The Quarterly Journal of Experimental Psychology (1986) **38A**, 705–737

The Uses of Short-Term Memory: A Case Study

Brian Butterworth

Psychology Department, University College London, U.K.

Ruth Campbell

Department of Experimental Psychology, Oxford University, U.K.

David Howard

Psychology Department, University College London, U.K., and Speech Therapy Department, Eastern Hospital, Hackney

It has been widely claimed that the systems employed in tasks of immediate memory have a function in the comprehension of speech; these systems, it has been proposed, are used to hold a representation of the speech until a syntactic analysis and interpretation have been completed. Such a holding function is meant to be especially important where the sentences heard are long or complex. It has thus been predicted that subjects with impaired short-term memory performance would show deficits in comprehension of such materials.

In this study, one subject with impaired phonological processing and a severely reduced digit span was tested on a range of tasks requiring the syntactic analysis, memory and comprehension of long and complex material. She was found to be unimpaired on syntactic analysis and comprehension, but not on sentence repetition. The implications for models of short-term memory are discussed.

Requests for reprints should be sent to Brian Butterworth, Psychology Department, University College London, Gower Street, London WC1E 6BT, U.K.

These results were reported to the Experimental Psychological Society at their London meeting in January 1985. We are grateful to members of the Society for the opportunity to discuss them, and for helpful comments and suggestions. Tom Leuchars, Cathy Price and Rachel Spence helped us to run the control experiments. We particularly thank RE for her patience in undergoing these experiments. Ruth Campbell and David Howard were supported by separate grants from the Medical Research Council.

GENERAL INTRODUCTION

In this study, we re-examine the role of phonemic processing in carrying out short-term memory (STM) tasks, and the role of short-term memory in the memory and comprehension of prose, in a subject, RE, who has demonstrable impairments on tasks of both phonemic processing and STM (see Campbell and Butterworth, 1985).

There is extensive evidence that normal performance on short-term memory tasks is mediated phonemically. Items whose names are phonemically similar are recalled worse than those that are phonemically dissimilar; this holds true for visual as well as for auditory presentation of linguistic materials (Conrad, 1964) and for pictures (Conrad, 1972). Neuropsychological evidence supports these results: patients with a phonemic impairment also show reduced STM span. JS, a patient reported by Caramazza, Berndt and Basili (1983), suffered an impairment of phonemic processing as evidenced by an inability to carry out rhyme judgements; he was unable reliably to select the nonword that sounded like a word from the set of letter strings containing visual distractors—like *Leef*, *Leab*, *Luaf*. Concomitantly, there was a severe impairment in serial recall of unrelated visually presented words. His STM difficulties extended to probe memory tasks as well: recognition times for a probe from a four-word presentation was not only worse than controls, but, also unlike controls, showed no relative slowing for phonological distractors. Another patient studied by Caramazza and his colleagues (Caramazza, Basili, Koller and Berndt, 1981), MC, whose phonemic ability was not tested directly but who showed impairments on tasks widely believed to involve phonemic processing—like nonword reading and word repetition—showed severely impaired performance on a range of STM tasks. Caramazza et al. suggests that the particular loss of order information and the absence of normal serial position effects was due to the abnormal functioning of phonological encoding. Three patients described by Allport (1984), who could not discriminate reliably between speech sounds nor detect mispronunciations in words, had a concomitant deficit in immediate and delayed serial recall and in matching span.

On the other hand, impairment in phonemic processing does not seem a necessary condition for a STM deficit. Vallar and Baddeley (1984a, 1984b) report a patient, PV, who has a severely restricted memory span, but who shows normal, usually errorless, performance on discrimination of natural speech sounds, on assigning stress to written words and making rhyme judgements.

The involvement of phonemic processing in STM tasks is generally construed as the coding in which material to be remembered is held in a single short-term store (STS). The locus of this store in relation to other language processes has not always been made clear. Waugh and Norman (1965) and Atkinson and Shiffrin (1968) locate the STS at the input end of language and memorial processes; they are not explicit about its relation to lexical access, but the superiority of word over nonword recall requires that it must be postlexical. In the original "working memory" model of Baddeley and Hitch (1974), the STS component is part of a central executive, mediating a number of processes, although the storage component itself would presumably have to be an output buffer with respect to lexical access in order to explain word superiority effects in STM. This store is supported by an "articulatory loop" that can hold and rehearse material made available by the central executive. In Morton's (1970) model the STS, called the "response buffer", also has an output locus and is identified with the buffer holding words to be spoken in spontaneous speech or in reading texts. The contents of the response buffer, which are phonologically coded, may be either postlexical—i.e. the output of the logogen system—or nonlexical—via the translation of graphemes into phonemes, or through direct repetition of heard speech.

