LEARNING AND REMEMBERING INTERACTIVE COMMANDS

P. Barnard, N. Hammond, A. MacLean and J. Morton

MRC Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF, England

There is a rich and expanding folklore concerning the consequences of inappropriate naming of computer commands. The problems are particularly acute for occasional users of interactive systems who may be unfamiliar with the jargon of computing. While "naming" has long been of interest to philosophers, linguists and psychologists [2], there is little systematic research on the psychological processes involved in the understanding and acquisition of the vocabularies of interactive computer systems.

Since the names for interactive commands tend to be drawn from the wider vocabulary of natural language, occasional users are faced with the task of understanding, learning and remembering new meanings for the words. Such considerations suggest that names which reflect users' own conceptions of command operations should facilitate learning. In one relevant study, Landauer, et al. [6] showed that potential users invoked a different vocabulary to describe manual text editing operations from that commonly employed by existing computer editors. Somewhat surprisingly they then showed that making names more or less natural had little effect on the time taken to learn a miniature editing system. An important qualification to the interpretation of this result is that only a very small command set was employed.

While Landauer et al. were unable to show that "naturalness" has a demonstrable effect on ease of use, other studies have found that names do influence ease of use. Carroll [3] has shown that command sets which include congruent terms [e.g. opposites such as RELEASE-TAKE] were rated as being preferred and were also better learned than were sets including non-congruent terms [e.g. UNHOOK-GRAB]. In this case the systematicity of having congruent pairs of terms provided redundancy in the naming scheme which facilitated use.

Other types of semantic relationships between words in a command set may be expected to have interfering rather than facilitatory effects. It might, for example, be expected that words in the same command set which express similar meanings in natural language would be confusable [e.g. CHANGE, ALTER]. In fact, Hammond et al. [5] have shown that the extent to which individual command names are ambiguous in relation to the relevant set of underlying operations correlates with the pattern of errors made in interactive dialogue. Their study also suggested that command sets consisting entirely of words whose meaning is relatively specific or constrained (e.g. "substitute") may precipitate different strategies for understanding and learning than those precipitated by use of command words whose normal meaning is more generic or less constrained (e.g. "change").

The relationship between the normal meaning of a command word and its meaning in interactive dialogue forms the focus of the investigation to be reported here. The central issue is straightforward. The semantics underlying generic words like "change" or "move" represent information covering a whole class of potential actions or operations. More specific terms such as "substitute" or "advance" invoke additional semantic information concerning the precise nature of the change or movement. For example, "substitute" presupposes "change" and also conveys additional information such as "replace with an entity of a similar type"; while "advance" presupposes "move" with the addition of directional information. Such additional information, if appropriate to a particular command operation, might well be expected to facilitate learning and remembering of command names and their operations. Laboratory memory tasks indicate that this general/specific contrast is likely to be a potent variable ([7] and [4]).

However, simply arguing that specific command names might facilitate performance could result in other relevant considerations being overlooked. Generic names also have potential advantages. The most obvious derives from their very generality. This means that the contexts in which they can be used are much broader than those in which more specific names are used. Three consequences might follow. First, the actual names, which tend to be of higher frequency of occurrence in the language, may be easier to generate from memory at a point when a user is uncertain of the precise command word needed. Second, the fact that they occur in widely varying contexts may enable them to be used in the novel computing context more readily than words which tend to be used in more restricted contexts. Third, since generic words do not implicitly convey a precise meaning, their meaning in context may have to be explicitly assigned. In doing so the relationship between a command operation and its name may be given a greater amount of explicit semantic processing by the user in the course of learning. Such explicit semantic processing, when compared with specific names which implicitly convey additional semantics, could plausibly improve subsequent retention.

Since there are arguments in favour of both "types" of command name, the issue is plainly an empirical one. Although the naming issue constitutes the central focus of the present study two further considerations motivated the design. First, there is evidence both from standard laboratory paradigms and from research on the use of computer commands that the wider task demands also influence the learning of command dialogues [5]. Accordingly, the study was designed to examine learning and memory where initial training occurred under different task demands. Second, it is entirely plausible that users of varying ability or cognitive style adopt different learning strategies. Accordingly, differences between individuals within the user sample were also monitored.

