WHY JOHNNY CAN'T READ CHINESE*

John DeFrancis, Seton Hall University

In recent years, most notably at the 1964 convention of the Association for Asian Studies in Washington, D.C., the teaching of Chinese has been the subject of intensive discussions. The general tenor of these discussions can be grasped from the following summary remarks made by Professor Frederick W. Mote at the Washington meeting:1

The panel discussion on the teaching of Chinese and Japanese . . . indicates that the general mood of the profession at this moment is one of despair and frustration. A large number of eminent scholars, both those on the panel and members of the audience, indicated extreme dissatisfaction with the present state of language teaching and gloom about the future.

In point of fact the dissatisfaction noted by Professor Mote does not apply uniformly to the whole field of Chinese language teaching but is more pronounced in some areas than in others. By and large things are not too bad in the teaching of spoken Chinese. There exist some adequate and perhaps even excellent textbooks and programs in this area. The situation is quite otherwise as regards reading, the teaching of which is almost universally considered unsatisfactory and is actually far worse than generally recognized.

The reason for this disparity between speaking and reading, though not commented upon directly by Professor Mote, is illuminated by the following remarks which also throw some light on how we might go about seeking a remedy for the situation:2

. . . it is apparent that there is remarkably little activity in the field of Chinese linguistic research today, and that the field, brilliantly and actively engaged in the war effort twenty years ago, now rests largely on the remarkable advances made at that time. Should we not now turn to the tasks of research and experimentation with new energy and initiative and refuse to accept the obviously discouraging present state of affairs as inevitable?

I hope in this paper, which embodies the results of a number of years of research and experimentation, to make some contribution toward dispelling the aforementioned gloom by pinpointing a few basic shortcomings in our reading programs and presenting concrete suggestions for overcoming them.

---

*Paper presented at the annual meeting of CLTA in Chicago, December 29, 1965, as part of a panel on Standards in Written Chinese.
Apart from the intrinsically greater difficulty of written as against spoken Chinese, it is my contention that the failure of reading programs to match the progress achieved in teaching spoken Chinese since World War II is due primarily to two basic errors. These mistakes permeate every aspect of our work. They have dominated every Chinese program, conditioned the design of our textbooks, determined the standards for various levels of achievement, and generally barred the way to the thorough overhauling required in our approach to written Chinese.

Simply stated, the errors consist in making too rigid a connection between reading and speaking and in placing too much emphasis on learning characters.

As everyone knows, during World War II the old system of teaching Chinese, which minimized conversation and utilized the traditional writing system as the medium of instruction, was replaced by a new method which gave priority to conversation and used a transcription system—Wade-Giles or Y. R. Chao's National Language Romanization or an entirely new system, George A. Kennedy's Yale Romanization. The swing of the pendulum went so far that the audio-lingual method has even come to dominate the teaching of reading.

The way in which this manifests itself can be seen by examining the programs of the two main approaches to the teaching of Chinese, which for purposes of convenience I shall label the Chao approach and the IFEL approach, i.e., that of the Institute of Far Eastern Languages at Yale. The Chao program is so designed that the student has an option as to when he initiates character study. He can if he so chooses first study a lesson in the transcription text, Y. R. Chao's Mandarin Primer, and then immediately thereafter take up the identical material in character form. Or he can study a number of lessons (at Princeton I believe it is nine) in the transcription text before he takes up Lesson One in the character version. In any case in the initial stage of his studies he learns in character form no more nor no less than is available in transcription form.

In the IFEL program there is a mandatory time lag in initiating the study of characters. Students finish half of the basic conversation text and only then start on the beginning reading text. The latter does not go beyond what the student has already learned in transcription form, and indeed does not even present all the material available in the conversation text.

The relationship between speaking and reading in the beginning texts of the two programs can be represented graphically as follows:

**Chao approach**

- Speaking and reading

**IFEL approach**

- Speaking only
- Speaking and reading

As regards the difference in time in initiating the study of characters, something can be said for both approaches. A delay in starting character study assures that the student receives the benefits of the audio-lingual method without the distraction and possible harm of too early immersion in character study. Yet it is possible, though more difficult, to start character study at the very beginning of the learning process without undue interference with the aim of giving the student a foundation in the spoken language.

Indeed the question needs to be raised as to whether the very aim of providing a foundation in the spoken language is itself desirable. Supporters of the audio-lingual approach not only claim that theirs is the best method of producing readers but often seem to imply that it is the only one. Yet some very able workers in the Chinese field, including such a master of written Chinese as Arthur Waley, have achieved high levels of competence by other means.

Scholars at the University of Chicago make the criticism, in which I fully concur, that the audio-lingual approach has not demonstrated its vaunted superiority over other methods in the field of reading. They therefore advocate starting Chinese not with conversational Chinese but with written Chinese, and with the wen-yen style to boot.

My own opinion is that, given the requisite ability and drive, it is possible to learn to read Chinese by virtually any method, including the method which is no method. In special circumstances with particular individuals one approach may prove more practical than another. Apart from these special cases, however, for most individuals, and in programs where a choice of methods is possible, I believe that reading, whether in the pai-hua or wen-yen styles, can potentially be taught most efficiently if it is based on the use of the audio-lingual method.