Although experimental studies of normal subjects have not been successful in distinguishing between these models, neuropsychological evidence has pointed clearly to an input locus with respect to postlexical language processes for the phonologically coded STS. Shallice and Butterworth (1977) reported a patient whose spontaneous speech was unimpaired while her performance on STM tasks showed striking deficits, demonstrating that the defective STS cannot be routinely involved in speech production. Moreover, STM patients regularly showed evidence of other deficits attributable to a deficit of an input STS (Shallice, 1979, for a review; and see below); indeed, the most recent reformulation of the working memory model accepts that the STS has an input locus in that, whatever else the central executive might have on hand, auditory material is held to have "obligatory access" to it (Salamé and Baddeley, 1982).

One general problem with these accounts of STM performance has been to explain what role the STS might play outside the confines of laboratory tasks. It seems ecologically implausible to postulate a structure or process whose sole function is to mediate the immediate recall of strings of unrelated items. Morton's attempt to link it to speech output processing was an early recognition of this problem.

Work on sentence comprehension in normal populations has attributed to STS a role in sentence comprehension, either as an obligatory and passive first-stage buffer holding phonologically coded material for subsequent lexical, syntactic and semantic processing (Clark and Clark, 1977; Garrett, Bever and Fodor, 1966), or where surface structure needs to be retained for some reason (Baddeley, 1979). In Baddeley's original formulation, order information was held in the articulatory loop, hence comprehension of sentences where word order was critical to interpretation (e.g. reversible passives) depended on the correct functioning of the loop. Wingfield and Butterworth (1984), on the basis of an experiment described below, argue that STS is not just a fixed-capacity receptacle for input, but rather its contents are actively controlled by online syntactic analysis. Another suggestion is that the STS contains not only the words in an input sentence, but abstract syntactic labels of the constituents; this store also mediates performance on span tasks (Savin and Perchonock, 1965; Wanner and Maratsos, 1978).

Neuropsychological investigations have so far supported the link between STM and speech comprehension: patients with poor STM performance show impairments in the comprehension of sentences, especially where these are long, syntactically complex or parsable only by full syntactic analysis of the surface structure. Thus, Shallice, reviewing the neuropsychological evidence, writes:

> There are functional reasons for postulating that a sizeable store should exist to retain the surface structure of speech in case the initial parsing falls behind in real time or fails completely as in Lashley's (1951) writing/righting example, so that backtracking becomes necessary . . . that surface structure (in contrast to gist) is retained in a short-term store is a well-known finding (Johnson-Laird and Stevenson, 1970). That such a store would play a major role in span performance has been suggested many times. (1979:270)

Vallar and Baddeley (1984a) similarly maintain that "the phonological short-term store is useful for the comprehension of long sentences with a complex syntactic structure, containing too much information to be processed during presentation. Under these conditions the phonological short-term store holds the sentence while the subject processes it." (126). And, indeed, their patient, PV, who shows a deficit in STM, also shows relatively intact comprehension of short sentences, but impaired comprehension of long sentences. In a similar vein, Caramazza et al. (1983) write, of their STM patient, MC, "Comprehension of these [long] sentences is dependent on the normal functioning of phonological working memory" (1983:160), while Allport (1984) concludes a discussion of the relation between STM and comprehension more cautiously: "a defect of verbal repetition span could be associated also with impaired encoding at a lexical and/or semantic level of representation."

The patients JS and MC reported by Caramazza and his colleagues show impairments in the comprehension of sentences needing syntactic analysis or order information (Caramazza et al., 1981, 1983). IL, a patient with a STM impairment, studied by Saffran and Marin (1975), has particular difficulty with sentences, like reversible passives, needing full syntactic analysis, or at least the retention of constituent order information, for their proper interpretation. And Shallice, in his review of this issue, claims that "subtle comprehension deficits exist for ALL STM patients studied in this respect. Thus our patients had a deficit on the Token test of De Renzi and Vignolo (1962) in which patients have to obey instructions containing much nonredundant information such as, 'Put the red circle between the yellow triangle and the green triangle'" (1979:270).

Note that, in different patients, rather different kinds of comprehension disorder are attributed to the STM impairment. JS, MC and IL are all said to have difficulty with simple reversible active or passive sentences (e.g. "the cat was chased by the dog"); Vallar and Baddeley's patient, PV, had no difficulty with sentences of this type. Shallice and Warrington's patients have difficulty with long but syntactically simple materials in the Token Test. In contrast, PV scored only one point below the normal cut-off on a shortened version of this test; PV's difficulties are said to emerge only when she is tested with long sentences whose constituent order has been subtly changed. It is not clear why these differences between patients exist. It is possible that they reflect differences between the patients in the severity of the STM impairment. The differences could be due to differences in the way in which auditory sentence comprehension is tested. Or, it may be that some or all of these patients have difficulties in sentence comprehension which are not related to their deficits in STM tasks. The information available does not allow us to discriminate between these three possibilities.

