterance. Finally, one may note the sentence, The pupil studies
Figure 1. The two alphabets of the Serbo-Croatian language.
Table 1. Letters of the Serbo-Croatian alphabet.

<table>
<thead>
<tr>
<th>Serbo-Croatian</th>
<th>Roman</th>
<th>Cyrillic</th>
<th>Printed Upper Case</th>
<th>Printed Lower Case</th>
<th>Cursive Upper Case</th>
<th>Cursive Lower Case</th>
<th>Letter Name in IPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>a</td>
<td>A</td>
<td>a</td>
<td></td>
<td></td>
<td>a</td>
</tr>
<tr>
<td>B</td>
<td>b</td>
<td>b</td>
<td>B</td>
<td>b</td>
<td></td>
<td></td>
<td>b̊a</td>
</tr>
<tr>
<td>Ć</td>
<td>č</td>
<td>č</td>
<td>Ć</td>
<td>č</td>
<td></td>
<td></td>
<td>ts̊a</td>
</tr>
<tr>
<td>Đ</td>
<td>đ</td>
<td>đ</td>
<td>Đ</td>
<td>đ</td>
<td></td>
<td></td>
<td>t̊ja</td>
</tr>
<tr>
<td>Đž</td>
<td>dž</td>
<td>dž</td>
<td>Đž</td>
<td>dž</td>
<td></td>
<td></td>
<td>d̊ja</td>
</tr>
<tr>
<td>E</td>
<td>e</td>
<td>e</td>
<td>E</td>
<td>e</td>
<td></td>
<td></td>
<td>e</td>
</tr>
<tr>
<td>F</td>
<td>f</td>
<td>f</td>
<td>F</td>
<td>f</td>
<td></td>
<td></td>
<td>f̊a</td>
</tr>
<tr>
<td>G</td>
<td>g</td>
<td>g</td>
<td>G</td>
<td>g</td>
<td></td>
<td></td>
<td>g̊a</td>
</tr>
<tr>
<td>H</td>
<td>h</td>
<td>h</td>
<td>H</td>
<td>H</td>
<td></td>
<td></td>
<td>x̊a</td>
</tr>
<tr>
<td>I</td>
<td>i</td>
<td>i</td>
<td>I</td>
<td>i</td>
<td></td>
<td></td>
<td>i</td>
</tr>
<tr>
<td>J</td>
<td>j</td>
<td>j</td>
<td>J</td>
<td>j</td>
<td></td>
<td></td>
<td>j̊a</td>
</tr>
<tr>
<td>K</td>
<td>k</td>
<td>k</td>
<td>K</td>
<td>k</td>
<td></td>
<td></td>
<td>k̊a</td>
</tr>
<tr>
<td>L</td>
<td>l</td>
<td>l</td>
<td>L</td>
<td>l</td>
<td></td>
<td></td>
<td>l̊a</td>
</tr>
<tr>
<td>LJ</td>
<td>lj</td>
<td>lj</td>
<td>LJ</td>
<td>lj</td>
<td></td>
<td></td>
<td>l̊ja</td>
</tr>
<tr>
<td>M</td>
<td>m</td>
<td>m</td>
<td>M</td>
<td>m</td>
<td></td>
<td></td>
<td>m̊a</td>
</tr>
<tr>
<td>N</td>
<td>n</td>
<td>n</td>
<td>N</td>
<td>n</td>
<td></td>
<td></td>
<td>n̊e</td>
</tr>
<tr>
<td>NJ</td>
<td>nj</td>
<td>nj</td>
<td>NJ</td>
<td>nj</td>
<td></td>
<td></td>
<td>n̊ja</td>
</tr>
<tr>
<td>O</td>
<td>o</td>
<td>o</td>
<td>O</td>
<td>o</td>
<td></td>
<td></td>
<td>o̊a</td>
</tr>
<tr>
<td>P</td>
<td>p</td>
<td>p</td>
<td>P</td>
<td>p</td>
<td></td>
<td></td>
<td>p̊a</td>
</tr>
<tr>
<td>R</td>
<td>r</td>
<td>r</td>
<td>R</td>
<td>r</td>
<td></td>
<td></td>
<td>r̊a</td>
</tr>
<tr>
<td>Š</td>
<td>š</td>
<td>š</td>
<td>Š</td>
<td>š</td>
<td>š</td>
<td>š</td>
<td>s̊a</td>
</tr>
<tr>
<td>T</td>
<td>t</td>
<td>t</td>
<td>T</td>
<td>t</td>
<td></td>
<td></td>
<td>t̊a</td>
</tr>
<tr>
<td>U</td>
<td>u</td>
<td>u</td>
<td>U</td>
<td>u</td>
<td></td>
<td></td>
<td>ů</td>
</tr>
<tr>
<td>V</td>
<td>v</td>
<td>v</td>
<td>V</td>
<td>v</td>
<td></td>
<td></td>
<td>v̊a</td>
</tr>
<tr>
<td>Z</td>
<td>z</td>
<td>z</td>
<td>Z</td>
<td>z</td>
<td></td>
<td></td>
<td>z̊a</td>
</tr>
<tr>
<td>Ž</td>
<td>ž</td>
<td>ž</td>
<td>Ž</td>
<td>ž</td>
<td></td>
<td></td>
<td>j̊a</td>
</tr>
</tbody>
</table>
reading, which is written in Cyrillic ĐAK УЧИ ДА ЧИТА but in Roman as, ĐAK УЧИ ДА ЧИТА. Regardless of which alphabet has been used, the phonetic transcription is the same in both cases: [dzjak uči da čita], as is the meaning.