My criticism is not that this method is used, but that it is abused. I think it a mis-application of the method that in both of the main approaches to reading the content of the reading program is so rigidly determined by the content of the conversation program.

The fact of the matter is that in all languages, whether it be the language we command as native speakers or the language we acquire as foreign learners, the areas of speech and of reading do not coincide. While a completely literate native speaker is able to read anything he can say and is also able to say anything he can read, in actual practice he speaks more than he reads about some things (e.g. everyday matters such as the state of one's health), and reads more than he speaks about other things, and when he does both read and speak about the same matters he often uses a different style. I don't believe even lawyers speak the language of written legalese.

The differential between reading and speaking is even greater when these are done in a foreign language, and perhaps greatest of all if the foreign language is Chinese, in which the gap between spoken and written forms is particularly wide. The Western student of Chinese may have occasion to talk about his illnesses and injuries, but he can succeed in handling these matters without recourse to reading and writing. And no Western student of the language has ever carried on a conversation in the language of Mencius—or, if we examine
closely, even in the language of Lu Hsün. In a language written in an alphabetic script the amount of work required to master both the written and spoken forms is much less than in the case of Chinese. Here the amount of work is so great that a rigid economy of effort is called for.

What these comments add up to is the recommendation that the contents of our Chinese programs should from the very beginning be divided into three parts: one, for speaking only; another, for reading only; and a third, for both speaking and reading. Graphically we can represent this as follows:

```
  speaking only
  speaking and reading
   reading only
```

The chart is intended to show only the general relationship and not the absolute magnitude of the three parts of the total program. Furthermore, the areas in which the parts are identical and in which they differ should be carefully noted. The common ground consists chiefly of the basic structural patterns and the basic vocabulary common to both written and spoken Chinese. The area reserved exclusively for speaking consists chiefly of vocabulary which is not likely to be encountered in the kind of written Chinese that interests most of our students. In both these areas the goal is an active command by the student. The area reserved exclusively for reading, in which the goal is a passive command, consists chiefly of vocabulary which the student can do without and yet be able to converse about essential matters. Ideally, the whole program should be devised in such a way that the student has the option of achieving both a speaking and reading mastery of all three parts if he so desires, and if he does not so desire will nevertheless not be forced, as he is at present, to read some comparatively useless things and to postpone too long the reading of more useful material.

The achievement of these goals will be impossible unless we are able to rectify the second major error in our work, namely the distorted emphasis on learning characters. This error manifests itself in somewhat different ways in the Chao and IFEL approaches.

The Chao approach to reading is characterized by the use of largely unstructured material. Its exponents seem to disdain the idea of making things easy and are almost self-consciously hard on their students. At Harvard after a year of half-time study students are expected to be familiar with 2000 char-
acters and to have an active command of 90 per cent of the 1372 characters appearing in Mandarin Primer. The transcription version of this text sets an enviable standard of linguistic sophistication and has many outstanding qualities as a conversation text. The character version, however, is an excellent illustration of the fact that what makes for a good conversation text does not necessarily make for a good reading text. As a reading text the character version suffers from the double fault of being too closely tied to the conversation text and of being unnecessarily difficult and inefficient. Teaching reading on the basis of the philosophy underlying such material is a little like teaching swimming by throwing a novice into the middle of a river. Those who survive this pedagogical technique can doubtless claim by this fact alone to have attained an acceptable standard of achievement. In view of the frequently excellent results achieved in programs following Chao's approach, one hesitates to suggest that even superior students might attain these results more quickly and efficiently by less draconian methods.

The second approach to reading, which was pioneered by the late Professor George A. Kennedy, is characterized by the controlled introduction of characters selected with regard to their frequency of use in general Chinese publications. The significance of frequency was expressed by Kennedy in the following terms:

The difficulty of memorizing a Chinese ideograph, as compared with the difficulty of learning a new word in a European language, is such that a rigid economy of mental effort is imperative. The student cannot afford to load his mind with useless freight, and, in the early stages of Chinese study at least, an ideograph which recurs no oftener than once in ten thousand must be definitely labelled useless. In the selection of an efficient vocabulary, imagination is a treacherous guide. All too often, language books for beginners start the student off with words like "dog, cat, chair, table, coffee, sugar," in the plausible belief that these are among the most common words. But the fact that an object is common does not imply that the word for it is common in literature. Objects may become so common that we do not write about them. Where economy of effort is desired, therefore, a more scientific standard of selection must be followed.

The idea advanced by Kennedy and partially exploited in IFEL publications that learning to read can be made easier and more efficient by studying commonly used characters first and rarer characters later, if at all, is rejected by the Chao school. It is not disturbed by the fact that the beginning text contains such rarities as the character 大 (in the term for 'knee cap'), a character which frequency counts indicate is not likely to occur even once in a million characters of running text.