In summary, it would seem that an impairment to phonological processing entails a deficit in the functioning of the hypothesized single STS, and hence on standard STM tasks, but not vice versa. Hence a plausible *first working hypothesis*: intact phonological processing is necessary, but not sufficient, for the normal functioning of the system mediating performance on STM tasks, perhaps because accurate phonological encoding is necessary to set up the representation held in the STS.

Moreover, it has been claimed that impaired STS entails a deficit in the comprehension of at least some kinds of linguistic material. The critical theoretical claim made by the authors cited above is that comprehension of complex sentences depends on the listener's ability to hold a literal representation of the sentence (or at least its surface structure) until the grammatical analysis of the sentence and its interpretation have been completed. Thus our *second working hypothesis*: the normal functioning of the STS is necessary, although not sufficient, for sentence recall and sentence comprehension. It follows as a corollary that intact sentence recall is necessary, but not sufficient, for sentence comprehension. In particular, the appropriate structural analysis of word order information will depend on the listener maintaining a verbatim version of the sentence. Thus, where STM list tasks show order errors, then order errors should be found in sentence recall, and sentence comprehension should show evidence of incorrect structural analyses. In other words, a deficit in STS functioning is sufficient for a deficit in both sentence recall and comprehension, at least where order information is critical.

To assess the validity of these two hypotheses and the corollary, we carried out a further series of tests on a subject, RE, whom we had previously shown to have impaired performance on some phonological tasks and on some STM tasks. A full case description and report of the tasks can be found in Campbell and Butterworth (1985).

In brief, RE is a 23-year-old woman with no history of neuropsychological disorders and with apparently normal speech and hearing. She has achieved 8 "O" levels, 3 "A" levels with good grades (required for university entrance in the U.K.), a BA honours degree in Psychology and certificates of proficiency in music (Associated Music Board, Grade IV or above, in piano, flute, recorder and violin), which indicate competent sight reading and musical memory. Her reading and spelling of real words are within the normal range for her age and educational level, although Campbell and Butterworth (1985) show that she has considerable difficulty in reading and spelling even simple nonwords (see next section).

Phonological Processing

Although peripheral auditory discrimination of synthesized speech sounds is entirely normal, she presents symptoms of a more central phonological disturbance. She performs poorly on rhyme judgements presented aurally or visually; her ability to segment words into phonemes is impaired as tested on a variety of tasks—like counting phonemes, exchanging the initial phonemes of two names, and "auditory acronyms", where the first phoneme of each of three words is to be combined to make a new word or pseudoword; and she offers the introspective report that she cannot "hear words in her head".

Reading and spelling tasks that require the generation and analysis of word sounds are also poorly performed. Thus, she is worse than normal controls on oral reading of pseudowords, which depends on the ability to segment words into phonemes in order to establish grapheme–phoneme correspondences. She is also worse on silent reading tasks involving judgements of word sounds, such as homophone and pseudohomophone detection. Her spelling errors are, unlike those of our undergraduate controls, often phonologically implausible (e.g. for "terrestrial", she wrote *terrisetrial*, whereas the modal mispelling by our controls was *tirrestrial*). We describe her condition as "developmental phonological dyslexia and dysgraphia" and hypothesize that a central phonological deficit underlies these disturbances.

On a range of tasks that can only be performed accurately on the basis of a phonological representation, RE relies exclusively upon visual orthographic strategies. Thus, in phoneme counting she is confused when the number of phonemes is different from the number of letters; on rhyme judgements she will say that *lemon* rhymes with *demon*, but that *rough* does not rhyme with *fluff*; when asked to exchange the initial phonemes of "Phil Collins", she offered "Chill Pollins", exchanging the initial letters, instead of the correct "Kill Follins"; and on auditory acronyms, she will invariably compose the target from the initial letters not the initial phonemes (e.g. "chair-otter-pail"→"cop", and not "chop"; "sing-able-gentle"→"sag", not "sage").

However, while these tests all show that RE is bad at using phonological information overtly, we have not yet demonstrated that she is sensitive to experimental manipulations of heard speech. Allport's patients, for example, were all impaired at detecting mispronunciations—single phoneme alterations—in words, which, he suggests, indicates that phonemic representations were impaired and hence constitutes a sufficient precondition for failures in overt phonemic manipulation.

In this paper we report RE's ability to detect mispronunciations as a direct indicator of her capacity for accurate input phonemic registration; we also report on her ability to reproduce rare multisyllabic words as a test of the intactness of her phonemic output capacity.

