British Journal of Psychology (1992), **83**, 517–531 *Printed in Great Britain*

The organization of the lexicon in Japanese: Single and compound kanji

John Morton*

MRC Cognitive Development Unit, 17 Gordon Street, London WC1H 0AH, UK

Sumiko Sasanuma

Tokyo Metropolitan Institute of Gerontology

Karalyn Patterson

MRC Applied Psychology Unit, Cambridge

Naoko Sakuma

Tokyo Metropolitan Institute of Gerontology

A word written in Japanese kanji can be single (one kanji character) or compound (more than one character). We investigated the mental representation of single-kanji and two-kanji compound words by means of a long-term facilitation paradigm. Over two experiments, the pattern of results indicated that identification of a target word was facilitated only by prior exposure to that identical word and not by a different word (single or compound) sharing a character with the target word.

The experiments reported in this paper address the structure of orthographic representations, for native Japanese speakers and readers, of words written in kanji, the Japanese ideographic (or morphographic) writing system. We will begin this paper with a brief description of the nature of the kanji script. Next we will introduce the theoretical framework within which the experiments are to be located. Finally we will outline the alternative models for kanji recognition which the experiments are designed to test.

(*a*) The structure of kanji

Words in kanji can be divided into two kinds, *single* and *compound*, with the great majority of lexical items being compounds. Compound words are made up of a sequence of single characters, and both the meaning and the pronunciation of the compounds are determined by the nature of the components. This will be illustrated

* Requests for reprints.

	目 /me/ (eye)	花 /hana/ (flower)	波 /nami/ (wave)
KUN-reading	/me/ : /medama/ 目玉 (eyeball)	/hana/ : /hanaya/ 花屋 (florist)	/nami/ : /tsunami/ 津波 (tidal wave)
ON-reading	/moku/ : /mokuteki/ 目的 (purpose)	/ka/ : /kokka/ 国花 (national flower)	/ha/ : /boHhatei/ 防波堤 (breakwater)

Figure 1. Examples of compound words where the individual kanji may have ON or KUN pronunciation.

shortly. Kanji characters usually have two pronunciations or 'readings'. These are called the KUN and the ON readings. Historically speaking the KUN reading is the original Japanese word for the particular concept and the ON reading (or readings) is the Chinese name of the Chinese character which was the precursor of the kanji character. The KUN reading is the one which is almost always used when the single character occurs in isolation. The ON reading is that which is usually, but by no means always, used when the character is found in a compound.

The pronunciation of a compound word is determined by its components. However, there appear to be no rules which determine whether it is an ON reading (often there are several) or the KUN reading of the component characters which are used in a particular compound. In Fig. 1 we give some examples of compounds in which both forms are used. Thus the character for *eye*, which is pronounced /me/ in isolation can occur either as /me/, the KUN reading, or as /moku/, the ON reading, in compounds. The other two examples show that the reading which is used, ON or KUN, is not determined by whether a character is at the beginning or end of the compound.

The meaning of a compound is usually related to the meanings of the components. Thus *eyeball*, /medama/, is made up of /me/, *eye* and /dama/, *round object*. The word /mokuteki/ is made up of /moku/, the ON reading of *eye* and /teki/ meaning *target*. The result has the meaning *purpose*, which is less transparently related to the components. Some further details on the structure of kanji can be found in Morton & Sasanuma (1984).

The question which we wish to address is that of whether compound words are recognized as integrated units or whether their recognition is contingent upon the recognition of their components. The theoretical framework within which we are operating is that of information processing and the methodology to be used is that which was employed in the development of the logogen model (Clarke & Morton, 1983; Jackson & Morton, 1984; Kempley & Morton, 1982; Morton, 1979, 1981; Morton & Patterson, 1980).