EXPERIMENTAL DESIGN AND METHOD

The experimental task was of a word-processing type in which users were shown a distorted proverb (e.g. sti tch in a time xpqy aves nine). Their job was to produce the correct proverb by issuing a sequence of commands. Forty-eight naive computer users, in four independent groups of twelve, performed this task. For two groups the command names were relatively general (e.g. Move, Add) whilst for the remaining two groups the names were more specific (e.g. Front, Insert). The two sets of names did not differ significantly on a standard measure of word frequency.

For one group of users under each naming condition, the solution path was constrained in an ordered step-by-step manner. In this task condition users were shown a "target state" for each step of the solution. The target state could always be achieved by issuing a single command. As soon as this target was achieved, a new target was shown, each time one step closer to the corrected proverb. Thus, in these conditions the subjects achieved a solution by progressing from one predefined "subgoal" to another. For example, users would initially be presented with "sti tch in a time xpqy aves nine" and the subgoal target "stitch in a time xpqy aves nine", then the subgoal "a stitch in time xpqy aves nine" and so on. For the other task condition in each naming condition, users were free to determine their own solution paths. Here only the "supergoal" target of the correct proverb was shown. A minimum of six commands were always required to complete a trial. Figure 1 shows an example display during task solution for the supergoal condition with specific command names. In this example, the subject is in the course of entering a "join" command.

The experiment was conducted in two sessions. In the first session, following initial instructions, subjects were required to correct eight proverbs. They had twelve commands at their disposal. Two of these commands involved fetching the proverb from a file and returning the final

```
----------------- COMMAND OPTIONS -----------------
 (Type 'help' for further details)

 append   cancel   end     fetch
 front    insert   join    prefix
 rubout   send     split
---------------------------------------------------
           TARGET SENTENCE

 a stitch in time saves nine
---------------------------------------------------
 sti tch in a time xpqy aves nine
 sti tch in a time aves nine
→ a sti tch in time aves nine                     ←

---------------------------------------------------
         Type your next command
join_
```

Figure 1. Example display during proverb solution [specific names and supergoal target].

version. Eight of the commands performed manipulations on the proverb itself. For any one proverb, a minimum of four of these were required to attain the correct version. Non-optimal solutions involving more operations were possible. The two remaining commands were CANCEL and HELP. The CANCEL operation revoked the current or the preceeding command step. Thus, users could return to a prior state along their solution path. The HELP command was used to display a command menu or, if the menu was already shown, to obtain full definitions of the command operations. The command menu automatically disappeared upon completion of a command step. Thus, users had to actively request either the menu or the full definitions whenever they required more information for determining what to do next. In the example of Figure 1, the subject must have invoked help once prior to typing "join" as the command menu is shown.

A fortnight later, the subjects returned for a further session. They had not been told the purpose of the second visit. This session consisted of (1) a memory test, (2) two tests to assess individual differences in the user sample and (3) the correction of four more proverbs using the same editing system. During the memory test, the subject first recalled as many of the command names as they could, without any assistance. They were then asked to write down a description of the operation associated with each command name. Next, the subjects' recognition memory was tested. They were presented with a list of thirty-six names and asked to indicate the twelve which had made up the system they had used a fortnight earlier, and how confident they were of each choice. Finally, given the twelve names they had marked, they were asked to describe any further operations they could recall.

Two individual difference tasks were then administered. In the first task, the Object and Function Questionnaire (OFQ), they were shown a sequence of 48 definitions from which target words were to be derived. For example, the definition

"To make up one's mind, to resolve or settle a question" required the target "to decide". Targets were either nouns or verbs, concrete or abstract and of high or low frequency. There were six definitions in each of the eight sub-conditions resulting from combinations of these variables. The example above is a high-frequency abstract verb. The aim of this task was to produce an independent measure of the ease with which people could derive words from an appropriate semantic description. Skill at this kind of task was considered particularly relevant to the similar problem of mapping the meaning of a command operation onto a required command name. The second task was developed by Broadbent et al. [1] to assess self-reported failures in everyday life of perception, memory and action. This task, the Cognitive Failures Questionnaire (CFQ), requires answers to 25 questions concerning how frequently the respondent forgets things like names, appointments or what they are currently doing, as well as aspects such as their distractibility and clumsiness. Again, cognitive failures of these types might well be related to the individual's ability to learn and remember computer commands.