Her performance on a range of STM tasks is not only substantially worse than normal, with serial recall and matching spans of at best 4 digits correct, it is also different in kind. The following results constitute powerful evidence of the exclusive use of an orthographic or visual strategy:

1. Visual presentation is better than auditory presentation.
2. There are no recency effects or suffix effects, which are also signatures of phonemic coding.
3. Forward span is no better than backward span.

4. There is no effect of phonemic similarity in the serial recall of letter strings.
5. Matching span with eyes closed is better than with eyes open.
6. In this task we introduced the catch trial "nought zero nought nought/nought nought zero nought". She was unable to detect that these two lists sounded different. However, further catch trials of this sort were ineffective, for RE said that she recoded "nought" and "zero" into different orthographic forms.

These data are consistent with our first hypothesis—namely, that intact phonological processing is necessary for normal STM performance. Moreover, impaired phonological processing has led to a qualitatively different way of handling STM tasks.

According to the working memory framework of Baddeley and Hitch (1974), word length effects are mediated by an articulatory rehearsal loop, which can maintain about 1.5 sec of processed phonological or articulatory material, leaving more capacity available for other processes. In our previous study of RE, we could find no evidence of a word length effect. One possibility, therefore, is that RE is unable to use the loop to store words, and hence her overall STM capacity will be that much reduced. Concurrent articulation is held to pre-empt the articulation process, so that the loop cannot be deployed in STM tasks; this results, in normal subjects, in reduced span. Baddeley and Hitch further claimed that the loop has a role in sentence comprehension, as performance here deteriorates with concurrent articulation. If RE does not use the loop, then this manipulation should have no effect on her already poor STM performance. But such a finding would predict that RE's sentence recall and sentence comprehension would be impaired, especially for materials requiring the exact maintenance of the surface structure. That is to say, it would support the second hypothesis, about the role of STS in comprehension, and locate the functional deficit more exactly.

To evaluate these proposals, six tests involving the syntactic analysis of complex sentences and their comprehension were employed. (1) The first is the Token Test (De Renzi and Vignolo, 1962); this is a standard clinical test of sentence comprehension, which has been administered to many of the STM subjects already reported. Written, as well as spoken, presentation is available, so that concurrent articulation can be used to assess the particular involvement of the articulatory loop in translating graphemic input and maintaining it as a phonemic representation, which, Salamé and Baddeley (1982) claim, is necessary for deriving order information from sentences. (2) The second test was Bishop's TROG (Test for the Reception of Grammar), which involves matching sentences of increasing complexity to one of four carefully matched pictures; sentence presentation can be spoken or written (Bishop, 1982). (3) The third test is a more difficult test of sentence comprehension using a sentence–picture verification procedure. (4) The fourth test requires judgements of grammaticality of long and complex word strings, with both spoken and written presentation. (5) The fifth test assesses RE's performance in immediate recall of grammatically well formed versions of the same long and complex sentences. (6) The sixth test is one of running text memory, using a procedure described by Wingfield and Butterworth (1984). Essentially, this is a text repetition task, but the span of text to be repeated is under the subject's control; it therefore allows us to see whether RE employs a strategy to reduce the load on the STS. To perform at normal levels and in the normal way, syntactic analysis and comprehension of the text is required. In one condition, the usual intonational cues to text structure are eliminated, forcing the listener to carry out more online parsing. This should put further strain on an impaired working memory system.

Our second working hypothesis predicts that RE would show abnormal and poorer performance on all six tests as compared with our control subjects, and would be particularly poor in maintaining constituent order information for use in recall and comprehension of sentences.

FURTHER TESTS OF PHONEMIC PROCESSING

Mispronunciation Detection

To assess whether, despite good discrimination of speech sounds, RE suffered an impairment of phonological lexical access, we auditorily presented her with 15 names of well-known personalities, each containing one consonant mispronounced by one feature—e.g. "Michael Heselti*me*" for "Heseltine" (the former British Defence Secretary). She made only one error on this task.

Immediate Reproduction of Long Words

We asked RE to repeat 22 long polymorphemic words, containing derivational suffixes and prefixes, like "parasitically", "underhandedness". She gave correct pronunciations for all of them. This again indicates excellent auditory access to lexical items and unimpaired pronunciation of them. In our previous study, reproduction of long pseudowords was good, though not entirely perfect.

These tests demonstrate that her receptive phonological capabilities, at least to the level of lexical access, are intact, and that she has no output problem for familiar multisyllabic sequences.

TESTS OF SENTENCE RECALL AND COMPREHENSION

Token Test

As Shallice (1979) has pointed out, all STM patients so tested have shown impaired performance on the Token Test (De Renzi and Vignolo, 1962). We tested RE and 10 matched controls on Parts 4 and 5 of the Token Test. This requires the subject to manipulate cardboard tokens according to complicated instructions, for example:

"Touch the large white circle and the small green square."
(A complex instruction in a simple syntactic form.)