(*b*) Information processing and the logogen model: The experimental paradigm

The logogen model is one of several major theoretical frameworks used in experimental psychology, psycholinguistics and neuropsychology over the last few decades to account for phenomena of single-word recognition, comprehension and production. Logogen units are lexical representations corresponding to familiar words (or perhaps morphemes: Murrell & Morton, 1974). There are separate logogen systems for spoken and written words; it is only the visual input logogen system that will concern us here. Logogen units have threshold values whose resting levels are determined by the reader's long-term experience with words. When the reader encounters a word, logogen units which have significant overlap with the stimulus string receive some degree of activation. Under normal conditions, the logogen corresponding to the presented word will of course provide the best match, reach its threshold, and therefore send activation to other parts of the system concerned with word meaning and pronunciation.

The major experimental paradigm used to explore the properties of logogen units in English has been long-term facilitation of word identification (Clarke & Morton, 1983; Jackson & Morton, 1984). The basics of this paradigm involve (1) a 'pre-training phase' where subjects are exposed to words on which they perform some task or judgement (for example, rating the words for pleasantness), followed some time (perhaps 30 minutes) later by (2) a test phase where words are identified, typically under conditions of reduced exposure duration. The crucial experimental manipulation concerns the relationship between items in the pre-training and test phases. Summarizing over a number of experiments with English speakers/readers, the general pattern emerging from this paradigm is that test identification of a word (such as SEEN) is significantly facilitated relative to control conditions if and only if the identical word (SEEN) or, alternatively, a morphologically related word (e.g. SEES) was included in the pre-training phase. Significant facilitation refers to successful recognition either at a lower exposure duration or with a higher probability for a fixed exposure duration.

The interpretation of this pattern has been that exposure to a word lowers the threshold of the logogen unit corresponding to that word (or morpheme), an effect which lasts for minutes or hours during which time the threshold gradually returns to resting level. There are of course components of the word-processing system other than the logogens *per se* at which facilitation effects could occur, for example the visual analysis (of features and/or letters) that precedes logogen activation, and the semantic and phonological representations of words which are retrieved or computed after logogen activation. Word recognition can indeed be primed by prior presentation of a different word sharing orthographic (Evett & Humphreys, 1981), phonological (Humphreys, Evett & Taylor, 1982) or semantic (Henderson, Wallis & Knight, 1984) features with the target word. However, such priming effects from non-identical words (typically explored with the lexical decision paradigm rather than perceptual identification) have a very short life expectancy. Orthographic, phonological or semantic priming has normally dissipated by the time that a few seconds, or a few intervening items, have elapsed. It is therefore our assumption that changes in logogen threshold, which necessarily affect only recognition of that specific word (or a morphological variant thereof) are responsible for long-term facilitation effects.

Since the properties of logogens are putatively revealed by the long-term facilitation paradigm, one can use this paradigm to ask questions about the relationship between logogen units and various kinds of word structures. For example, is a compound word represented by a single logogen unit corresponding to the whole compound, or is it recognized via activation of logogens for the individual components of the compound? An experiment by Osgood & Hoosain (1974) suggests that familiar nominal compounds such as STOCK MARKET or TRAP DOOR are represented by whole-compound logogen units. Pre-training with STOCK MARKET provided subsequent facilitation of the whole compound (STOCK MARKET) but not of one of its components (MARKET). On the other hand, a non-compound noun phrase (such as STREET MARKET) presented in pre-training did facilitate test identification of MARKET.

The present paper is addressed to similar questions regarding recognition of words written in Japanese kanji. As explained above, the great majority of kanji words are compounds consisting of two (or more, but most typically two) characters. If the Japanese reader's logogen system were organized along the same formal principles as the English reader's, then we might predict from Osgood & Hoosain's result that logogens would correspond to kanji compounds. However, the structure of the kanji writing system differs in many ways from an alphabetic writing system like the one used in English; therefore the logogen model as derived for English makes no predictions about the equivalent information-processing description of Japanese. These questions are, rather, empirical ones which can only be answered by experiments using the equivalent paradigm with Japanese subjects reading familiar Japanese words. We will now describe some possible patterns of facilitation effects in identification of kanji compound words.

(*c*) Possible results and models for kanji recognition

Notation

Kanji characters will be referred to with italicized capital letters. Compound kanji words will be referred to by a string of capital letters. Thus the compound word AB is made up of two single characters A and B. We also have to distinguish between the KUN and ON readings for a character. This will be done with lower-case letters. Thus $A(x)$ indicates that the single character A is read in isolation as 'X' which will normally be the KUN reading. We will then distinguish between $A(y)B$ as a compound in which A is given the ON reading and $A(x)D$ in which the character A is given the KUN reading.