After completing the memory assessment and the two individual difference tasks, users corrected a further four proverbs using the interactive system. All users received the same type of command names as they had encountered in the first session. However, all users corrected the proverbs in session 2 under the supergoal version of the task. This permitted at least a partial assessment of the longer-term influence of training in session 1 under constrained [subgoal] or free [supergoal] solution paths.

RESULTS

Analyses of variance were carried out on measures of time and on command usage. In the case of session 1 these analyses were confined to proverbs 2-8 since an experimenter was available to provide advice and encouragement during solution of the first proverb. The extent of a user's knowledge of commands and their operations was assumed to be reflected in the frequency and duration of consultations of the HELP facilities [HELP 1 - command name menu; HELP 2 - command names plus descriptions of the operations]. In addition, inappropriate or unintentional consequences of commands were assumed to be reflected in the use of the CANCEL command [although other operations, notably insertion and deletion, could also be used to the same effect]. The same measures were correlated with individual difference scores and the data from the memory tests were analysed using conventional measures of accuracy. The present paper will provide information on general characteristics of performance and the effects of the learning task. The principal effects of type of command name, recall and individual differences will then be summarised. The full range of data will be reported in detail elsewhere.

General Characteristics of Performance and Goal Effects

Several aspects of the data can be briefly described. First, on all measures performance generally improved across the course of session 1 and during session 2. Second, the patterns of command usage [excluding HELP and CANEL] were similar across conditions. Although six commands were required in an optimal solution, users, on average, issued around ten commands to solve the first two proverbs of session 1. For the last two proverbs of session 1 this reduced to around seven or eight commands. The different commands also varied significantly in their frequency of use. However, this was entirely attributable to disproportionate use of the Rubout and Insert operations. In all conditions these commands were issued at least twice as often as the other commands which did not differ among themselves when scaled against optimal usage.

The effects of the two types of task condition used in session 1 [subgoal and supergoal] can also be summarised briefly. This manipulation did not influence most measures of performance during session 1. Nor did the initial learning task influence most measures of subsequent performance in session 2 where only the supergoal condition was employed. Two findings are nevertheless worth noting. First, during session 1 task conditions did not influence the frequencies with which users consulted HELP 1 or HELP 2. However, when HELP 1 was consulted [command name menu] the subgoal condition led to a greater frequency of moving directly on to enter a command in session 1 than did the supergoal condition [$F(1,44) = 4.56$, $p < .05$, see Figure 3]. At this stage the alternative course of action is to consult HELP 2. Second, for both session 1 and session 2, users initially trained in the subgoal condition were more likely to issue CANCEL commands to "undo" the effects of a prior operation [$F(1,44) = 4.87$ for session 1 and 4.41 for session 2; $p < .05$ in both cases].

Although the subgoal condition provided constraints likely to increase the use of CANCEL to achieve the required individual steps of a solution, the evidence suggests that an impulsive strategy of "operate-cancel if inappropriate" was more likely to be established in the subgoal condition. The effect in session 2 indicates that aspects of this strategy carry over to session 2 where the initial task constraints were removed. Most importantly, the effects of initial training in the subgoal and supergoal tasks in no way interacted with effects of the different types of command name. Accordingly, the following effects of command name hold for both types of initial training.

The Influence of Command Names on Interactive Performance

On several measures, interactive performance followed a similar pattern in both session 1 and session 2. These measures included the total time taken to "solve" a proverb, the frequencies of consulting HELP 1 and HELP 2 and the frequency of entering a command following consultation of HELP 1 [the command name menu].

The total times taken, on average, to solve each proverb are shown in Figure 2. In session 1 the only statistically reliable effect involved training [trials]. The overall time taken was rather longer for the general names [243.5s] than for the specific ones [209.5s] but this difference was not statistically reliable. In session 2 the times taken were significantly longer overall with the general names [$F(1,44) = 4.19$, $p < .05$]. However, this difference changed across trials [$F(3,132) = 4.31$, $p < .01$]. The figure indicates that the advantage for specific command names is greatest for the solution of the first proverb of session 2.

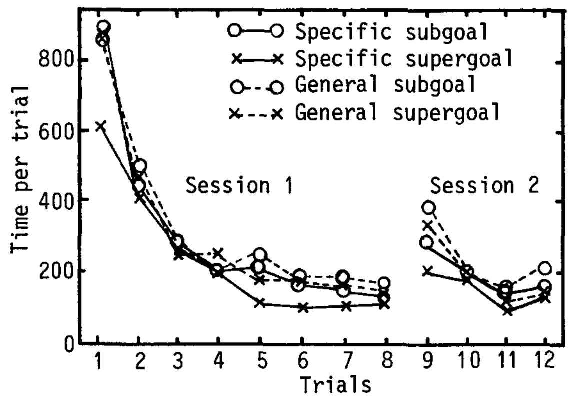

Figure 2. Mean times taken to solve each proverb.