A most central feature is that both alphabets are taught in the schools and by most accounts the letter forms and the letter-to-sound correspondences of both alphabets are learned by the end of the second grade. The children are taught one alphabet in the first year and a half and then master the other by the end of the second year. In the western part of the nation the Roman alphabet is learned first and in the eastern part of the nation it is the Cyrillic alphabet that the children master initially. This geographically based ordering of acquisition of the two alphabets provides a model for examining the relation of two separate symbol systems, learned at different times—a bi-alphabetism if you wish—of which bilingualism is the fashionable example. It deserves reemphasizing that the two alphabets map onto the same phonemic and semantic structure.

At this juncture let us collect the preceding discussions of the phonemic regularity and the bi-alphabetism of Serbo-Croatian in order to highlight several important contrasts with English orthography. First, where it can be claimed that the English orthography more directly represents the morphology, it can be claimed that the Serbo-Croatian orthographies more directly represent the phonology. Common to the views of Chomsky and Venezky, a reader of English often needs to know more about a word than its surface orthographic structure in order to pronounce it. One would say of Serbo-Croatian that knowledge about any word's surface orthographic structure is generally all that is needed in order to pronounce it. Second, English spelling more than occasionally reveals the etymology of words but the radical reworking of the Serbo-Croatian writing system according to Karadžić's injunction ensured that the contemporary orthography would be essentially ahistorical. Third, because of the virtually invariant relation between letter and sound there are no true homophones in Serbo-Croatian. (Situations such as tale/tail, crews/cruise, wait/weight could never arise.) We emphasize true because the bi-alphabetic nature of Serbo-Croatian permits homophones of a very special kind, precisely, letter sequences that are visually quite distinct—for one is composed mainly of uniquely Cyrillic and the other of uniquely Roman letters—but which are identical in pronunciation and meaning.

It is the case, however, that Serbo-Croatian, like English, allows true homographs. It is for this reason that a reader can generally, rather than always, pronounce a word correctly on the basis of knowing only its surface orthography. Two words may be written the same way, but, owing to different assignments of vowel length and accent type, can be pronounced differently and mean different things. In Serbo-Croatian a vowel can be short or long and its accent can or can not extend into the following syllable. Sometimes these contrasts are noted by diacritical marks. More commonly, however, the ambiguity must be resolved, as in English, by sentential context. The language gives rise additionally to a special kind of homography, again made manifest over the two alphabets. Thus a given letter sequence such as POTOP can be read one way in Roman and another way in Cyrillic (see Table 1), and mean two entirely different things (respectively, inundation and rotor).
There is a further feature of the Serbo-Croatian language on which we now pass remark by way of concluding our delineation of the language's special properties. It is that inflection is the principal grammatical device in the language in contrast with English, which uses inflection for grammatical purposes only sparingly. Thus for nouns, all grammatical cases in Serbo-Croatian are formed by adding to the root form an inflectional element, namely, a suffix consisting of one syllable of the vowel or vowel-consonant type. The Serbo-Croatian nouns, pronouns, and adjectives are declined in seven cases of singular and seven cases of plural whereas verbs are conjugated by person and number in six forms.