"Put the red circle between the yellow square and the green square."
(A simple instruction in a complex syntactic form.)

The materials were presented under three conditions: aurally, visually and visually with concurrent articulation. This last condition was a further test of the loop hypothesis discussed above. As can be seen in Table I, RE scored slightly better than the control means in all but one condition: with concurrent articulation, on Part 4, the controls

Table I

Performance[a] on the Token Test[b]

		Conditions		
			Written	
		Spoken	Silent	Concurrent articulation
Part 4	RE	100	95	95
	Controls	97	92	75
Part 5	RE	100	100	95
	Controls	98	94	90

[a] In percentages.
[b] De Renzi and Vignolo, 1962.

performed worse with than without concurrent articulation ($t = 6.8$, $p < 0.01$); RE, on the other hand, was unaffected. This difference is significant.

She was given the same set of sentences for auditory–verbal repetition. She repeated correctly 0.80 of the sentences from Part 4 and 0.76 of the sentences in Part 5 of the Token Test; she had made no errors in comprehension of these same sentences when they were auditorily presented. Four of her errors were within class substitutions (colour substituted for colour or shape for shape; e.g. "put the green square beside the red circle"→"put the green square beside the blue circle"; the other three errors were misordering of lexical items within the sentence (e.g. "before picking up the yellow circle, pick up the red square"→ "before picking up the red circle, pick up the yellow square". In this task RE is able to demonstrate correct comprehension of sentences that she cannot repeat correctly.

TROG—Test for Reception of Grammar

This test (Bishop, 1982) requires the subject to match a sentence to one of four carefully selected coloured line drawings. For example,

"The boy chasing the horse is fat."
Pictures: horse chasing fat boy
fat horse chasing boy
boy chasing fat horse
fat boy chasing horse

"The boy the dog chases is big"
Pictures: boy chases big dog
big dog chases boy
big boy chases dog
dog chases big boy

The sentence types are graded for difficulty. Two tests were carried out, the first with spoken presentation, and the second, several months later, with written presentation. On the final 52 items in the test, RE scored 51 out of 52 for spoken sentences and all correct for the written sentences. This test has not been standardized for adults, but an all-correct performance would be at the 99th centile for 12–13 year olds.

Sentence-Picture Verification

The third sentence comprehension task was designed to be rather more demanding: deciding whether a sentence is a correct description of a given picture. All the sentences involved two clauses: one was of the form NP is PP (e.g. "The horse is above the circle"), the other was a reversible sentence—NP precedes/follows NP—which could be in the active or the passive form. One clause was embedded in the other as either an object or a subject relative clause. There were two prepositions—"above"/"below"—and two main verbs—"precedes"/"follows". With all permutations of these factors, 64 distinct sentences were generated; for example:

"The bus is preceded by the train which the triangle is below."
"The car which is below the triangle follows the train."
"The dog which the circle is above follows the bus."
"The train is preceded by the horse which is above the square."

Each sentence had a matching picture. This had two objects facing to the left, one in front of the other, in the middle of the page. Each of them had a shape above and a shape below; the same four shapes occurred in each picture (circle, square, triangle, star), but there were two different objects (from the set horse, cat, dog, train, bus, car) in each picture. For 24 items the picture was accurately described by the sentences. The mismatches differed in one of three ways (12 mismatches of each type):

Lexical mismatch: one of the items in the sentence was replaced by a different object from the same semantic category. E.g. "The bus which the triangle is above is followed by the cat" with a picture of a dog following a bus which is below a triangle.

Verb mismatch: the sense of the precede/follow verb was reversed. E.g. "The circle is above the car which is preceded by the cat" with a picture of a cat preceding a car with a circle above it.

Preposition mismatch: the sense of the locational preposition was reversed. E.g. "The star is below the car which is preceded by the bus" with a picture of a bus preceding a car which is below a star.

Correct comprehension of these sentences which are multiply reversible depends critically on knowledge of constituent order. If, as a consequence of her STM deficit, RE has difficulty in using knowledge of constituent order in comprehension of spoken sentences, she should find this verification task involving syntactically complex reversible sentences very difficult.

On the 64 sentences, RE made 3 errors. Ten Oxford women undergraduates served as control subjects; they made an average of 1.3 errors, with a range of 0–5. In this sentence comprehension task RE is well within the normal range: she has no difficulty in understanding syntactically complex, reversible sentences presented auditorily.