The first experiment to be described investigated the interfacilitation of single kanji characters and compounds in which that same character occurs either in the first or the second position. In the compounds, the single character could have either the ON reading or the KUN reading. The three kinds of stimuli would thus be represented respectively as $A(x)$, $A(y)B$, $CA(y)$ and $A(x)$, $A(x)D$, $EA(x)$.

There are two broad classes of model differing in respect of whether recognition of the compound word is contingent on the recognition of its components.

Model 1

In this model, illustrated in Fig. 2a, the logogens corresponding to compound characters have no inputs other than those from the logogens corresponding to their components. If and only if the latter succeeded in identifying the components would the compound be recognized. From such a system we can state directly that under given stimulus conditions, the probability, $p(AB)$, of identifying a compound AB will be equal to the product of the probabilities of identifying the components. Thus:

$$p(AB) = p(A) . p(B)$$

Within this model, the categorization of a compound word would result from the categorization of its components by a simple combination rule. This logical computation would take place after the logogen system and would not be subject to

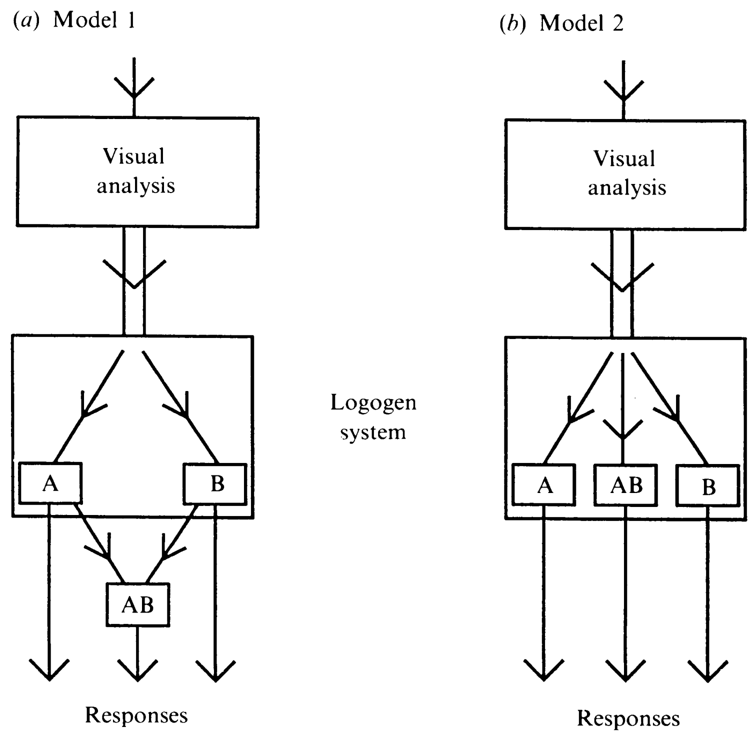

Figure 2. Possible models of the organization of the visual input lexicon in Japanese. In Model 1, the recognition of the compound *AB* is contingent on the recognition of the component kanji *A* and *B*. The unit *AB* is effectively an AND gate. In Model 2, recognition of the compound *AB* is effected independently of recognition of the components *A* and *B*. The logogen system is defined as the level at which long-term facilitation of perceptual recognition takes place. A similar definition would apply in other frameworks.

facilitation effects. On the other hand, one might expect the logogens for the single characters to be liable to facilitation effects as with single words in English. That is, we would expect to find an effect of the prior presentation of *A*, on its subsequent recognition in a T scope. Suppose, then, that as a result of identity priming, performance on a single character goes up to $p'(A)$. The same value should be found for priming by *AB* and *CA* on the subsequent recognition of *A* (which we will write as *AB-to-A* and *CA-to-A*) since the involvement of the *A* logogen would be identical in all cases.