**ERROR PATTERN IN BEGINNING READING**

Where other languages with a close match between sound and writing have been examined, the evidence is that children learned very rapidly to read aloud letter sequences congruent with the orthographic rules of the language (Elkonin, 1963; Venezky, 1973). Nevertheless, it can be noted that indifferent to the script-to-utterance correspondence reading differences emerge early (Gibson & Levin, 1975) and that some children will continue to have problems even where the spelling of the words on which they are instructed is phonetically regular and maps to sound directly (Savin, 1972). Reading skill, in the long run, appears to be largely indifferent to the language being read (Gray, 1956). A not overly venturesome claim is that different writing systems induce differences in acquisition of reading and differences in the reading process without necessarily affecting the ultimate proficiency of reading. The point to be emphasized, perhaps, is that of Carroll (1972): "A perfectly regular alphabetic system may facilitate word-recognition processes but its use does not alter the fact that the learning of reading entails the acquisition of skills in composing word units from their separate graphic components and practice, large amounts of it, in recognizing particular word units."

Given the orthographic distinction between English and Serbo-Croatian one can ask: In what ways does the beginning reader in Serbo-Croatian differ from his counterpart in English and in what ways are they the same? One can ask, in short, with respect to the acquisition of reading, what changes across orthographies and what remains invariant? We are examining this question in relation to research already conducted and currently underway at the Haskins Laboratories.

A point of departure for the reading research of the Haskins Laboratories' group is that reading is somehow parasitic on speech. One recent focus has been the notion of "linguistic awareness" (Mattingly, 1972). A child might try to read words by the medially of shape. But this nonanalytic strategy, while useful to a point, is far from optimal; the child cannot benefit from the fact that the alphabet permits its users to generate a letter string's pronunciation from the spelling. But what is required of the child to know how the alphabet works? I. Y. Liberman and Shankweiler (1979) argue that the child must realize that speech can be segmented into phonemes and he must know how many phonemes any given word in his vocabulary contains and their order. He must know that the letters of the alphabet represent
phonemes, not syllables or some other unit of speech (see also Gleitman & Rozin, 1977; Rozin & Gleitman, 1977).

The difficulty and significance of phonemic segmentation has been frequently noted (e.g., Elkonin, 1973; Gibson & Levin, 1975; Rosner & Simon, 1971); the inability to analyze syllables into phonemes marks the child who has failed to learn how to read or, at least, who reads poorly (I. Y. Liberman, Shankweiler, A. M. Liberman, Fowler, & Fischer, 1977; Savin, 1972).

Exemplary of the difficulty with phonemic segmentation is the pattern of errors a child makes in reading syllables. For simple English consonant-vowel-consonant structures the error rate on the final consonant is larger than that on the initial consonant while the error rate on the vowel is largest of all (Shankweiler & I. Y. Liberman, 1972). Moreover, the form of the vowel and consonant errors differ in nontrivial ways (I. Y. Liberman & Shankweiler, 1979). To what extent, one might ask, are these patternings of errors orthographically based? Are they indigenous to the writing system of English or would they be as likely in the orthographies of Serbo-Croatian? For example, the greater error rate on vowels might be owing to the fact that in English vowel pronunciation is extremely context conditioned. On the other hand, it might be owing to the differential status of vowels and consonants in the perception and production of speech; in which case one might treat the different error rates of vowels and consonants and the direction of the difference as indexing a universal property of phonographic writing systems.

We have begun an examination of these questions through an experiment that is closely comparable to one previously conceived and conducted by the Haskins Laboratories group.

The 65 subjects in the experiment all tested within the normal range of intelligence. They were selected from the first grade population of an elementary school system located in Belgrade. Their ages ranged from 6.5 to 7.5 years. They had completed their first semester and had an active knowledge of the Cyrillic alphabet.

We devised two lists of the CVC-type monosyllables written in Cyrillic. One hundred CVCs were words and 100 CVCs were pseudowords. The words were familiar to first graders. In the word and pseudoword lists the 25 Serbo-Croatian consonant phonemes that can occur in both the initial and in the final positions of a word appeared twice in each position. In the majority of the trigrams the medial letter was one of the five Serbo-Croatian vowels (/i/, /e/, /a/, /o/, /u/) as in ДИВ 'giant,' ЦИВ 'pipe,' ДАР 'gift,' СОК 'juice,' and БУК 'wolf.' In some trigrams, however, the medial letter was the semi-vowel /r/. In Serbo-Croatian monosyllabic words of the type consonant-semivowel /r/-consonant, as in БРХ 'top,' ТРН 'thorn,' ГРБ 'emblem,' are not infrequent. And finally, it should be noted that of the 100 words, 25 could be reversed to produce other words: For example the word БОР 'pine' if read from right to left reads РОБ 'slave.'