Syntactic Judgements of Long Sentences

The previous task tests comprehension of one particular type of complex sentence; the syntactic dependencies involved are over relatively short distances. We therefore devised a test of difficult grammaticality judgements, some of which would require the evaluation of long-distance syntactic dependencies. We included a range of kinds of syntactically complex sentences and a variety of different kinds of grammatical violation that might be able to detect specific difficulties in sentence parsing, of the kinds that previous researchers have ascribed to difficulties in short-term memory.

Method

Materials

Salamé and Baddeley (1982) claim that the phonological STS is essential for maintaining information about the *order* of words in spoken or written sentences. We therefore included two types of strings that were deviant because of this misordering of their constituents:

function words were transposed—see (2) below;
centre-embedded clauses were in the wrong voice—for example, passive instead of active (8)

Under Salamé and Baddeley's hypothesis, function word deletions and substitutions will be much easier to detect than function word transpositions—(1) and (3). Also, Caramazza et al. (1981, 1983) claim to have found in their patients with STM and phonological deficits a marked disorder of function word processing relative to nouns and verbs.

Linebarger, Schwartz and Saffran (1983) showed that a group of four agrammatic patients performed with reasonable accuracy in judging the grammatical well-formedness of spoken sentences of a variety of types. Two types of grammatical violation, however, presented them with particular difficulty: errors in reflexive pronouns (7) and errors in tag questions (6). Linebarger et al. speculate that their patients' troubles with these violations may not reflect a grammatical deficit, but rather parsing difficulties due to a restricted phonological STS. We included also errors on grammatical suffixes (4 and 5), since these have been associated with impaired nonword reading and spelling (see Campbell and Butterworth, 1985, for a discussion).

1. *Functor deleted*:
   "Can Caesar's invasion of Britain be described as one of * (the) most brilliant operations ever?"

2. *Functor transposed*:
   "The electricity supply failed because of two wires that *have should (should have) been touching each other but weren't."

3. *Wrong functor*:
   "The backs of chairs were never designed to be suitable places *of (for) hanging fur coats."

4. *Inflected suffix deleted*:
"It was only last week that I was stopped by the police while *walk (walkING) on Hampstead Heath."

5. *Wrong inflexional suffix*:
"The prisoner was involved in some dreadful crimes but has not yet *admittING (admittED) responsibility for his actions."

6. *Wrong tag*:
"Airline pilots should never forget that the safety of their passengers is their paramount concern, *OUGHT (should) they?"

7. *Wrong reflexive*:
"Though I never tell a lie, I might be willing in certain circumstances to perjure *OURSELVES (myself)."

8. *Wrong voice of centre embedded clause*:
"The lion, who *WAS EATEN BY (ate) some raw steak, terrified the child."

The sentences ranged from 14 to 21 words in length, with a mean of 17.6 words; the eight different sentence types were all of equal length.

Procedure

We presented RE and 15 matched controls with 5 sentences in each of 8 categories of ungrammaticality, along with 40 matched grammatical sentences in each of two conditions, auditorily and written. Subjects were asked to say whether the word string presented was wrong in some way and given an example of one type of grammatical error.

Sentences that were grammatical for a given subject in one condition were altered to provide the ungrammatical string in the other condition. Thus, if one group of subjects heard (6) above, the other group would read (6′):
(6′). "Airline pilots should never forget that the safety of their passengers is their paramount concern, should they?"

To check the reliability and stability of RE's performance, we repeated the test several months later, but with the strings originally presented in spoken form now given in written form, and the original written strings presented in spoken form. These tests are denoted below as RE1 and RE2, respectively.

Results

We computed d′ for each type of ungrammaticality for RE and for the controls. These are shown in Table II.

In RE1, RE is able to discriminate the grammatical version from the ungrammatical version for all types of spoken material and falls within the normal range for each type. For written presentation, she performs just worse than chance for Type 2, and her overall discrimination measure is just outside the normal range, though for all types except (2) she is within normal range and shows good discrimination.

In the replication, RE2, performance on all types is within normal

Table II

Accuracy in Detecting Ungrammatical Sentences in d′ *Units*

	Mode of Presentation					
	Spoken			Written		
	Controls (Range)	RE1	RE2	Controls (Range)	RE1	RE2
1. Functor deleted	2.09 (−0.59 to 4.66)	3.17	3.17	2.87 (0.59 to 4.66)	2.08	1.09
2. Functors transposed	3.58 (1.09 to 4.66)	2.58	2.58	2.74 (0.00 to 4.66)[a]	−1.09	2.08
3. Wrong functor	1.94 (0.59 to 3.17)	1.68	2.58	3.25 (1.09 to 4.66)	3.17	1.09
4. Suffix deleted	2.29 (0.00 to 4.66)	4.66	0.00	3.38 (2.08 to 4.66)	4.66	3.17
5. Wrong suffix	2.76 (0.00 to 4.66)	4.66	0.50	2.25 (0.00 to 4.66)	2.08	4.66
6. Wrong tag	1.95 (−0.59 to 3.17)	2.08	2.33	1.27 (−0.23 to 4.66)	1.89	2.58
7. Wrong reflexive	2.52 (0.50 to 4.66)	3.17	3.17	3.62 (1.68 to 4.66)	1.68	1.68
8. Wrong voice	3.05 (0.59 to 4.66)	2.58	2.58	3.37 (1.68 to 4.66)	2.08	3.17
Overall *d′*	2.86 (1.66 to 3.56)	2.96	2.30	2.99 (2.26 to 3.38)	2.14	2.62