While rejecting a structured approach to character study, the Chao program, as in the case of most programs, utilizes a structured approach to the study of phonology. Now outside of a classroom no one ever says such artificial sequences as mā, má, mū, mā. Yet most of us use, with clear benefit to
the student, special pronunciation drills of the same general form as the excellent, carefully planned drills in Mandarin Primer. Structured character materials based on the ordered introduction of certain graphs and the elimination or postponement of others are of course less natural than, say, selections from Hu Shih or Lu Hsun—or Mandarin Primer. The main point is, however, that characters can be presented in an ordered sequence not only without harm to the student’s ability to handle unedited Chinese but with distinct benefit in this direction. This is a question of methodology on which there is an irrefutable difference of opinion between many teachers of Chinese, including supporters of the IFEL approach, on the one hand, and those who follow the Chao approach, on the other.

Important though this difference between the Chao and IFEL schools is, of even greater significance is the emphasis common to both, and indeed common to the teaching of Chinese in general, on learning as many characters as possible as quickly as possible. The number of characters learned is universally taken as the measure of achievement toward the goal of learning to read. The ubiquitous question is: “How many characters do you know?”

This distorted emphasis on characters is an error which constitutes the main obstacle to any significant advance in the teaching of reading. A full discussion of the error requires some examination into the reasons why it has come about.

One important reason is a quite simple one—the countability of Chinese characters. No other feature of Chinese is so readily subject to measurement. There is no agreement as to what constitutes a word or a structure in Chinese, but a novice can point to a character and can come up with a nice round figure for the number of graphs he knows, that is the number of graphs contained in the textbook he has studied. It gives a satisfying feeling of exactitude to be able to say that student A knows 1000 characters and student B knows 800 and hence A has learned 25 per cent more than B.

Another reason is a simplistic concept of language. It is a common misconception to exaggerate the role of vocabulary relative to other linguistic features. The misconception is not confined to the area of Chinese. It is conspicuously present in the arguments for Basic English. The prevailing approach to the teaching of Chinese, which seeks to lay a foundation of what is in effect a sort of Basic Chinese, shares some of the fallacies of Basic English. It will be instructive to expand on this point.

The very concept of vocabulary is oversimplified. A character in Chinese and a word in English is something with white space around it on the printed page. As thus defined the units are easy to count. Basic English proudly boasts that its 850 words can be printed on a single page. One of the words, one of the 200 ‘picturables,’ is the word ‘chess.’ The picture in my mind is not single but kaleidoscopic: fifteen men singing ‘Yo-ho-ho’ on a wooden box, a horde of civic-minded individuals seeking funds for community projects, Jane Mansfield parading her most outstanding possession. The Basic English people are cheating: there are at least three English words, not just one, spelled chess.

Even if one knows all the dictionary meanings of a printed word this does not guarantee full familiarity. The printed word chest has an initial which is pronounced differently from the same initial in chemistry. The word occurs singular or plural, in various syntactical patterns, in all manner of special usages. Here are a few:

- a chest of Chinese tea
- a chest of solid mahogany
- a chest of ample proportions
- a chest of drawers
- a hope chest
- a hairy chest
- a community chest
- a lovely chest

The preceding list has been drawn up by me with malice aforethought. That the examples are representative of actual occurrences can be seen from the following specific cases drawn from the literature in Basic English:

- go on the air
- go by air
- put on the radio
- run on the bank
- use measured language

- man in the street
- keep in one’s head
- be off one’s head
- Two circles, please.
- That’s in my line.

The fact that expressions of this sort are used in Basic English without explanation is due to another failing in vocabulary-oriented reading programs, namely the failure to distinguish between the capabilities of a native speaker and the problems of a foreign learner. If you are a native speaker of English—at least of British English—the examples cited present no difficulty. You need only to learn the written form of a word to know it in all or most of its uses. But if you are a foreigner beginning the study of English the examples contain a host of pitfalls. There is nothing automatic or instinctive about acquiring a feeling for these problems. As young Cable in “South Pacific” pointedly remarks in another connection, you’ve got to be carefully taught. Some of the terms can be taught as idioms. In many cases an adequate explanation would require a fairly sophisticated linguistic analysis, including especially insights that can be provided by the techniques of transformational grammar. The main point is that they are not easy for a foreigner.

The situation is the same in Chinese. Take the following examples of compounds formed from common characters:

- 學會 (1) to master; (2) scholarly organization
- 細纖 (1) learn to speak; (2) theory
- 會 (1) able to speak
- 會來 likely to come
- 會長 director of an organization

- 長大 grow up
- 長工 long-term employment
- 小說 novel
- 小人 mean person

6
Native Chinese who know the individual characters will in general encounter little difficulty with the combinations containing these characters. But the ability to recognize each character does not guarantee that the foreign student will know the meaning of the compounds formed from them. Part of the problem is to identify the change of meaning which a character undergoes in different combinations. Essential to this is the ability to identify the relationship between the characters. Between two characters A and B it is possible to have many types of relationships. Here are a few examples:

<table>
<thead>
<tr>
<th>A and B</th>
<th>男女, men and women</th>
</tr>
</thead>
<tbody>
<tr>
<td>A or B</td>
<td>早晚, sooner or later</td>
</tr>
<tr>
<td>A's B</td>
<td>我頭, my head</td>
</tr>
<tr>
<td>A number of B's</td>
<td>三天, three days</td>
</tr>
<tr>
<td>A th number of B</td>
<td>三號, third of the month</td>
</tr>
<tr>
<td>B made of A</td>
<td>木船, wooden boat</td>
</tr>
<tr>
<td>B which produces A</td>
<td>火藥, gunpowder</td>
</tr>
<tr>
<td>B which is A</td>
<td>大人, adult</td>
</tr>
<tr>
<td>to B in an A manner</td>
<td>大吃, eat a lot</td>
</tr>
<tr>
<td>A does B</td>
<td>他說, he says</td>
</tr>
</tbody>
</table>

Failure to recognize that ability to identify a printed word or character is only a small part of learning to read leads to absurdities such as the following, expressed by one of the leading exponents of Basic English:

The average rate of learning words in a foreign language is about thirty words an hour. The average learner is willing to give an hour a day for a month to learning the fundamentals of a new language. Hence it is possible for anyone whose natural language is not too remote from English to read anything written in Basic in less than 30 hours.