A string of three uppercase Cyrillic letters arranged horizontally at the center of a separate 3" x 5" white card defined a stimulus. The cards were placed face down in front of the subject and were turned over one by one by
the examiner. The subject was asked to read each letter string aloud as it was presented. Responses were written down by the examiner and were recorded simultaneously on magnetic tape. A complete list was presented in a single session with each child participating in two separate sessions. If in the first session the child read the word list, then in the second session he read the pseudoword list and vice versa. The order of presentation was balanced across children.

The responses to the stimuli revealed several types of errors: 1) substitution, 2) addition, 3) omission, and 4) reversal of sequence when a letter string or a part of it was read from left to right. Single letter orientation errors did not occur because the Cyrillic uppercase letters did not provide opportunity for reversing letter orientation.

The analysis of errors showed that sequence reversals accounted for only a small proportion of the total of misread letters, although the lists were constructed to provide ample opportunity for the complete reversal of sequences. (As noted, 25% of the words were "reversible"; and 13% of the pseudowords were words if read from right to left, for example, the pseudoword НИС would become СИН'сон').

The complete sequence reversals are distinguished from the partial and the total reversal scores for words and pseudowords are given in Table 2. Proportions of opportunity for error (in percentages) are presented within parentheses. We note that sequence reversals were rare.

Single letter omission errors were also quite rare. Their distribution on initial and final consonants and on the medial vowel/semivowel is presented in Table 3. Omissions of the final consonant in words seem to be more frequent than in pseudowords, but the respective proportions of opportunity are too small to allow any reliable conclusion on their distribution.

Additional errors were distributed in a nonrandom manner (see Table 4). Additions of a single phoneme in front of the final consonant (FC1) were more frequent than after the final consonant (FC2), other types of additions being relatively infrequent.

In words and pseudowords of the consonant-semivowel /r/-consonant type, additions of a single phoneme in front of the final consonant were relatively the most frequent. For example, the word ГРБ was often misread as /grab/, /grub/, /greb/, or /grob/. In four words (ГРБ, ВРХ, ТРГ, ТРН) there were 45 single vowel additions, and in four pseudowords (БРС, ДРН, КРП, ПРК) there were 47 single vowel additions of the FC1 type. Viewed in terms of opportunities for this particular error in the four words, the percentage amounts to 17% and in the four pseudowords up to 18%. This is a notable result. Apparently, to facilitate the phonetic rendition of the letter string, the child inserted a vowel between the medial semivowel and the final consonant.

Substitutions of single phonemes were the major source of errors in the experiment. Distribution of substitution errors on initial and final consonant and on the medial vowel/semivowel is presented in Table 5. Raw error
### Table 2

**Sequence reversals**

<table>
<thead>
<tr>
<th></th>
<th>Complete sequence reversal</th>
<th>Partial sequence reversal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Words</strong></td>
<td>17% (1.1%)</td>
<td>6 (0.0%)</td>
<td>23</td>
</tr>
<tr>
<td><strong>Pseudowords</strong></td>
<td>21 (2.5%)</td>
<td>13 (0.0%)</td>
<td>34</td>
</tr>
</tbody>
</table>

### Table 3

**Omission errors**

<table>
<thead>
<tr>
<th></th>
<th>Initial consonant</th>
<th>Medial vowel</th>
<th>Final consonant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Words</strong></td>
<td>1</td>
<td>4</td>
<td>11 (0.2%)</td>
<td>16</td>
</tr>
<tr>
<td><strong>Pseudowords</strong></td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>

### Table 4

**Additions of a single phoneme**

<table>
<thead>
<tr>
<th></th>
<th>Initial consonant</th>
<th>Medial vowel</th>
<th>Before final consonant FC1</th>
<th>After final consonant FC2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Words</strong></td>
<td>6</td>
<td>10</td>
<td>52</td>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td><strong>Pseudowords</strong></td>
<td>1</td>
<td>9</td>
<td>52</td>
<td>25</td>
<td>87</td>
</tr>
</tbody>
</table>
Table 5