[a]0.00 is chance performance, 4.66 is perfect performance

range, and overall discrimination measures are within normal range. However, both she and one control perform at chance on strings containing deleted suffixes auditorily presented.

This test turns out to be quite difficult even for our controls, with some performing close to or below chance for most types of strings (as can be seen from Table II). RE's performance is entirely comparable to the controls and shows no consistent or particular difficulty with some kinds of material or with one mode of presentation.

Sentence Repetition

With the token test materials, RE was able to demonstrate normal comprehension of sentences that she could not repeat accurately. In the syntactic judgements task, RE was no worse than control subjects in rejecting syntactically deviant sentences. This test investigates whether she is as good at repeating them as control subjects.

The materials were 40 sentences from the syntactic judgements test presented in their correct form. Their length ranged from 15–21 words (mean 17.6). The sentences were presented auditorily for immediate oral repetition. Ten Oxford women undergraduates acted as controls.

The results are presented in Table III. Errors are classified into word omissions, substitutions, additions and order errors (misordering of single words or constituents). There is considerable variation among the controls. The best subject repeats 83% of the sentences correctly and makes only 11 errors, while the worst is correct with only 37% of the sentences and makes 84 errors. In terms of the number of sentences

Table III

Repetition of Long Sentences

		Controls[a]		
	RE	Mean	SD	Range
Correct repetition (maximum 40)	11	24.6	6.9	15–33
Errors				
Word omissions	31	13.1	8.7	4–29
Word substitutions	36	16.3	10.5	4–37
Word additions	12	4.8	4.6	0–16
Order errors	2	0.9	0.9	0–3
Total errors	81	35.1	22.7	10–84

[a] $N = 10$.

repeated correctly, RE is worse than any of the controls; she is outside the control range in the number of omissions, but (just) within the control range for the numbers of substitutions, additions and order errors. RE's total number of word errors is just better than the worst control. The relative proportions of different types of error in RE's performance is almost identical to the control means; thus, although she makes quantitatively more errors than controls, the pattern of different error types is qualitatively normal.

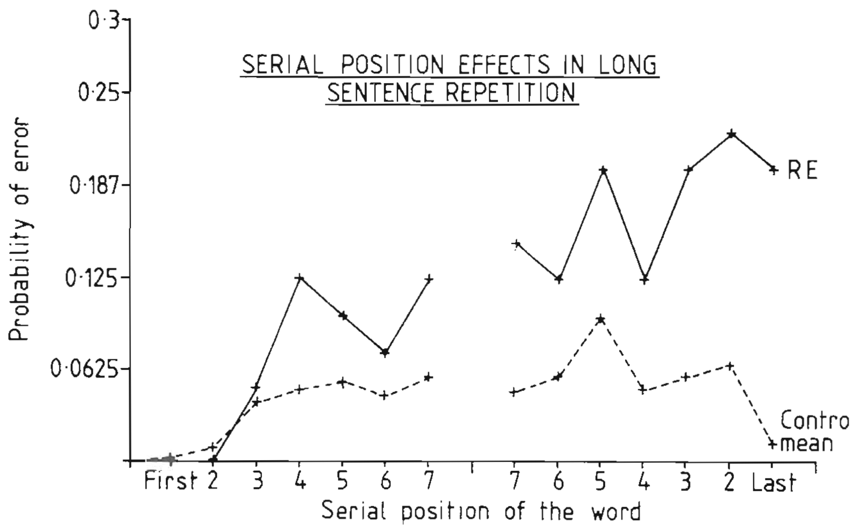

Figure 1. Serial position effects in sentence repetition; the probability of an error on the first seven and the final seven words in the sentence.

The probability of error (substitution or deletion) at each position in the sentence, for the first seven items and the final seven items, is shown in Figure 1. Error rates for the control subjects differ significantly across the first seven items [χ^2 (6) = 32.5, $p < 0.001$], owing to an advantage for the first two lexical items in the sentences. RE shows the same primacy effect. There are also significant differences across the last seven items [χ^2 (6) = 29.1, $p < 0.001$], with superior performance with the final word. RE, in contrast, shows no recency effect; she, in fact, makes more errors in the final position (five deletions and three substitutions) than all ten control subjects combined (five deletions and one substitution).