Chinese characters are more difficult than English words. Therefore to learn some 800 of them—a mystic number which crops up frequently in Chinese programs—a period not of 30 hours but of some 30 weeks, i.e. an academic year, is frequently allotted. The details are a little different for Chinese and English. The reasoning is the same. It is equally fallacious.

As a remedy for this simplistic view of language it would be instructive to remind ourselves of some of the problems encountered by a student in attempting to learn the written form of a foreign language. In order to be able to read something in Chinese the student must, among other things,

1. identify each character
2. grasp its relationship to other (not necessarily adjacent) characters
3. recognize special idioms and turns of phrase
4. be familiar with the subject-matter under discussion and with cultural details such as place names, book titles, names of people, and other specific references.
5. be aware of such subtleties as play on words, unconventional (perhaps even "ungrammatical") manipulation of language, rhythm and balance, and a host of other things which make for literary quality.

Not merely the first item, but all of them and still others are involved in any meaningful definition of "knowing" a character. Such a definition for knowing characters can also serve as the definition for the ability to read. I think, however, that we should recognize the ability to read as including still another aspect which in a way is the very antithesis of the emphasis on knowing characters: The ability to read includes the ability not to read. I mean by this that students should be able to act on Bacon's advice to change one's style of reading to accord with the subject-matter and purpose in reading. The student should be able to read some things word for word so as to extract all the content, but to skim over other things, reading a phrase or sentence here and another there to determine whether to go back and read in greater detail or whether to be content with the information obtained by this selective reading. In many cases, including geographical or personal names, for example, the reader can dispense with knowing the pronunciation or meaning of a specific character, provided of course he is aware of the function of the character in question. It requires a firm grasp of the total structure for a student to be able to do in Chinese what he does continuously in his own language—to read with understanding even when he encounters an occasional new word, especially since in many cases he does not even need to look up the word but can grasp its meaning from the context. (And usage in context, be it noted, rather than definitions in a dictionary, are the best guide to the meaning of words.)

If we are to make any progress in the teaching of reading we must take a good deal more account than has been done to date of these other factors involved in the ability to read. And in order to accomplish this we must carry forward the research and experimentation of pioneers like George Kennedy. A convenient starting point would be to extend somewhat the information regarding frequency of characters to which Kennedy drew attention. Most important in this connection is the study by Dr. Ch'en Ho-ch'in, a graduate of
Teachers College at Columbia who was at one time Commissioner of Education in Shanghai. Ch'en tabulated almost a million characters of running text in various kinds of publications. He found 4,719 different characters. These he published as a list with the characters arranged by order of frequency. The list is headed by the character 大, which occurred 29,592 times, and ends with a number of characters which occurred only once in all the material he counted.

Analysis of Ch'en's data reveals a startling picture of the relative frequency of Chinese characters. The 400 most frequently used characters comprise 73.1 per cent of the total text counted by Ch'en. The least frequently used half of all the different characters comprise a mere 2.5 per cent of the total text. These and other relevant figures are presented in the following table and chart:

<table>
<thead>
<tr>
<th>number of different characters</th>
<th>order of frequency</th>
<th>per cent of total text</th>
<th>cumulative per cent of total text</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>1 - 400</td>
<td>73.1</td>
<td>73.1</td>
</tr>
<tr>
<td>400</td>
<td>401 - 800</td>
<td>12.4</td>
<td>85.5</td>
</tr>
<tr>
<td>400</td>
<td>801 - 1200</td>
<td>5.8</td>
<td>91.3</td>
</tr>
<tr>
<td>400</td>
<td>1201 - 1600</td>
<td>3.3</td>
<td>94.6</td>
</tr>
<tr>
<td>400</td>
<td>1601 - 2000</td>
<td>1.9</td>
<td>96.5</td>
</tr>
<tr>
<td>400</td>
<td>2001 - 2400</td>
<td>1.0</td>
<td>97.5</td>
</tr>
<tr>
<td>2319</td>
<td>2401 - 4719</td>
<td>2.5</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The startling nature of the frequency curve based on Ch'en's material would probably be even more startling if we took into account subsequent trends in the direction of increased vernacularization of style and simplification of characters. On this basis I estimate the proportion of the total text comprised by various numbers of characters in contemporary publications to be roughly as follows:

- first 400 characters: 75 - 80 per cent
- second 400 characters: 85 - 90 per cent
- third 400 characters: 92 - 95 per cent

In each case the figure is probably closer to the top of the range. After the 1200th character each character is likely to occur no more than once in ten thousand characters of running text, after the 2400th no more than once in fifty thousand, and after the 4800th no more than once in a million.