<table>
<thead>
<tr>
<th></th>
<th>Initial consonant</th>
<th>Medial vowel</th>
<th>Final consonant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Words</strong></td>
<td>172 (2.6%)</td>
<td>93 (1.4%)</td>
<td>264 (4.1%)</td>
<td>529</td>
</tr>
<tr>
<td><strong>Pseudowords</strong></td>
<td>213 (3.3%)</td>
<td>113 (1.7%)</td>
<td>368 (5.7%)</td>
<td>693</td>
</tr>
</tbody>
</table>

scores and the respective percentages (within parentheses) indicate that final consonant (FC) errors exceed initial consonant (IC) errors. A Wilcoxon signed-rank test on proportions of correct responses revealed that this difference was significant ($T_{52}=252$, $p<0.001$), a result that agrees with the findings for beginning readers of English. The occurrence of phoneme substitutions on medial vowel segments was, however, less frequent than on initial ($T_{53}=273$, $p<0.001$) or final ($T_{57}=202$, $p<0.001$) consonant segments. Serbo-Croatian differs from English: consonants cause more difficulty for beginning readers than vowels. In an attempt to understand this finding one is reminded that the vowel set in Serbo-Croatian comprises only five vowels and that the Serbo-Croatian vowels are neatly distinctive in the $F_1-F_2$ plane. On the contrary, within some groups of the Serbo-Croatian consonants the distinctiveness is poor. For example, within the group of four affricates /tʃ/, /tʃj/, /dʒ/, /dʒj/ the phoneme boundaries are extremely fragile. Moreover, in some regions of Yugoslavia the native population replaces the voiced affricates /tʃ/ and /dʒ/ by their respective voiceless mates /tʃj/ and /dʒj/.

In our opinion the result of this experiment indicates that the substitution errors (both the initial consonant and final consonant) were phonetically biased. By far the more frequent errors were the substitutions within the group of the Serbo-Croatian affricates. All proportions of opportunity for substitution in Table 5 are small in comparison with the corresponding figures in the report of Shankweiler and I. Y. Liberman (1972).

A last but not the least interesting finding of this experiment is the fact that the final consonant substitution errors (see Table 5) were more frequent for pseudowords than for words. This suggests that even at an early stage of learning to read the process of decoding is sensitive to lexical content and that the child may possess both nonlexical (orthographic) and lexical routes to the phonology (Baron & Strawson, 1976; Forster & Chambers, 1973; Patterson & Marcel, 1977).
LEXICAL DECISION AND PHONOLOGICAL ANALYSIS

It is commonplace to underscore the fact that English spelling is a less
than perfect transcription of the phonology. Nevertheless, English is an
alphabet in spite of its apparent phonological capriciousness—for each
spelled English word provides strong hints as to its pronunciation. Some
students of reading (e.g., Smith, 1971), however, have felt that the hints are
so obscure, the relation between script and phonology so opaque, that the
fluent reading of English by-passes what must be the complex and arduous
process of converting the letter patterns into their related phonological
forms. The idea that the fluent reading of English may proceed without
reference to the phonology is buttressed by the claim that the English
spelling often preserves morphological relatedness, that is, similar meaning
(Chomsky, 1970). Given this claim, it is a simple step to supposing that the
fluent reading of English proceeds as one might suppose that the fluent
reading of logographic writing proceeds, that is, without a phonological
intermediary between the printed word and its meaning (e.g., Goodman, 1973).

But forceful arguments can be made and have been made by Rozin and
Gleitman (1977) to counter these denials of a phonologic strategy. Indeed, as
Rozin and Gleitman (1977) take pains to point out, the observations question­
ing a phonological mediary cut two ways and when looked at carefully add
strength to, rather than weaken, the notion of phonological involvement in the
reading of English.

It is evident from what has been said about Serbo-Croatian writing, that
neither of the two foregoing arguments against a phonological encoding is
especially compelling from the perspective of that orthography. Indeed, if an
opaque relation between script and phonology and a preserved transcription of
the morphology are advanced as reasons against phonological involvement in the
reading of English, then a transparent relation between script and phonology
and an optimal transcription of the phonology should be received as reasons
for phonological involvement in the reading of Serbo-Croatian.