The frequencies of consulting HELP 2 are shown in Figure 3. Overall, users required the definitions of HELP 2 very much more often with the general command names than with the specific names [$F(1,44) = 14.45$, $p < .001$ in session 1; $F(1,44) = 12.6$, $p < .001$ in session 2]. The advantage for the specific names also diminished in both sessions as trials progressed [$F(6,264) = 4.72$, $p < .001$ in session 1; $F(3,132) = 7.22$, $p < .001$ in session 2]. The frequencies of consulting HELP 1 across trials showed similar effects with HELP 1 being used more frequently with general names than with specific names.

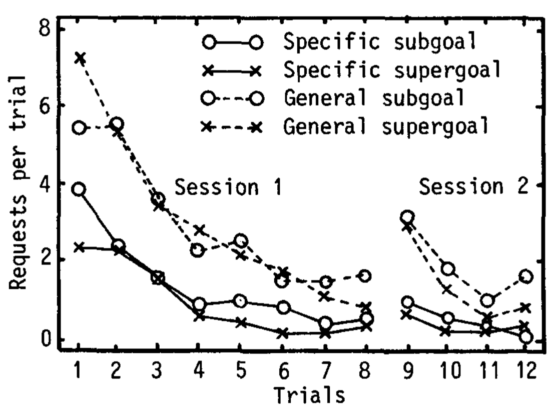

Figure 3. Mean frequency of use of HELP 2 during solution of each proverb.

One problem is that it was necessary to access HELP 1 en route to consulting HELP 2. In consequence, these data included instances where users accessed HELP 1 and then entered a command as well as instances where HELP 1 was consulted en route to HELP 2. Figure 4 shows those cases where access to HELP 1 was followed directly by entry of a command. As outlined earlier, command entry was more frequent following use of HELP 1 in the subgoal conditions than in the supergoal conditions. However, the type of command name had no overall effect in either session [$F < 1$ in both cases]. There was nevertheless a significant interaction with trials in both sessions 1 and 2 [$F(6,264) = 2.36$, $p < .05$]; $F(3,132) = 5.36$, $p < .002$ respectively]. Figure 4 suggests that direct command entry following use of HELP 1 was greater for the specific terms at the beginning of each session and reduced more rapidly with the specific names than with the general ones.

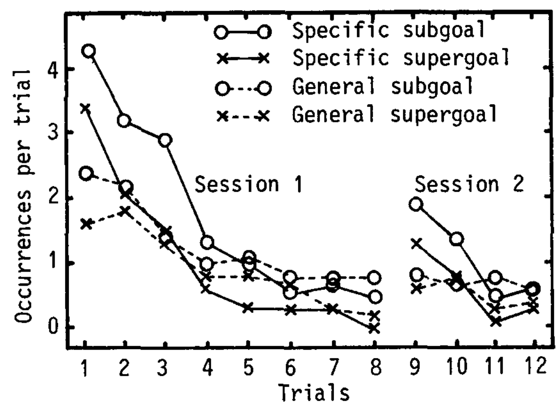

Figure 4. Mean frequencies of command entry following use of HELP 1 during solution of each proverb.

Since the use of the HELP facilities suggested that users of the general names required more information prior to command entry than users of the specific names, the times spent in the various states were analysed in more detail. These analyses (for session 1 only) showed that the total times spent in each state roughly mirrored the frequency with which a state was entered. However, when the mean times spent in each state were computed an interesting pattern emerged. First, for the mean times to enter a command without any recourse to HELP, there were no significant differences between the various conditions. In contrast, if the command name menu was requested [HELP 1], then users of the specific names took rather more time to issue this request than did users of the general names [$F(1,44) = 3.55$, $p = .066$]. Although this effect was of marginal significance, the same pattern held when considering times for the actions from the HELP 1 state itself. From here users could either type a command or request to see the full definitions. Again, users in the specific and general groups took an equal time to decide to enter a command, whereas the users of specific command names took more time when deciding to request the full definitions of HELP 2 [$F(1,44) = 7.30$, $p = .01$]. In addition, although

HELP 2 was consulted more often by the users of general names, the users of specific names actually took more time for each consultation [F(1,44) = 15.21, p < .001]. Subsequent to use of HELP 2, the time taken to enter a command was also greater with the specific names [F(1,44) = 15.27, p < .001].

