Research on the Variety of Scripts

Insup Taylor

Follow this and additional works at: https://scholarsarchive.byu.edu/dlls

BYU ScholarsArchive Citation
Available at: https://scholarsarchive.byu.edu/dlls/vol12/iss1/3

This Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Deseret Language and Linguistic Society Symposium by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Research on the Variety of Scripts

Insup Taylor
Life Sciences, Scarborough Campus
University of Toronto, Canada

In my keynote address I described four writing systems and ten scripts that are used today in different parts of the world. Assuming that you have read the written version of the address (Taylor, 1986), for this seminar I will discuss examples of experiments carried out on a few of these scripts, notably English (alphabet), Chinese (logography), Japanese (logography and syllabary), and Korean (alphabetic syllabary). These examples are intended to delineate some methods by which problems can be attacked, rather than to review the research in the field.

I limit my coverage to three research problems common to these writing systems and scripts and a few problems specific to particular systems and scripts.

Research Common to Different Scripts

The three common problems discussed are phonetic recoding, eye movements, and cortical processing.

Phonetic Recoding

In reading a sentence silently for comprehension, readers may store its individual words in working memory and integrate them to extract the meaning of the sentence as a whole. Such storing and integrating appear to be carried out in working memory in a phonetic code, as revealed in experiments.

The above idea was first suggested by Conrad (1964), who presented a string of letters to subjects for immediate recall. A letter string visually presented might look like:

B I K T C V R

When recall errors occurred, they were similar to the stimuli in sound, rather than in shape. Thus, the letter T ("tee") was recalled by error as P ("pee"), and not as F.

In later research, sentences have been constructed with words that are phonetically either similar or dissimilar. In one such experiment, subjects were asked to detect order anomaly in visually presented sentences, such as:

"1. Crude rude chewed Jude stewed food"
"2. Dark-skinned ate Ian boiled meat"

The subjects found it easier to detect anomalies in sentence 2 with its varying phonetic forms than in sentence 1 with its similar, and hence confusing, sounds (Baddeley & Lewis, 1981).

A similar study can be done with Chinese sentences. The subjects were asked to judge whether or not sentences such as following were grammatical and meaningful:
The two sentences are similar in meaning: The first means 'The stupid husband and wife chop up tree', and the second means 'The stupid husband and wife pick flowers'. The first sentence has characters with similar sounds, while the second has characters with dissimilar sounds. The subjects found it easier to judge the second sentence than the first (Tzeng, Huang, & Wang, 1977).

So, sentences containing phonetically similar words are more difficult to process than those containing phonetically dissimilar words. This event occurs only because the words of a sentence, whether in a phonetic or logographic script, are processed using a phonetic code.

Why, for linguistic materials, is a phonetic code preferred to a visual code, even when they are presented visually? The phonetic code seems to be primary in the following senses. Historically, humans have had oral speech since time immemorial, whereas they have had writing for only about 4,500 years. Even today, all human races and tribes have oral languages but only some of them have written ones. No one learns to read and write before learning to speak and listen to speech. Even after they learn to read, most people spend more time speaking than reading. (For more about phonetic recording, see Taylor & Taylor, 1983, chap. 10.)

To convince yourself that phonetically similar words are indeed confusing and difficult to remember, read and then recall the following “story,” which is the first lesson of the primer prepared by the renowned American linguist Bloomfield (Bloomfield & Barnhart, 1961).

Nan can fan Dan.
Can Dan fan Nan?
Dan can fan Nan.
Nan, fan Dan.
Dan, fan Nan.

**Eye Movements in Reading**

As one reads, a saccadic jump brings a target word into the fovea, where acuity is sharpest. The eyes then fixate on the word for about a quarter of a second, during which time the image of the object is more or less stationary upon the retina. It is mainly during the fixation that a reader acquires information on the fixated word. At the end of the fixation, the eyes saccades to the next target word. (For more about eye movements, see Taylor & Taylor, 1983, chap. 7).

The readers allot their attention differentially over text, giving more to content words and less to function words. Differential processing is aided if grammatical morphemes are visually distinguishable from content morphemes by being short (e.g., English function words), by being written in simple phonetic scripts (Japanese and South Korean), by being followed by a space (in all-Hangul Korean), and so on. These visual features of grammatical morphemes must be noticeable in peripheral vision so as to guide the
eye's saccades.