For the compound words there will be an effect of identity priming (*AB-to-AB*) as a result of the effects on the logogens for the single characters. Performance should be given by:

$$p'(AB) = p'(A) . p'(B)$$

Furthermore there should be a priming effect from single to compound characters (*A-to-AB* and *A-to-CA*). This will be smaller than compound word identity priming, giving a value of $p'(A) . p(B)$. The same effect should be found for *CA-to-AB* and vice versa.

Model 2

In this model, shown in Fig. 2b, both single and compound characters are represented at the logogen level. Furthermore, there are no direct interconnections at this level, all logogens receiving input solely from the visual analysis system. With such a model there are no *a priori* expectations with regard to the relative levels of identity facilitation for single and compound characters. We would expect no facilitation from single to compound characters – that is, there would be no lasting effects on the *AB* logogen of presenting the character *A*. For the same reasons we would not expect facilitation between compound characters such as *AB* and *CA*. Whether or not there was facilitation from compound to single characters would depend upon whether the stimulus *AB* led to the operation of the logogen for *A* as well as that for *AB*. An alternative would be that compound logogens inhibited the operation of logogens corresponding to their components.

The logogen model has been influential in developing conceptions of the processes involved in word recognition; but (*a*) there are components of these processes, such as feature analysis/letter recognition, which the logogen model leaves largely unspecified; and (*b*) there are, of course, other models of word recognition. One such model which addresses particular attention to the interplay between word- (or logogen-) level representations and the earlier levels of feature and letter analysis is the interactive activation model of McClelland & Rumelhart (1981; Rumelhart & McClelland, 1982; McClelland, 1985). As none of the experiments or issues dealt with in this paper concerns differences between logogen and interactive-activation conceptions of written word perception, it would be inappropriate to discuss these differences. Indeed, we only mention the McClelland & Rumelhart model to point out that our major question (that is, whether compound words are recognized as units or via their components) could also be framed quite naturally within the interactive activation theory. The question would be whether the network needs only the existing three levels of feature, letter and word units, with compound words having word-level representations just like any other words; or whether a better characterization would include a fourth, top level for compound words receiving activation from the individual component words located at level three.

EXPERIMENT 1
Method

Selection of stimulus words

Stimulus words were selected such that triplets could be found of the form *A*(*x*), *A*(*y*)*B*, *CA*(*y*) for the ON reading set and *A*(*x*), *A*(*x*)*D*, *EA*(*x*) for the KUN reading set. The following other restrictions were also observed.

1. For the *A* words: these words were of high frequency usage, defined by having been learned in the course of primary school education. They were all nouns and of a graphic complexity ranging from 5 to 16 strokes. As a rule these are read as KUN when found as single kanji words.
2. *B*, *C*, *D* and *E* characters: these single characters did not overlap at all with the *A* set. In the compounds, *B* and *C* characters should be given the ON reading while *D* and *E* characters should be given the KUN reading.
3. Compounds: these were rated for familiarity by a group of college students using a seven-point

rating scale. Each word was rated by 40 students. Only words with a mean rating of 3.0 or above were used in the experiment.

Sixty sets of words were found for the ON reading list – comprising triplets of *A*, *AB* and *CA*. A further 60 sets were found for the KUN reading list, comprising *A*, *AD* and *EA*. The *A* words were identical in the two lists with the exception of four characters, the properties of the language plus the constraints we had imposed on selecting stimuli restricting us to this extent. All the stimulus words were typeprinted on to white cards using the Mingcho typeface with a size of 1 cm squared per single character.

Experimental design

The experiment was divided into two phases. The first phase was a pre-training session and the second phase was a test session. In the ON test session subjects were presented with items of all three forms: *A*, *AB* and *CA*. For each of these three kinds of stimuli there were four possible pre-training conditions differing as to whether the subject was presented with *A*, *AB*, *CA* or was given no pre-training germane to that particular triplet. For the single characters, then, we could measure both identity facilitation and the extent of facilitation from compounds to the single character. For the compound characters we could estimate identity facilitation, single-to-compound facilitation and compound-to-compound facilitation. Exactly the same design applied in the KUN session, giving estimates of identity, part-to-whole, and whole-to-part priming for *A*, *AD* and *EA* words.