Thus, in terms of overall frequency of consulting HELP and the summed times spent viewing it, the users of general command names required more information than the users of specific names. Furthermore, it seems that users of the general names acknowledged this and adopted a strategy of consulting help frequently and without delay. In contrast, when the users of specific terms were unsure, they tended to take more time before deciding to consult help or, after using HELP 2, over entering a command. Under circumstances where users possessed sufficient knowledge to proceed without HELP, command entry times were comparable.

Memory for Command Names and Operations

A fortnight after initial learning and prior to session 2, users in three of the conditions were able to recall approximately one third of the command names. In the fourth, specific names under the supergoal condition of initial learning, recall rose to an average of 52% but this effect was not reliable. Likewise, there were no reliable effects of the various conditions on the overall accuracy of recognition memory for command names. All conditions gave rise to around 80-85% accuracy in recognition. However, a greater proportion of the specific names were recognised with a high confidence level [F(1,44) = 7.26, p < .01].

In contrast, memory for the meanings of commands was clearly more accurate with the specific names than with the general names. As scored by three independent raters, recall of the operations performed was superior with the specific names both after initial free recall of the command names [F(1,44) = 7.88, p < .01] and including those additional items recalled after the recognition test [F(1,44) = 5.47, p < .025]. Total accurate recall for the description of the ten major operations [excluding HELP and CANCEL] averaged 7.96 for the specific terms and 6.58 for the general terms. Finally, users in the supergoal condition recalled more of these descriptions accurately (5.38) than users in the subgoal condition (3.88) immediately after initial recall of the command names [F(1,44) = 7.88, p < .01]. However, for the subgoal condition a greater number of descriptions were added after the recognition test [F(1,44) = 9.22, p < .01] and hence there was no difference in total recall for the operations as a function of goal condition (supergoal = 7.42; subgoal = 7.13).

Individual Differences

Scores from the two individual difference tests were correlated [Pearson Product Moment] with performance data normalised to compensate for differences between experimental conditions. Of particular relevance to this presentation are relationships between the measures of cognitive style and use of the HELP facilities in the interactive task. Users with high scores for correctly deriving words from their definitions [OFQ] consulted the HELP facilities less frequently than users with lower scores for correct derivations. [E.g. for overall mean correct on OFQ and mean use of HELP 2; r = -.382, p < .01 for session 1; and r = -.327, p < .05 for session 2; df = 46]. Broadbent's index of cognitive failure was also related to the use of the HELP facilities, although the correlations were significant only for the performance data for session 2 [session 1, r = -.251, 0.1 > p > .05; session 2, r = -.414, p < .01, df = 46]. Users with relatively high cognitive failure scores consulted level 2 HELP less often than users with lower cognitive failure scores. Possibly individuals with a high cognitive failure score were more impulsive and therefore less likely to capitalise on the available help.

Although cognitive failure and OFQ accuracy scores were both related to the use of command definitions [HELP 2], these two measures of cognitive style were not themselves intercorrelated [r = -.131, p > .1]. In order to establish whether there were two separate influences on the use of HELP 2, the users were divided up according to their scores on the two dimensions. Table 1 provides a breakdown for the mean number of times that command definitions [HELP 2] were consulted (per proverb) during session 2. The numbers of users in each cell are given in parentheses. These data are broadly consistent with two independent factors. For the general names the effects of the two dimensions appear to be roughly additive. With the specific names, users with Low OFQ scores show the same trend but users with High OFQ scores referred to HELP 2 very infrequently irrespective of their CFQ score (i.e. less than once per user in the whole of session 2).

	SPECIFIC NAMES		GENERAL NAMES	
	Low OFQ	High OFQ	Low OFQ	High OFQ
High CFQ	.59(8)	.20(5)	1.46(6)	.71(6)
Low CFQ	1.00(4)	.18(7)	2.71(6)	1.83(6)

Table 1. Frequencies of HELP 2 use per proverb in Session 2.