Some grammatical morphemes are less important, and hence more likely to be skipped, than are others. For example, in English the articles (a, the) are "good" or prototypical function words (short, frequent, have little semantic and syntactic function, etc.), whereas long infrequent prepositions (between, despite) have substantial content as well as syntactic function (Taylor, in press).

The following sentence is a part of a paragraph; the numbers over some words are gaze durations, which are summed durations of consecutive fixations on the same word by an individual subject. The gaze durations tend to be longer on content words, and are either shorter or non-existent on function words (Just & Carpenter, 1980).

1586 267 40 80 267 617, 467 450 450
"Flywheels are one of the oldest mechanical devices known to man."

For guiding the reader's eye, the ideal may be a mixed-script text in which informative words are in a visually prominent script, whereas less informative words are in a less prominent script. Precisely this kind of mixed-script text is used in Japan and South Korea: Visually complex Kanji are used for key content words and simple phonetic signs are used for grammatical morphemes, as shown in Fig. 3 of Taylor, 1986.

Sakamoto reports that mixed text is read twice as fast as all-Hiragana text (in Sakamoto & Makita, 1973). A new study is required to show that (1) Japanese people, even those without an established habit of reading in mixed text, read faster mixed than all-Hiragana text, and (2) the readers indeed fixate longer on content words in Kanji and shorter on grammatical morphemes in Hiragana.

The Chinese language has a handful of particles that indicate plural number of people, completed action, and so on (Chao, 1968; Kratochvil, 1968). Each particle is written in one Chinese character, just as is a content morpheme. Thus, particles are not particularly distinguishable visually from content morphemes. Perhaps partly for this reason, Chinese readers averaged 10 saccades per line (frequent fixations), compared with English readers who averaged 4 saccades (Stern, 1978).

If characters for particles are simplified more drastically than those for content morphemes, a Chinese reader might be able to notice them in peripheral vision, and thus skip them.

Cortical Processing

The human cortex is composed of two hemispheres, left (LH) and right (RH), connected by the corpus callosum. The two have different but complementary processing modes. Fig. 1 caricatures the processing preferences of the two. It also shows how the left visual field (LVF) from each eye projects to the RH and the right visual field (RVF) projects to the LH. The corpus callosum transfers information between the hemispheres so that much, though not all, of what is received by one becomes available to the other.

The preferred modes of the LH and RH are conventionally accepted in the literature on physiological and behavioral measures of people with intact brains, people with damage in either LH or RH, and people with split brain. A portion of the literature on normal people is presented below; the literature on damaged or split brains is presented elsewhere (Taylor, in press). (For damaged brains, see, for example, Ogden, 1984; Paradis, Hagiwara, & Hildebrandt, 1985; for split brains, see Gazzaniga, 1983; Zaidel, 1978, 1983;...
Which hemisphere processes Chinese characters? The answer depends on experimental procedures. In Japan, Hatta (1978) prepared three types of materials: (a) single Kanji, (b) 2-3 Hiragana for the sounds of the Kanji, and (c) 2-Kanji words that include the single Kanji as component.

Each stimulus was presented in a tachistoscope either to LVF or RVF of right handed males and females. There was LVF advantage for single Kanji, RVF advantage for multi-Kanji words, and again RVF advantage for Hiragana words. In short, single Kanji were processed better by the RH, Hiragana by the LH, and multi-Kanji by the LH.

Similar results were obtained with Chinese speakers. In one experiment, subjects verbally identified single characters shown in a tachistoscope (Tzeng, Hung, Cotton, & Wang, 1979). A strong LVF advantage (RH processing) was obtained, regardless of whether the characters contained a “phonetic component” (Table 3, Taylor, 1986). When stimuli were multi-character words the opposite result—namely, evidence of LH processing—was obtained. In another experiment, Chinese speakers decided, by pressing a key, yes or no to the question, “Do characters form a meaningful word?” In this task that did not involve verbal identification, evidence of LH processing was again obtained.

A single character or Kanji is processed as a whole visual pattern by the RH whereas a two-character/Kanji word is processed as a sequential object by the LH.