If the figures presented above are accepted as being even very approximately correct, the conclusion seems to me inescapable that efficiency in teaching demands initial concentration on high-frequency characters and temporary if not permanent rejection of low-frequency items. The acceptance of this general principle still leaves a great deal of leeway as to the order, rate, and procedure in introducing characters. I do not at all mean to suggest, for example, that we should start with the character of highest frequency and work our way mechanically down the list. There are various factors which need to be taken into account in determining the details of the reading program.

One of the most important factors has to do with combinations of characters as opposed to single characters, for, as I tried to emphasize above, knowing a character does not necessarily give a student access to the combinations of which it is a part. Combinations of characters are unfortunately not as readily identifiable and countable as single characters. Simply determining whether or not a given sequence of characters constitutes a compound is a complicated matter on which there is not likely to be unanimous agreement. But the situation is not hopeless. Rough figures can be compiled which will throw some light on the general problem of combinations of characters in Chinese.

Supposing we start with a question: "How many compounds are there in Chinese?" Since we don't know what a compound is, we can hardly give a fully satisfactory answer to this question. We can, however, answer it in part by asking a related and readily answerable question: "How many entries are there in Chinese dictionaries?" There are about 100,000 in Tzu Hsi, 104,000 in Mathews' dictionary, over 200,000 in the McGraw-Hill Modern Chinese-English Technical and General Dictionary. For purposes of comparison it may be useful to note that the latest unabridged Webster's dictionary of English contains 450,000 entries. In view of the backward state of Chinese lexicography and the fact that Chinese lexical entries include compounds which translate into English as phrases rather than as single words such as those forming the bulk of the entries in English dictionaries, I strongly suspect that a more complete dictionary would push the figures for Chinese beyond that for English.
But citing these figures gets us no further along than citing the equally frightening figure of 42,000 for the number of different Chinese characters contained in the K'ang-hsi dictionary. Clearly, a great many of these items will never be encountered in ordinary writing and therefore need not be studied.

Supposing we try another tack by asking “How many combinations of characters does a native Chinese know?” I have seen no figures whatsoever on this. Nevertheless I think we can answer the question with a useful degree of accuracy. We can do so provided we accept one assumption, namely that the number of combinations of characters known to native speakers of Chinese is approximately the same as the number of words known to native speakers of English. This seems a reasonable assumption, since the number of combinations in Chinese is probably of the same general order of magnitude as the number of words in English, and since we must suppose that speakers of Chinese and speakers of English of the same general educational and cultural level have about the same level of knowledge. If all this is granted, then we can change our question to “How many words does a native speaker of English know?” For this question there exist some specific and rather startling answers—startling because one of the most egregious errors is underestimating the size of vocabulary at the command of native speakers of a language.

According to educational psychologists who have studied the problem, the number of basic words (quick is a basic word, quickly a derivative word) which American students can recognize (i.e. can understand when spoken and/or written) is as follows:

- 16,000 for first grade child
- 35,000 for fifth grade child
- 60,000 for college undergraduate

These staggering figures suggest the magnitude of our task in bringing our students to anything like the level of competence possessed by native Chinese. However, although they throw some light on the problem of how many compounds we need to teach, they still do not inform us regarding the equally important problem of which compounds to select. I think it possible to get some help in this matter by looking in two directions.

The first is to examine lists of commonly used terms such as the list of 3000 words (te'yu) compiled by the Committee on Chinese Language Reform in Peking. This list, which contains about 800 monosyllabic words and 2200 polysyllabic words, seems to have been arrived at by subjective means, rather than by an objective count such as that undertaken for single characters by Ch'en Ho-ch'in. There are a number of terms of doubtful urgency for the Western student—e.g. such items as the words for ‘persimmon’ and ‘eggplant.’ There is also some question as to the extent to which the sources used for compiling the list affected the figures given for the frequency of the items in various kinds of literature. It is stated that the words included in the list of 3000 items comprise from 78.2 to 88.2 per cent of all words—including repetitions—in elementary school textbooks, 66.5 per cent of the words in selected news items, and 62.1 per cent of the words in an essay discussing qualifications for membership in the Communist Party. Even if we hesitate to project these figures into general literature I think the list is worth careful consideration as at least embodying the subjective views of a group of Chinese who have given some thought to an important aspect of Chinese language teaching.

Even greater help in selecting compounds can be obtained if we examine characters from the point of view of the extent to which they enter into combinations. There are in Mathews' dictionary 7773 characters and 104,000 combinations—an average of 13 combinations per character. The actual range is from zero to over 400. Some 1200 characters enter into no combinations, another 1200 enter into only one, another 900 into two. The distribution can be summarized approximately as follows:

<table>
<thead>
<tr>
<th>Mathew's dictionary</th>
<th>per cent of the 7773 characters</th>
<th>per cent of the 104,000 combinations</th>
<th>range of the number of combinations per character</th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>10</td>
<td>0 - 7</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>20</td>
<td>8 - 23</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>70</td>
<td>24 and over</td>
<td></td>
</tr>
</tbody>
</table>

It is apparent from these figures that some characters are much more productive of compounds than are others. In contrast to the 1200 characters which enter into no combinations, there are 1200 which enter into 24 or more each for a total of over 70,000.