In accordance with this design the 60 sets of kanji triplets in each list were divided into three sets of 20. These sets differed as a function of whether the *A*, the *AB* or the *CA* words (or their equivalents for the KUN experiment) were designated as the test stimuli. Each set of 20 was further subdivided into four sets of five which differed as to which pre-training stimulus was used. The pre-training stimuli were rotated for all three kinds of test stimuli to four subject groups such that the pre-training phase for each subject consisted of 15 *A* stimuli, 15 *AB* stimuli and 15 *CA* stimuli. All subjects had the same test running order and all subjects had the same notional pre-training order which was random with respect to the test order with certain restrictions. The four subject groups differed with respect to exactly which pre-training stimulus they saw. Thus, the first stimulus seen in pre-training by subject group 1 was *A* (the test stimulus), for subject group 2 it was the *AB* compound, and for subject group 3 was *CA*. For the fourth subject group the test stimulus was a control word for which there was no relevant pre-training stimulus. The design is illustrated in Table 1.

Table 1. Illustration of experimental conditions in Expt 1

Prime	Test words		Condition
	花	*O–A*	(Control)
花	花	*A–A*	(Identity)
花粉	花	*AB–A*	(Compound–single)
開花	花	*CA–A*	(Compound–single)
	坂道	*O–EA*	(Control)
道	坂道	*A–EA*	(Single–compound)
坂道	坂道	*EA–EA*	(Identity)
道順	坂道	*AD–EA*	(Compound–compound)

Apparatus

During the test session, the stimuli were presented in a Takei T scope with a timer which allowed control of exposure durations in 1 ms steps. The viewing field was one metre from the eyes, the viewing

Morton, John, ***The organization of the lexicon in Japanese: Single and compound kanji*** , British Journal of Psychology, 83:4 (1992:Nov.) p.517

field being 22 cm. The subjects were shown a fixation point for 1.5 s, followed by a 0.5 s gap before the stimulus was presented.

Subjects

Subjects were college undergraduate volunteers. There were 12 males and 12 female subjects aged between 18 and 24 years old. All subjects took part in both the ON and the KUN sessions.

Procedure

To determine an appropriate presentation duration for the test phase, the subjects were first tested with a mixture of 10 simple and 10 compound kanji which were not involved in the experiment. The test was preceded with the following instructions.

> I am going to show you some kanji words on a screen. They are either single or compound kanji words, and will be shown at the centre of the screen. In the case of compound words, the two kanji are arranged vertically. When I say 'hai', look at the fixation point at the centre of the screen and push the button in front of you. Let's practice ... all right? Your task is to tell what word you see. Guess if you cannot tell. We are not measuring the speed of your response, so you don't have to respond in a hurry.

The method of ascending limits was used, starting at 20 ms in steps of 2 ms to the criterion of two successive correct responses. For each subject, the means and SDs were taken of the durations at the first of the correct responses for single and compound words separately. In the testing session subjects were started off with durations one ms less than these means using a mixture of five single and 10 compound words. If a hit rate of 40 to 50 per cent correct was achieved, no change was made. Otherwise, the duration of exposure was increased or decreased within the range of one SD until the subject achieved the target hit rate.

In the pre-training session, which came next, subjects were presented with the list of pre-training words on individual cards with the instructions to rate them for their imageability on a seven-point scale. The stimuli were presented once every 10 seconds. This procedure was repeated twice, the subjects being told that we wanted to check their consistency. The 60 test stimuli were all presented at the exposure duration determined in the first part of the experiment. They were preceded by 15 warm-up stimuli, being a mixture of simple and compound characters. The same subjects participated in both the ON session and the KUN session, with an interval of over a month between the two.

Results

Table 2 shows the mean per cent correct for each of the three categories of word, under each of the four pre-training conditions for the ON and KUN sessions. One-way analyses of variance were carried out for the three categories of test word in the two sessions. The *F* values and significance levels for analyses by subject and by word are given in the table. Main effects of the pre-training conditions were found only for compound words in both analyses. To identify the source of these effects, the Newman–Keuls multiple comparison test was used to compare differences between pairs of the pre-training conditions for each of the two sets of compound words in both the ON and KUN experimental sessions. In all cases, apart from the *AD* stimuli, identity priming was significantly better than all other pre-training conditions in both analyses. With the *AD* stimuli the ANOVA was only significant for the by subject analysis; the by word analysis just missed significance. A *post hoc* Newman–Keuls test gave a significant difference only between identity priming and the control.