One of the most complex studies on hemispheric processing of a variety of symbols was carried out by Nishikawa and Ninna (1981). Japanese (and French) subjects had to respond yes or no, by pressing a key, to the question, “Are all the symbols the same/different in name/shape?” The symbols were alphabetic letters, upper- or lowercase; Hiragana or Katakana; and Kanji. All the stimuli were either upright or inverted, containing from 2 to 5 items. Table 1 shows the sample stimuli and responses on “same” decision. Fig. 2 shows the results for the “same” decision; the results for the “different” decision are similar but slightly slower.
Table 1. Types of Symbols and Responses

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Type</th>
<th>Shape</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>R R R R R</td>
<td>alphabet</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>R r R R R</td>
<td>case mix</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>あ へ い え</td>
<td>Hiragana</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>っ へ い え</td>
<td>Hira-Katakana</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>み し す ぬ</td>
<td>Kanji</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>げ げ げ げ</td>
<td>(inverted letter)</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

"Same" responses

![Graph showing time to make "same" decision on the variety of symbols](image)

Figure 2 — Time to make "same" decision on the variety of symbols

(Based on Table 1 of Nishikawa & Nlina, 1981)

First, look at the upright phonetic symbols (Kana and alphabet) in the left panel. Whether deciding on name or shape, reaction time is faster with LH presentation, and the time increases for added items, suggesting sequential processing. Decision is faster on shape (visual processing) than name (phonetic processing). Decision is faster on alphabet symbols than Kana, perhaps because the alphabet contains 26 choices whereas Kana includes 46 (or 71 including the secondary Kana; see "Kana," Taylor, 1986.)

Now look at Kanji and inverted symbols in the right panel. Reaction time is faster with RH presentation, and the time does not increase for added items, suggesting simultaneous processing. Simultaneous processing is faster than sequential processing. Reaction time is faster to inverted symbols than to their corresponding upright ones, perhaps
because the former are processed purely as visual objects.

Physiological measurements (EEG, blood flow, etc.) of normal people reading alphabetic text show that both RH and LH are involved (Ornstein, Herron, Johnstone, & Swencionsis, 1976; Lassen, Ingvar, & Skinhoj, 1978). Readers of other types of scripts may be expected to show a similar pattern of cortical processing. Recall that a single character/Kanji is better processed by the RH, but a word with two character/Kanji is better processed by the LH. And Japanese and Korean grammatical morphemes tend to be written in phonetic scripts, which are processed better by the LH. Text is a sequence of words and grammatical morphemes.

**Research Unique to Specific Scripts**

Some researchers investigate problems that are peculiar to particular writing systems or scripts.

**English orthography**

With English orthography, one can investigate the effects of a low letter-sound correspondence on pronouncing and spelling. One study on pronouncing (Baron & Strawson, 1976) and one study on spelling (Marsh, et al., 1980) have already been presented ("Learning to read in alphabet," Taylor, 1986).

Apparently, how well schoolchildren read aloud irregular words is a good index of their reading abilities. Adams and Huggins (1985) prepared a list of 50 irregular words, blocked in five levels of frequency. Examples of test words from most to least frequent were: ocean, whom, sweat, trough, fiance. All the words were within the children's listening vocabulary. The children were divided into two groups, above average (good readers) and below average (poor readers), based on WISC IQ tests, Stanford and Gates-McGinnite reading comprehension tests.

The number of correctly read words varied directly and strongly with reading ability, as shown in Fig. 3.

![Figure 3](image_url)

**Figure 3** — Mean percentage correct for good and poor readers from Grades 2 and 5

(Adams & Huggins, 1985, Fig. 2; with the authors' permission)

With sentence context preceding each test word, all the children read the test words better, but the pattern of differences between good and poor readers remained. In Fig. 3, percent correct decreases as frequency of words becomes lower. The apparent effect
of frequency may be partly due to the fatigue and interference from the earlier words that accumulate as more and more words in the list are read. The children read the words always in the same order.

Even so, one can surmise that the presence of many irregular words is an obstacle to full mastery of word decoding by Grade 5.