It may be supposed that in general the characters which enter into few combinations are of lesser importance, and those which enter into many combinations are of greater significance. A sampling check of the characters in Mathews' dictionary against those in Ch'en's frequency list confirms the supposition. By and large, characters which enter into few combinations rank low in frequency, those which enter into many combinations rank high in frequency. The converse is also generally true. There are, to be sure, a few exceptions. Chief among these are morphological and syntactical particles such as the plural suffix men and the question particle ma, which although entering into very few compounds in the dictionary nevertheless rank quite high in frequency. That there is a generally close correlation between frequency and compounding can be seen from the following table:

<table>
<thead>
<tr>
<th>Order of frequency of individual characters</th>
<th>Average number of compounds per character</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 50</td>
<td>103</td>
</tr>
<tr>
<td>551 - 600</td>
<td>47</td>
</tr>
<tr>
<td>1151 - 1200</td>
<td>26</td>
</tr>
<tr>
<td>2351 - 2400</td>
<td>13</td>
</tr>
<tr>
<td>3551 - 3600</td>
<td>5</td>
</tr>
<tr>
<td>4669 - 4719</td>
<td>4</td>
</tr>
</tbody>
</table>

A detailed count reveals that the first 1200 characters in Ch'en's list enter...
into close to 60,000 out of the total of 104,000 combinations in Mathews’ dictionary. A sampling of these 60,000 combinations further reveals that in about three-quarters of the cases, that is in about 45,000 combinations, the constituent characters do not go beyond Ch’en’s first 1200 characters themselves.

In other words, with only 1200 characters, i.e., about 15 per cent of the total in Mathews’ dictionary, it is possible to form 45,000 combinations, i.e., about 43 per cent of all combinations in the dictionary.

These figures make clear that in selecting compounds for study it would be well to concentrate on those formed from high-frequency characters, because in so doing we open the way to the acquisition of an enormous number of compounds without having to learn additional characters. It does not follow, to be sure, that because two characters A and B are both high frequency, therefore the compound AB is also of high frequency. Yet if our subjective judgment occasionally leads us astray in selecting compounds composed solely of high-frequency characters, no harm is done. Serendipity assures such useful by-products as further practice on the characters A and B, particularly on the diverse relationships mentioned earlier as existing among characters.

In addition to practice on the immediate relationships among characters we can also obtain practice on the wider situations in which these occur. Just as knowing two characters A and B does not guarantee command over the combination AB, in similar fashion knowing the meaning of two sentences S1 and S2 does not guarantee command over the sequence S1S2, as illustrated by the following simple example:

他不去，我也不去。 ‘He’s not going, and I’m not going either.’

他不去，我也不去。 ‘If he doesn’t go, I won’t go either.’

他不去，我也不去。 ‘Even if he doesn’t go, I won’t go.’

The infinity of relationships to be found among a finite number of characters is nothing less than the total problem of reading. This problem cannot be solved by over-emphasis on sub-aspects of the reading task such as learning characters or mastering sentence patterns. What is required is a many-faceted attack which can develop a sort of intuitive feeling for the language as a whole. Only the total act of reading brings into play all the abilities needed to solve the multitude of problems which constitute the total problem of reading.

It is an ironic fact, however, that the thing which seems to get the least attention in reading programs is reading. We emphasize learning large numbers of characters. We provide flashcards to assist in their memorization. We reinforce this by insisting that students know how to write the characters from dictation. We demand that the radicals and other constituent parts be memorized. We do everything except ask that the student do a great deal of reading. This we postpone until the student has learned “enough” characters. Small wonder that our Peking Duck approach to character study—a combination of forced feeding with little exercise—leaves the student with an acute case of indigestion at about the 1500 character level. The attempt to learn a great many characters is self-defeating when pursued without adequate practice in reading. The student “knows” 1500 characters, but curiously enough he probably spends more time looking up compounds formed from characters already “learned” than in looking up characters he has never seen before.

A large amount of reading will benefit the student even if for pedagogical reasons the material is presented in graded levels of difficulty. Even material of a limited range of difficulty and of more or less uniform content will inevitably contain many variations which help the student to deepen and extend his control over what he has studied. In particular such graded material, by involving the student extensively in the total act of reading, can contribute in a vital way to achieving fluency, as against the agonizing translation which too often passes for reading.