Table 2. Results of data analysis for Expt 1

	Correct responses (%) Prime				ANOVA by subject			ANOVA by word		
< ON >	*O*	*A*	*AB*	*CA*	$F(3, 92)$	p	Newman–Keuls 5%	$F(3, 76)$	p	Newman–Keuls 5%
A:	52	*58*[a]	61	59	1.01	.394		0.55	.652	
AB:	46	54	*80*	53	8.83	.000	*AB* > *O*,*A*,*CA*	7.35	.001	*AB* > *O*,*A*,*CA*
CA:	43	47	44	*65*	4.90	.003	*CA* > *O*,*A*,*AB*	3.50	.020	*CA* > *O*,*A*,*AB*
	O	*A*	*AD*	*EA*						
< KUN >										
A:	55	*61*	55	53	0.55	.647		0.31	.819	
AD:	33	47	*59*	38	6.91	.000	*AD* > *O*,*A*,*EA*	2.64	.056	n.s.
EA:	47	49	49	*73*	7.25	.000	*EA* > *O*,*A*,*AD*	4.99	.003	*EA* > *O*,*A*,*AD*

[a] Italicized values indicate the identity priming conditions.

The basic recognizability of single words was greater than the compound words when the baseline conditions were compared. The average for single kanji was 53.5 per cent and that for the compound kanji was 42.8 per cent. This difference was significant by subject ($F(2, 141) = 4.18$, $p = .017$) but not by word.

Errors. In 54 per cent of the error responses to unprimed compound words, the subjects mentioned one of the two characters of the compound. There was a strong trend for the response to be appropriate for the first (topmost) of the two characters, the ratio being about 3:1.

Discussion

The major findings of the first experiment can be summarized as follows:

1. For single characters there was no significant identity priming of recognition over the time intervals used.
2. For compound characters there was strong identity facilitation, but no cross-facilitation between compounds sharing a single character.
3. There was no priming effect from a single character to a compound word which contains it.
4. There was no facilitation at all from compound stimuli to the single characters they contain.

From this pattern of data we can rule out Model 1 in Fig. 2a. That is, compound words are not recognized by virtue of the categorization of the component characters. There are two kinds of evidence against this model. The first is the general lack of part-to-whole, whole-to-part and compound-to-(different) compound priming. In the present experiment, this evidence is weak due to the lack of significant identity priming with the single kanji. However, there is another piece of evidence against such an idea: recognition of the compound words was much better than would be predicted from the recognition of the individual components. In a system such as that shown in Model 1, which relies solely on the categorization of the components, as described in the introduction, the probability of recognizing *AB*, p(A . B), could not be better than the product of the probabilities of recognizing the components, p(A).p(B). This should give a value for the compound control conditions of about 29 per cent rather than the average of 43 per cent which was found. This result is independent of the question of priming which will be addressed in the next experiment.

We are thus left with Model 2, shown in Fig. 2b. In this model the logogens for single and compound words are fed separately from the stage of visual analysis and there are no interconnections between them. Such a model puts no constraints on the relative levels of the control recognition levels for compound and single characters. It is also clear that in such a model there would be no facilitation between compound words sharing a character.

Thus far the data resemble the effects found by Gipson (1984) with compound English nouns. His experiment was with *spoken* words and he showed that presentation of a component of a compound noun, e.g. 'stock' or 'market', had no facilitating effect on the subsequent recognition of the compound ('stock market'). He also found that presentation of the compound had no long-term priming effect on recognition of the components, in line with the results for written words reported by Osgood & Hussain (1974).

It should be remembered that the present experiment was carried out on high frequency compound words. If low frequency compounds were used we need not expect to find exactly the same results. Indeed, one possibility is that Model 1 predicts the facilitation patterns for low frequency words, the transition to Model 2 being the result of repeated exposure to the compound.