**Japanese Kanji and Kana**

Are complex Kanji more difficult to process than simple Kanji? This important question needs to be answered before a large number of characters are drastically simplified as in the Republic of China (Table 2, Taylor, 1986). In Japan, Kawai (1966) prepared a set each of simple Kanji and of complex Kanji. Although Kawai used his own index of complexity, his simple Kanji appear to have around 5 strokes while complex Kanji around 15 strokes. The two sets of Kanji varied in seven levels of frequency. Table 2 shows a sample of stimuli, as well as the results to be described shortly.

<table>
<thead>
<tr>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hi frequency Error</td>
<td>王 1.25</td>
</tr>
<tr>
<td>Low freq. Error</td>
<td>八 29.0</td>
</tr>
</tbody>
</table>

(based on Tables 4 and 5 of Kawai, 1966)

Adult Japanese were asked to give meaning and “Chinese and Japanese pronunciations” of the individual Kanji (see “Kanji In Japan and Korea,” Taylor, 1986). Errors were fewer on complex than on simple Kanji, and fewer also on high frequency than low frequency Kanji, as can be seen in Table 2. With nonsense configurations of lines, too, Kawai found that more complex figures were learned better than simpler ones (see also Fukuzawa, 1968, who tested children with similar results).

If a researcher were to carry out a new experiment on the same question, she or he should vary complexity in several levels, and measure not only error rates but also latencies to pronunciation and meaning access. The researcher should ask also how the six principles of Kanji formation (Table 3, Taylor, 1986) affect Kanji processing.

Another Japanese experiment compared recognizability of a word in three scripts—Katakana, Hiragana, and Kanji (Tanaka, 1977). A two-Kana/Kanji target word had to be recognized among a series of two-Kana/Kanji words. Recognition scores were higher for Hiragana for subjects aged less than 11, after which they were higher for Kanji. As children progress in grades, they learn more Kanji and use them more frequently. Kanji words, when they are learned well, are recognized better than Kana words.

**Korean Hangul**

In Hangul, each symbol codes a phoneme, as in an alphabet, but between two and four alphabetic symbols are packaged into a syllable-block, which is the unit of printing and reading.
One research issue is the relative efficiency of linear vs. packaged arrangements of the symbols in syllable-blocks. To study this issue, Taylor (1980) taught four CVCV (Set I) and four CVCCVC (Set II) Hangul words to English-speaking subjects, who learned to read them fluently in 5 min. The subjects initially read the words faster in a linear than packaged arrangement. But over 18 trials, which gave them about 80 min experience, the differences between the two arrangements gradually narrowed and then almost disappeared, as shown in Fig 4.

![Figure 4 — Pronunciation time for Sets I and II in packaged and linear arrangements (Taylor, 1980)](image)

A study with subjects without an established habit of reading in a linear arrangement, run over many trials, might show a clear superiority of the packaged arrangement over the linear one. Also, one might expect packaged arrangements to improve discrimination of longer sequences more than of shorter ones.

Another question is, "For discrimination and recognition of syllable-blocks, are three levels of complexity better than only one level?" To answer this question, Taylor (1980) used the same technique as Tanaka (1977, above), except that stimuli were Hangul syllable-blocks, and that there were two experimental conditions. In a "uni-level" condition, a target and a series of syllable-blocks among which the target had to be recognized were all of the same level of complexity, either CV or CVC. In a "mixed-level" condition, the series contained syllable-blocks in all three levels of complexity.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Target</th>
<th>&quot;Circle each occurrence of target&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uni-level</td>
<td>다 오 드 마 로 다 세 모 다 ...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>온 옥 들 막 온 닫 섭 문 닫 ...</td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>다 온 옥 들 마 온 닫 섭 문 닫 ...</td>
<td></td>
</tr>
</tbody>
</table>

More targets were recognized in the mixed-level than uni-level condition.

**Conclusion**

I have presented examples of research common to several different scripts, as well as of research unique to specific scripts. Linguistic, cognitive, and cortical processing seems to be similar for people reading sentences and text in different scripts. Processing seems to be dissimilar for people recognizing individual symbols in different scripts.
References

Adams and Huggins (1985) see the next page
Gazzaniga, 1983 (see next page)
Ogden, 1984 (see next page)
Paradis, Hagiwara, & Hildebrandt, 1985 (see below)


Zaidel, 1983 (see below)