Hence in addition to asking a student “How many characters do you know?” we need to ask “How much have you read?” Insofar as the student’s reading is confined to textbooks we can make very specific answers to this question—we need only count the number of characters of running text in a given book. For the main reading texts which have already been or are being developed in this country (i.e. Y. R. Chao’s Mandarin Primer, IFEL’s Read Chinese I–III, and the new series of Seton Hall materials entitled Beginning Chinese Reader, Intermediate Chinese Reader, and Advanced Chinese Reader) the figures are presented in the following table, which includes, for ease in comparison, figures on the number of different characters, number of combinations, number of characters of running text, and various ratios among these elements:

<table>
<thead>
<tr>
<th></th>
<th>MP</th>
<th>RC I</th>
<th>RC II</th>
<th>RC III</th>
<th>BCR</th>
<th>ICR</th>
<th>ACR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of characters</td>
<td>1372</td>
<td>300</td>
<td>300</td>
<td>400</td>
<td>400</td>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>Number of combinations</td>
<td>2200</td>
<td>390</td>
<td>620</td>
<td>500</td>
<td>1250</td>
<td>2400</td>
<td>2400</td>
</tr>
<tr>
<td>Characters of running text</td>
<td>26,000</td>
<td>24,000</td>
<td>23,000</td>
<td>20,000</td>
<td>120,000</td>
<td>120,000</td>
<td>120,000</td>
</tr>
<tr>
<td>Combinations per character</td>
<td>1.6</td>
<td>1.3</td>
<td>2.1</td>
<td>1.3</td>
<td>3.1</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Running text per character</td>
<td>18</td>
<td>80</td>
<td>77</td>
<td>50</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Running text per combination</td>
<td>12</td>
<td>61</td>
<td>37</td>
<td>40</td>
<td>95</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

These figures show that we have here three different sets of texts embodying quite different ideas on how to teach reading. Most radically different is the idea underlying the Seton Hall approach that even in beginning reading programs a better balance is required in which the emphasis is less on learning large numbers of characters and more on learning many combinations and doing a large amount of reading.
Basically we have two choices: In a given amount of time, we can ask students to learn, say, one or the other of the following:

- 1200 characters + 2000 compounds
- 800 characters + (2000 + X) compounds

I don’t know what number X stands for. I am convinced, however, that it is large, far in excess of the 400 characters not studied in the second approach, because it is easier (though not automatic) to learn new compounds formed from known characters than to learn new characters and new compounds formed from them.

At present our amateurish preoccupation with the popular numbers game of “How many characters do you know?” places a premium on superficial learning in the interest of racking up an impressive total of characters and gives a sense of accomplishment which on closer examination turns out to be largely spurious. If a student learns 1200 characters and 2000 compounds, he has learned only one-fourth of the characters known to a Chinese college undergraduate, but as compared to the latter he has acquired only one-thirtieth of the compounds and has mastered only a tiny fraction of the contextual situations in which the characters might occur, as indicated by the fact that he has done only an infinitesimal fraction of the reading accomplished by the native Chinese. (The fraction is even more infinitesimal if we take account of the closely related activity of speaking.) My reaction to programs which teach 2000 or 1200 (or perhaps even as “few” as 800) characters in a year of half-time study is much like Frederick Mote’s to the “fantastic” Air Force feat of recognizing number sequences spoken at high speed against distracting background noises. In differing degrees such feats reflect a danger against which Professor Mote warned—“the danger that language teachers will attempt to isolate too limited an aspect of the language.”

The justification usually given for the emphasis on learning characters is that knowledge of characters is the prime essential in being able to read, so that a student who knows 2000 characters is surely better equipped than one who knows only 1600, and both are better off than the student who knows only 1200.

I challenge whether this is necessarily true. Beyond the 1200 characters which comprise 82–95 per cent of the total running text in general literature, each additional character, according to the findings quoted earlier, is likely to occur no more than once in ten thousand characters of running text. Instead of stuffing himself with an undigested mass of relatively low-frequency characters the student would be better advised to absorb more compounds formed from characters already learned and to spend more time in actual reading. A regimen consisting of fewer characters, more compounds, and more reading will in general give the student access to a higher proportion of a given text than will a diet consisting of more characters, fewer compounds, and less reading.

Related to the point just made is the question of what can be written with a limited number of characters. Y. R. Chao points out that the 1,100 characters in a system of Basic Chinese proposed by the dramatist Hung Shen “allows much greater freedom of combination than the word list in Basic English. The result is that the language written within the limits of this list is much nearer normal Chinese and gives much less impression of special style than is the case with Basic English.” These remarks have important implications for textbook writing. They mean that with a limited number of characters it is possible to provide students with a great deal of reading matter in a style approaching that of real Chinese. Instead of the paltry amount of reading to which students are now exposed, I would suggest the desirability of forbidding students to learn, say, more than 2400 characters until they have read a million or so characters of running text—the equivalent of only half a dozen books the size of Sun Yatsen’s San-min chu-i.

The question may be raised whether all this reading is really necessary, whether it does not waste some of the student’s time which might more profitably be spent on other things. I concede the possibility but would point out that a literate Chinese can read a book the size of San-min chu-i, which contains 160,000 characters, in half a dozen to a dozen hours. At most the Western student would be wasting only this amount of time. Any time spent in excess of this would represent time needed to learn something already possessed by the native Chinese, whether it be characters, combinations, structures, or—a vastly important but too much neglected item—fluency.