At all events, this general issue of the contribution of phonological
encoding to reading is given particular expression in various laboratory
tasks. An extremely popular task is that of lexical decision, a task in which
the subject must decide as rapidly as possible whether a visually presented
letter string is a word. A finding often presented as evidence for phonologi­
cal involvement in accessing English lexical items is that rejection latencies
for nonhomophonic pseudowords are shorter than for homophonic pseudowords
(Rubenstein, Lewis, & Rubenstein, 1971). That is, it takes longer to initiate
response (say, pressing a telegraph key) to indicate "no" (it is not a word)
to a pseudoword that sounds exactly like a real word than to a pseudoword that
does not sound like any word (also Coltheart, Davelaar, Jonasson, & Besner,
1977). While, in general, lexical decision experiments support the idea of a
phonologically mediated access to English lexical items (e.g., Meyer, Schvaneveldt,
& Ruddy, 1974), other experiments that use other tasks imply no
phonological analysis or, at best, a phonological analysis that occurs
subsequent to lexical evaluation (e.g., Green & Shallice, 1976; Kleiman,
1975).
All things considered, however, the emerging orthodoxy appears to be that there is both a phonologically mediated route to the lexicon and a more direct, nonphonological route with the two modes of access relatively independent and possibly parallel in operation. As Gleitman and Rozin (1977) express it, reading probably proceeds at a number of grains of linguistic analysis simultaneously.

We wish to support the claim of phonological involvement in lexical decision. Evidence is presented that suggests that in lexical decision on Serbo-Croatian letter strings the phonological representation cannot be bypassed and that the phonological interpretation of a letter string is obligatory and automatic. Additionally, evidence is presented to show a complicity between the phonological evaluation and the lexical evaluation of letter strings that is of significance to the construction of a theory of word recognition.

Given the nature of and the relation between the two Serbo-Croatian alphabets it is possible to create a variety of types of letter strings. Thus, a letter string composed of uniquely Roman letters or of uniquely Cyrillic letters (in Figure 1) would receive single phonological interpretation and could be either a word or not a word. In contrast, a letter string composed of the common and ambiguous letters (see Figure 1) would receive two distinct phonological interpretations and could be either a word or not a word; more precisely, it could be a word in one alphabet and a pseudoword in the other or it could represent two different words, one in one alphabet and one in the other.

In a series of three experiments (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978) bi-alphabetic subjects were invited—by experimental design and by instruction—to relate to letter strings (block capitals) in the Roman alphabet mode. None of the letter strings seen by a subject were comprised of uniquely Cyrillic letters and relatively few of the letter strings were composed of common and ambiguous letters, that is to say, could even be read as Cyrillic. The conclusion on which all three experiments converged was that lexical decision to a letter string was slower when that string could be given two phonological readings (that is, could be read in either the assigned Roman alphabet mode or the nonassigned Cyrillic alphabet mode but if and only if the letter string was a word in at least one of the alphabets. Pseudowords that could be read in both alphabets were rejected no slower than pseudowords constructed from the set of letters unique to the Roman alphabet.

This result is nicely illustrated by a recent experiment in which there was no imposed alphabet bias: The adult bi-alphabetic subject (there were 48 subjects in the experiment) decides whether a string of (capital) letters is a word in the Serbo-Croatian language. In this experiment, unlike the previous ones, letter strings containing uniquely Roman letters and letter strings containing uniquely Cyrillic letters were presented. The types of letter strings (LS) examined are shown in Table 6 together with the correct lexical decision for each type. (The odd labeling of letter strings is to maintain consistency with the table of letter strings given previously in Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978; the present table is more inclusive). Table 6 is self-explanatory although it needs remarking that LS5
Table 6. Types of letter strings in the Roman and Cyrillic alphabet.

<table>
<thead>
<tr>
<th>Type of letter string (LS)</th>
<th>Lexical entry (L)</th>
<th>Phonological representation (P)</th>
<th>Symbolic representation</th>
<th>Is it a word? (in Roman or in Cyrillic)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L_R</td>
<td>L_C</td>
<td>L_R</td>
<td>L_C</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
and LS9 are composed solely from the common letters (see Figure 1) and are therefore read the same way and mean the same thing (in the case of LS5) in Roman and Cyrillic. The results of the experiment are shown in Figure 2. It is apparent from inspection of Figure 2 that lexical decision was impaired for those letter strings that could be given both Cyrillic and Roman interpretations but only if the letter string was a word. To give two of the relevant comparisons, decision times to LS4 were significantly slower than decision times to LS1 (F=11.72; df=1,26; p<0.01); decision times to LS3 were significantly slower than decision times to LS1a (F=33.4; df=1,27; p<0.001). The latter contrast is especially interesting since letter strings of type LS3 are words in both alphabets and since a general observation in the literature on English words is that letter strings with multiple meanings are accepted as words faster than letter strings with a single meaning (e.g., Jastrzembski & Stanners, 1975). Clearly, the present observation is counter to this general finding. It should also be noted that the slower decision time to LS3 was witnessed in our previous research (Lukatela, Savić, Gligorijević, Ognjenović & Turvey, 1978). Returning to the data represented by Figure 2, where the letter string was not a word, the lexical decision was not retarded by phonological bivalence: decision times to LS7 did not differ, for example, from those to LS8 (F=2.44, df=1,50).