The lack of significant facilitation from compounds to the single kanji might be thought surprising in that normal Japanese text offers no purely spatial cues to indicate whether a single kanji is part of a compound. (In the English equivalent, if we write STOCK MARKET, as opposed to STOCKMARKET, we would have no spatial indication that STOCK was part of a nominal compound.) However, in the (even more unexpected) absence of identity facilitation for the single kanji, it is impossible to draw any further conclusions.

It is possible that with the single kanji we were running into apparatus limits. With a paradigm where word recognition is the variable dependent on the duration of exposure of the stimuli, the control of exposure duration is crucial. The apparatus used in this experiment, a Takei T scope, uses fluorescent light. Rise and decay times for fluorescent lights are subject to random variation and it is possible that noise was added to the data by such effects. To check on this we carried out a partial replication of Expt 1.

EXPERIMENT 2

The main purpose of Expt 2 was to replicate the first experiment with tighter control over the exposure duration of the stimulus words. In addition, we were concerned that the separate KUN and ON reading sessions in Expt 1 might have led to special strategic effects. Accordingly, for Expt 2, ON and KUN compounds were combined in the same experiment. The special objectives of Expt 2 were to see whether we could find identity priming for single kanji under such conditions, and to confirm (or otherwise) the patterns of compound-word priming found in Expt 1.

Method

Stimulus words

A total of 72 single and compound kanji pairs were used. There were 18 cases of single kanji paired with each of the four compound types $A(y)B$, $CA(y)$, $A(x)D$, and $EA(x)$. Fifteen of each type were taken from Expt 1 and a further three were selected using the same criteria as before. The total 72 pairs were combined into one list. Additionally, 30 words were selected for threshold determination and a further 15 used as warm-up stimuli.

Experimental design

The basic design was the same as in Expt 1 with two exceptions.

1. Test word categories were reduced to two – single words and compound words – collapsing over type of compound.
2. Pre-training conditions were reduced to three – control, identical and cross-category.

The design is illustrated in Table 3.

Table 3. Examples of experimental conditions in Expt 2

Prime	Test words			*Condition*
	花	*O*	*A*	Control
花	花	*A*	*A*	Identity
花粉	花	*AB*	*A*	Compound–single
	坂道	*O*	*CA*	Control
道	坂道	*A*	*CA*	Single–compound
坂道	坂道	*CA*	*CA*	Identity

Apparatus

A slide projector (Kodak Carousel 5600) with a mechanical shutter driven by electrical switching (Nihon Koden Model SEN-11,01) was used. This device allowed us to control the opening and closing of the shutter in 1 ms steps. The brightness level was controlled to be 3.76 cd/m^2 with the shutter fully open.

Results and discussion

Table 4 shows the mean percentage of correct responses for each of the two categories of test words under each of the three pre-training conditions. Performance on ON and KUN words is listed separately.

Table 4. Mean percentage of correct responses for ON and KUN, Expt 2

	ON		KUN		Total	
	Single[a]	Compound	Single	Compound	Single	Compound
Control	48.2	59.3	61.2	53.2	54.7	56.2
Single	*61.2**[b]	51.8	*69.2*	55.0	*65.2**	53.3
Compound	51.2	*69.2*	66.0	*69.2**	58.7	*69.2***

*$p < .05$; **$p < .01$.
[a] These words receive KUN pronunciation when presented singly.
[b] Italicized values indicate identity priming.

Two-way analysis of variance (pre-training conditions × ON/KUN) was performed for single words and compound words separately. For single words, pre-training conditions yielded a main effect ($F(2, 161) = 4.32$, $p < .015$) and so did ON/KUN ($F(1, 161) = 16.45$, $p < .000$); while for compound words only pre-training condition yielded a main effect ($F(2, 161) = 8.85$, $p < .000$). The interaction of these two factors was not reliable in either case.

Pairwise comparisons (by means of the Newman–Keuls test) between the three pre-training conditions for the combined ON/KUN results yielded significant differences between control and identity priming only, for both single and compound words. However, when ON and KUN words were analysed separately, significant differences were restricted to identity priming for ON single words and KUN compound words.