DISCUSSION

A short paper obviously precludes detailed discussion of particular findings and issues. The more general implications of the whole study can nevertheless be summarised briefly.

First, the effects of the two different forms of command name were systematic. With the general command names, users accessed HELP information more often and after shorter time delays than users of the specific command names. Thus, where the command names provided only minimal clues [general names] as to the nature of the operation performed, users seemed to adopt the strategy of actively seeking more information. Where the command names provided more specific clues users required less information overall before issuing commands. However, at points in the task where they were unsure of what to do, users of the specific terms appeared to take more time over some kinds of decisions.

Second, although users were more confident in their recognition of specific command names, both recall and recognition accuracy did not show a reliable difference between general and specific names. In this context it should be pointed out that the sample size was relatively small given the high level of variability in recall performance. Nevertheless, recall of command operations was superior with specific command names. On the basis of the present data it is not possible to determine the relative contribution of two factors. On the one hand, the operations may have been better recalled with the specific names because more differentiated memory representations were formed in session 1. On the other hand, the specific names may simply have served as better prompts for retrieval than their more general counterparts.

The relative contributions to learning and memory of all these factors obviously require further clarification. However, the broad pattern of data suggests one potentially profitable way of thinking about the issues. Memory for the command operations was poorer with the general names in spite of the fact that users had greater total exposure to the definitions of the operations [via HELP 2] in session 1 than users of the more specific command names. One possibility is that use of the HELP facility constitutes a relatively passive cognitive strategy for learning which leads to less efficient retention than spending time actively considering options. This latter activity appeared more characteristic of performance with the specific terms.

The few effects of the subgoal and supergoal tasks also point to the importance of considering the likely consequences of task demands on cognitive strategies and subsequent retention. Users in the subgoal condition not only appeared to use an impulsive "operate-cancel" strategy than their counterparts in the supergoal condition but also showed inferior unprompted recall for command operations. Thus, the requirement in the supergoal condition for users to determine their own solution path may also have resulted in the use of more active cognitive strategies for learning.

Interestingly, the analysis of individual differences would suggest that differing demands imposed by the various experimental conditions in some sense modulate users' predispositions to adopt particular cognitive strategies. Furthermore, more than one factor would seem to be involved. For example, users who produced more target items on the definitions test (OFQ) used the HELP facility less. Users with a high cognitive failure score were also less likely to use the HELP facility. Thus, less frequent use of HELP might be due to a predisposition towards impulsive action and a failure to capitalise on available information [high cognitive failure]. In addition, less frequent use of the HELP might also reflect a greater tendency to rely on more productive learning strategies (high OFQ accuracy) which lead to constructive mapping of command names onto their meaning.

Taken together, the results of the present study suggest that alternative types of command names are likely to influence performance in a systematic and often subtle manner. In this respect global measures such as total time on task were not the most revealing. In consequence, findings concerning command names which are based mainly on such measures [e.g. 6] are best interpreted with great caution. The potential influences of alternative command names could be pursued more profitably if the kind of demands imposed by particular names are analysed in conjunction with broader task demands as well as the cognitive predispositions of the potential user population.

ACKNOWLEDGMENT

The research reported in this paper was supported by funding from IBM United Kingdom Laboratories Ltd.

REFERENCES

1. Broadbent, D.E., Cooper, P.F., Fitzgerald, P. and Parks, K.R. The Cognitive Failures Questionnaire (CFQ) and its Correlates. British Journal of Clinical Psychology, 1982, 21 (In Press)

2. Carroll, J.M. Names and naming. An interdisciplinary review. IBM Research Report RC 7370, 1978.

3. Carroll, J.M. Learning, using and designing command paradigms. IBM Research Report RC 8141, 1980.

4. Gumenik, W.E. The advantage of specific terms over general terms as cues for sentence recall: Instantiation or retrieval? Memory and Cognition, 1979, 7, 240-244.

5. Hammond, N., Barnard, P., Clark, I., Morton, J. and Long, J. Structure and Content in Interactive Dialogue. Paper presented at APA Montreal, September 1980; and IBM Hursley Human Factors Report HF034, October, 1980.

6. Landauer, T.K., Galotti, K.M. and Hartwell, S. A computer command by any other name: A study of text editing terms. Paper presented at APA Montreal, September 1980.

7. Thios, S.J. Memory for general and specific sentences. Memory and Cognition, 1975, 3, 75-77.