As anticipated, these data on bi-alphabetic lexical decision permit two conclusions of some significance to an understanding of the reading of Serbo-Croatian. (We are assuming like others—for example, Coltheart et al., 1977—that lexical decision is a laboratory task well suited to investigating the nature of the information extracted from a printed word for use of lexical access.) First, the data suggest strongly that phonological encoding of Serbo-Croatian words is an automatic and extremely rapid process; as we have seen, phonological bivalence interferes with lexical decision. Second, the data suggest that it is not phonological bivalence per se that retards lexical decision, rather the necessary contingency is that the phonologically bivalent letter string being evaluated must be a word in the Serbo-Croatian language.

There are a number of theories that could be pursued by way of explaining this curious result of bi-alphabetic lexical decision. They are not pursued here for there is little to be gained at this stage by adjusting the details of this or that account of lexical decision (e.g., Coltheart et al., 1977; Meyer & Ruddy, Note 1) so as to force a fit with the present data. It suffices, perhaps, to note the Coltheart et al. (1977) concluding lament that for English there is no compelling evidence for the view that the mapping from printed word to lexical entry references the phonology. They propose that:

Unequivocal evidence for this view would be obtained by demonstrating that the phonological code for a word is sometimes used in making the "yes" response to that word in a lexical decision or categorization task; such a demonstration remains to be achieved (Coltheart et al., 1977, p. 551).

Do the present data constitute such a demonstration for Serbo-Croatian?
Figure 2. Lexical decision latencies and errors for Serbo-Croatian letter strings that are readable in only one alphabet or readable in both alphabets.
THE PROCESSING RELATION BETWEEN THE TWO SERBO-CROATIAN ALPHABETS

A question that has been pursued at some length is how the Roman and Cyrillic alphabets relate psychologically. For the reader of Serbo-Croatian the alphabets must be kept distinct at some level (or in some manner) of processing in order to circumvent the ambiguous characters as a potential source of phonetic confusion. Might we therefore speak of an alphabet mode implying perhaps that the reader can be in one mode or the other but not in both concurrently? The experiments just described bear on this question.

And how are the two alphabets memorially represented? If there are two alphabet spaces are all the letters of the Roman alphabet stored in one space and all the letters of the Cyrillic alphabet stored in the other? Or is there a region of overlap, say, the representations of the common letters? Given that the meaning of one alphabet precedes the other, how is priority in learning manifest in either the processing or the representation of the two alphabets? These questions and others guided our attempts to understand the psychological fit between the two Serbo-Croatian writing systems (Lukatela, Savić, Ognjenović, & Turvey, 1978); a part of that research is reported here.

A very simple experiment proved exceptionally instructive. Native Eastern Yugoslavians (those who learn Cyrillic first) were presented individual Roman and Cyrillic letters in random order and pressed a key as quickly as possible in answer to the question "Is this letter Cyrillic?" or to the question "Is this letter Roman?" The results are given in Figure 3. It took considerably longer to verify the common letters (see Figure 1) were Roman in the "Is this letter Roman?" condition than to verify that the common letters were Cyrillic in the "Is this letter Cyrillic?" condition. The suggestion is that the subjects of the experiment viewed the common letters as essentially members of the Cyrillic alphabet and only indirectly as members of the Roman alphabet. Arguing in like style, the ambiguous characters would appear to inhabit both alphabet spaces. The most telling observation however was this: rejecting Cyrillic letters in the Roman alphabet mode took appreciably longer than rejecting Roman letters in the Cyrillic alphabet mode.