The results of Expts 1 and 2 combined suggest that both single and compound kanji words are facilitated only by pre-training with the identical word, indicating that the unit of kanji recognition is a word rather than a character. There are unknown factors that operate, however, such that the facilitation effect is far greater for compound than for single kanji.

General discussion

In the experiments reported here we investigated interfacilitation between two types of kanji words, single kanji and compounds containing those single characters either in the first or second position, for both ON and KUN reading test words. The major findings can be summarized as follows:

1. Identification of both single and compound kanji words is facilitated by pre-training with the identical word. No interfacilitation was observed either between compound word pairs which shared a character, or between single and compound words.
2. Kanji words composed of single characters are generally easier to recognize than kanji compound words but somewhat less likely to benefit from identity priming.

explanations which cannot be resolved here. For example, a single kanji, while (from our data) not affecting the recognition units of compound kanji which contain it, could affect the semantic and phonological representations of such compound kanji and its recognition unit could be inhibited by them by feedback, thus reducing or removing any identity priming effect. To confirm such conjectures would require much more extended data collection, but, if it turned out to be the case, would imply that the logogen model as presently constituted would not be able to account for the data.

In any case, our findings indicate that recognition units of kanji are formed on the word level rather than on the level of the individual character. In terms of the options in Fig. 2, Model 1 is not a possible representation of the Japanese lexicon.

Acknowledgements

We are deeply indebted to Dr Ken Goryo for his generous assistance in many phases of carrying out the experiments.

References

Clarke, R. & Morton, J. (1983). Cross modality facilitation in tachistoscopic word recognition. *Quarterly Journal of Experimental Psychology*, **35A**, 79–96.

Evett, L. J. & Humphreys, G. W. (1981). The use of abstract graphemic information in lexical access. *Quarterly Journal of Experimental Psychology*, **33A**, 325–350.

Gipson, P. (1984). A study of the long-term priming of auditory word recognition. Unpublished doctoral dissertation, University of Cambridge.

Henderson, L., Wallis, J. & Knight, D. (1984). Morphemic structure and lexical access. In H. Bouma & D. Bouwhuis (Eds), *Attention and Performance X*. Hove & London: Erlbaum.

Humphreys, G. W., Evett, L. J. & Taylor, D. E. T. (1982). Automatic phonological priming in visual word recognition. *Memory and Cognition*, **10**, 576–590.

Jackson, A. & Morton, J. (1984). Facilitation of auditory word recognition. *Memory and Cognition*, **12**(6), 568–574.

Kempley, S. & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, **73**, 441–454.

McClelland, J. L. (1985). Putting knowledge in its place: A scheme for programming parallel processing structures on the fly. *Cognitive Science*, **9**, 113–146.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, **88**, 375–407.

Morton, J. (1979). Word recognition. In J. Morton & J. C. Marshall (Eds), *Psycholinguistic Series II*. London: Elek Scientific Books.

Morton, J. (1981). The status of information processing models of language. *Philosophical Transactions of the Royal Society of London B*, **295**, 387–396.

Morton, J. (1982). Disintegrating the lexicon: An information processing approach. In J. Mehler, E. C. T. Walker & M. Garrett (Eds), *Perspectives on Mental Representation*, pp. 89–110. Hillsdale, NJ: Erlbaum.

Morton, J. & Patterson, K. (1980). A new attempt at an interpretation, or an attempt at a new interpretation. In M. Coltheart, K. Patterson & J. Marshall (Eds), *Deep Dyslexia*. London: Routledge & Kegan Paul.

Morton, J. & Sasanuma, S. (1984). Lexical access in Japanese. In L. Henderson (Ed.), *Orthographies and Reading*. Hillsdale, NJ: Erlbaum.

Murrell, G. A. & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, **102**, 6, 963–968.

Osgood, C. E. & Hoosain, R. (1974). Salience of the word as a unit in the perception of language. *Perception and Psychophysics*, **15**, 168–192.

Rumelhart, D. E. & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, **89**, 60–94.

Received 23 November 1990; revised version received 11 March 1992