The proposed approach to reading would make it possible for us more rapidly to wean students away from textbookese and to start them on a more solid diet of real Chinese. This process does not need to wait until some indefinite future when students have acquired enough characters. It can begin at a very early stage, particularly if we provide students with closer imitations of or slightly modified excerpts from actual publications. Such material should be sharply distinguished from the kind of watered-down rewriting which is so insipid that all taste of the original literary flavor is lost. It can be illustrated by a selection presented in Beginning Chinese Reader at the level of 360 characters (AND 1066 compounds). The selection is a slightly modified excerpt from Act I of Ts’ao Yu’s Sunrise, the changes consisting only of replacing the names of characters in the play by na ‘male,’ mǔ ‘female,’ and tā ‘he, she,’ and of making some omissions (indicated by . . . ) of things which the student is not yet ready to cope with. The selection is as follows:

女——在××旅館住宿的一個女人,二十五歲。
男——他從前的朋友,二十五歲。
男：他是誰?這個女是誰?這個東西是誰?——我不明白你為什麼跟這麼大的東西來往?他是不是誰?
女：你要知道嗎?——他——我有幾個錢。
男：可是你為什麼跟這麼個東西來往?—
女：我没有告诉你吗？……他有幾個錢……
男：怎麼你現在會……
女：得了……我知道你心裏是不是說我有點太隨便？……
男：我……我……我……
女：你說老實話是不是？
男：對了……你簡直不是我以前想的那個人……你也知道我這一次到這裏來是為什麼？
女：為什麼？我不知道！
男：我不喜歡看你這樣……我要你跟我回去
女：回去？回到那兒去？……
男：車票就在這裏……坐十點的車我們就可以離開這兒……
女：不等等我只問你一句話
男：甚麼？
女：你有沒有錢？
男：我不懂你的意思
女：不懂？……你不要這樣看我！你說我不應該這麼說話？……你難道不明白？

Notes
5. Ch'en Ho-ch'in, Yü-t'ien ying-yung tsu-hui (Shanghai, 1928).
7. In the study by Seashore and Eckerson it is estimated that if both derivative and basic words are counted, the average undergraduate knows 155,736 words. This figure is challenged by another researcher, who comments as follows: "... the statistics of the present writer suggest that this apparently fantastic figure is itself likely to be a gross underestimate by about one hundred thousand words..." (George H. Hartmann, A Critique of the Common Method of Estimating Vocabulary Size, together with some Data on the Absolute Word Knowledge of Educated Adults, Journal of Educational Psychology, 31 (1941), 358.)
8. Han-yü san-ch'ien ch'ang-yung ts'u-piao. There are various editions of this work. The one I have consulted was published in Tokyo in 1962.
9. These figures were computed by looking up a representative sample of characters (1000 out of 7775) and noting the number of compounds listed under each character. This procedure, while adequate enough for our purposes, is of course very crude. In Matthews' dictionary the compounds listed under a character include chiefly those of which the character is the initial. In some cases compounds of which the character is the final are also listed. Compounds in which the character occurs in other positions are rarely listed under the character. As a result the picture of mutual exclusiveness which is presented in particular by the second column of the table is not accurate. It would be possible to improve on the present study by setting up a procedure for obtaining accurate information on such matters as the number of compounds into which a character enters (regardless of position in the compound), the number of compounds limited to certain specific characters, and other relevant data. This could be done by using a computer in conjunction with the Modern Chinese-English Technical and General Dictionary. In this dictionary a four-digit number from the Chinese Telegraphic Code is assigned to each of the 9,500 characters which constitute the 212,000 entries. This information is also available on magnetic tape. A program could be written for the computer requesting information such as that noted above. Even with a computer, to be sure, getting this information and presenting it in a usable form would be a time-consuming task.

I should like to thank Mr. Simon T. H. Chang and Miss Josephine Young for help in compiling the figures presented in this paper.
10. In Chinese, as in linguistics generally, linguistic analysis has been largely confined to the phonological, morphological, lexical, and syntactic levels, with virtually no study being made beyond the level of the sentence. (Zellig Harris has done pioneer work in this area in his papers on "discourse
and teaching in the Chinese field have completely neglected this important
area, the student has no alternative but to make up for it as best he can
by extensive reading.

10. Yuan Ren Chao, Mandarin Primer (Cambridge, 1948); Fang-yu Wang,
Read Chinese Book I (New Haven, 1961); Richard I. Chang, Read Chinese
Book II (New Haven, 1962); Fred Wang and Richard Chang, Read Chinese
Book III (New Haven, 1961); John DeFrancis, *Beginning Chinese Reader*
Reader are scheduled for later publication.

11. For a further comparison, it is worth noting that Sun Yatsen's *San-min
chu-i* contains 2,134 different characters and 163,296 characters of
running text, or roughly 77 characters of running text per different character—
a higher ratio than many works specifically designed as textbooks for
foreign students. (For these figures see Huang Te-shih, *Han-tzu tsung-shu
yu ch'ang-yung tz'u-shu* [Taipei, 1962], p. 18.)

12. These remarks apply with even greater force to textbooks designed to
teach Chinese children how to read. In the first grade, according to one
study, Chinese children learn over 500 characters but less than 300 com-
pounds. (Ta Fang, "Ch'ang-yung tz'u yu ch'ang-yung tz'u shu-liang te
t'ui-ts'e," in Ta-lu tsa-chih yi-wen ts'ung-shu, yi-yen wen-tzu-hsueh
[Taipei, 1963], pp. 13–18.) This preposterous and completely unnecessary
imbalance results in reading matter every bit as inane as that in some of
the first-grade textbooks in the United States.

14. Chao, p. 15.