We have come to look at these data in the following way. We reasoned that the average latency for rejecting a Cyrillic character as Roman is an index of the degree to which a description of a Cyrillic character is, on the average, similar to a description of a Roman character. In the notation of Tversky (1977) this similarity may be written as $s(c,r)$ where the perceptual representation of the target Cyrillic letter $(c)$ is the subject of the relation and where the memorial representation of an individual Roman letter $(r)$ is the referent. Similarly, the average latency for rejecting a Roman character as Cyrillic indexes $s(r,c)$. It follows, therefore, that $s(c,r) > s(r,c)$. In other words, for speakers of Serbo-Croatian who have learned the Cyrillic alphabet first, the perceptual descriptions of Cyrillic characters are, on the average, more similar to the memorial descriptions of Roman characters than the perceptual descriptions of Roman characters are, on the average, similar to the memorial descriptions of Cyrillic characters.

What is the basis for this asymmetry? By Tversky's (1977) argument asymmetric similarities such as $X$ is more similar to $Y$ than vice versa hold if and only if $Y$, the referent term, is more salient on some nontrivial dimension
Figure 3. Mean latencies and their range of variation for the alphabet decision task performed by subjects who learned the Cyrillic alphabet first.
from X, the subject term. The putative salience of (processing) the Roman alphabet may arise because the dimensions of description of the Roman alphabet include those of the Cyrillic; or that the descriptors of the Roman alphabet distinguish the Roman characters more efficiently than the descriptors of the Cyrillic alphabet distinguish Cyrillic characters. In short, the basis for the asymmetry may lie in some absolute property distinguishing the structure of the two alphabets. If true, the direction of the asymmetry should be indifferent to the order in which the alphabets are acquired. On the other hand, the basis for the asymmetry may just be the order of acquisition. To this purpose, the alphabet-decision task described above was replicated with subjects who had acquired the Roman alphabet first and the Cyrillic alphabet second. The results are shown in Figure 4. They reveal that under the two question regimes ("Is this letter Roman?"; "Is this letter Cyrillic?") these subjects behaved differently, as did the subjects in the first experiment. But most importantly the behavior of the subjects indigenous to Western Yugoslavia was diametrically opposite to that of the subjects indigenous of Eastern Yugoslavia (compare Figure 4 with Figure 3). By the same reasoning as outlined above we conclude, for subjects who learned the Roman alphabet first, that $s(r,c) > s(c,r)$. That is, for Roman-first subjects, processing Roman letters is more similar to processing Cyrillic letters than vice versa. More generally we conclude that the alphabet-processing asymmetry is owing not to a fixed structural property of the alphabets but to their order of acquisition. One tentative conclusion to be drawn is that the procedure developed by the child to decode the letters of the first acquired alphabet is modified for the second acquired alphabet so that decoding the second acquired alphabet necessarily entails the procedure for decoding the first acquired alphabet but not vice versa.

But perhaps the more outstanding, although equally tentative, conclusion to be drawn is that the order in which the alphabets are acquired, and the concomitant early bias in reading toward one of the alphabets, leaves a profound impression on the letter decoding processes of adult readers of Serbo-Croatian. This conclusion is not unrelated to some results recently published by Jackson and McClelland (1979). In the view of some students of reading (e.g. Kolers, 1969; Smith, 1971) individual differences in the reading ability of experienced readers are solely differences in comprehension ability. The research of Jackson and McClelland brings this view into question by showing individual differences in the ability of American college student readers to access letter codes, an ability that accounts for a significant portion of the variance in effective reading speed. What has been noted with mature Serbo-Croatian readers is that in the alphabet decision task there is an interaction between the alphabet first learned and the alphabet being decided upon. The pattern of decision times for Roman-first subjects is, on the significant contrasts, a mirror image of the pattern for the Cyrillic-first subjects. What is surprising about this interaction is that the subjects have been reading in the two alphabets for between 12 and 16 years and yet on a simple decision task the alphabet learned first makes its mark. The point on which our data and those of Jackson and McClelland would appear to converge is that the basic encoding processes by which letters of the alphabet are distinguished and named are not necessarily asymptotic in mature readers; nor is mature reading indifferent, perhaps, to the manner of their acquisition.
Figure 4. Mean latencies and their range of variation for the alphabet decision task performed by subjects who learned the Roman alphabet first.
REFERENCE NOTE


REFERENCES


Kolers, P. A. Reading is only incidentally visual. In K. S. Goodman & J. Fleming (Eds.), Psycholinguistics and the teaching of reading. Newark, Del.: International Reading Association, 1969.


FOOTNOTES

1In a subsequent analysis of these data (see Lukatela, Popadić, Ognjenović, & Turvey, this volume), the detriment to performance incurred by phonologically bivalent letter strings occurred both for words and pseudowords.