In the previous task, RE was normal in judging whether these sentences were grammatically well formed; she is, however, worse at repeating the sentences than control subjects. Her pattern of error types is normal, and, like control subjects, she shows a primacy effect; but RE differs from the controls in that she shows no recency effect in sentence repetition.

Text Comprehension and Memory

The sentence repetition task involves isolated sentences that are syntactically correct and semantically plausible. With each sentence, subjects can therefore use syntactic and semantic material to support their repetition performance, but this structural information spans only the individual sentences. To assess RE's memory for material presented as meaningful text, we used the method and materials employed by Wingfield and Butterworth (1984). In this task, subjects have to report verbatim a tape-recorded prose text of more than 100 words presented over headphones; subjects are allowed to stop the tape whenever they wish and report immediately the last chunk. In their second experiment, Wingfield and Butterworth presented one set of passages with normal intonation, one set with artificially flattened intonation and one set of passages read as a list of unrelated words. Intonational information is highly correlated with sentence structure. Sentence starts have a higher pitch than sentence ends, and terminal contours form a small and highly recognizable set of cues to sentence boundaries. These cues should enable listeners to determine at least some basic syntactic boundaries—like clause and sentence endings—without having to compute a full syntactic analysis. Insofar as a STS is implicated in syntactic processing, this should impose a lighter burden on it and leave more capacity available at any point in time for other syntactic and semantic processes. Wingfield and Butterworth expected that the absence of pitch cues in the flattened condition would mean more segmentations at nonsyntactic boundaries and poorer recall accuracy. However, overall accuracy was identical for the two conditions at around 80%, segmentations fell overwhelmingly at major syntactic boundaries (70–75%) and chunk length was identical at around 11 words. Some chunks of 18 or more words were recalled with complete accuracy. The length of chunks and their syntactic character indicates that subjects were utilizing parsing, and probably semantic interpretation, to achieve these high levels of accuracy. In the list intonation condition, segmentations did not reflect syntactic structure to nearly the same extent: only about 35% of segmentations were at clause or sentence boundaries, and only about 25% of such boundaries attracted segmentations. Mean chunk length dropped to around 8 words, although recall accuracy remained high

(90%). Although subjects performed identically with both normal and flattened intonation, it is possible that with normal intonation, comprehension reflects resource-limited performance (Norman and Bobrow, 1975), and when intonational cues are not present, more resource capacity may be allocated to the online parsing to maintain normal levels of performance. However, if a deficit in the STS reduces available resources, then RE may not be able to compensate for the lack of acoustic cues to sentence structure. Thus the second hypothesis would predict that not only would impaired comprehension result in poorer recall accuracy, or perhaps a strategy of taking shorter chunks, but also that there would be even poorer performance when parsing load is increased by the lack of intonational cues. With list intonation stimuli, which our subjects seemed to treat as successive STM tasks, with a classic 7 ± 2 items per chunk, RE should be impaired roughly to the extent that she was on serial digit recall—that is, she should recall just 4 ± 2 items.

Method

Speech Stimuli

Our stimuli consisted of 10 passages of normal English prose, on average 124 words in length. These were recorded by a male speaker with close to received pronunciation (Dr. Francis Nolan, a phonetician in the Linguistics Department at the University of Cambridge), at a rate of approximately 200 wpm. All passages were passed through a channel vocoder, and for one set the channel containing the fundamental frequency was kept at a monotone chosen to match the average fundamental frequency of the original recordings. In a third treatment the ten passages were read by the same speaker as if they were lists of unrelated words, with equal stress on each word and with no timing, amplitude or pitch variation across the words of the passage.

Procedure

The recorded passages were presented at a comfortable listening level (65 dB, re 0.002 dynes/cm^2), with a counterbalanced order of normal flat and list intonation. RE was given control of an interrupt switch, which allowed her to stop the tape whenever she wished, but she was told not to shadow the speech, but to report only when she had stopped the tape. Our instruction stressed only accuracy of recall.

This was the procedure used by Wingfield and Butterworth; and data from the 18 students in their experiment were used as controls.

Results

These are summarized in Table IV, and in Figures 2 and 3. RE's overall recall accuracy was indistinguishable from the means for controls in all

Table IV

Performance on Wingfield and Butterworth's task

		Controls[a]	RE
Mean segment length (words)	Normal Intonation	10.6	10.8
	Flat Intonation	10.6	11.8
	List Intonation	7.9	7.3
Percent words correct	Normal Intonation	80	79
	Flat Intonation	79	79
	List Intonation	90	88

[a]Data from Wingfield and Butterworth, 1984.

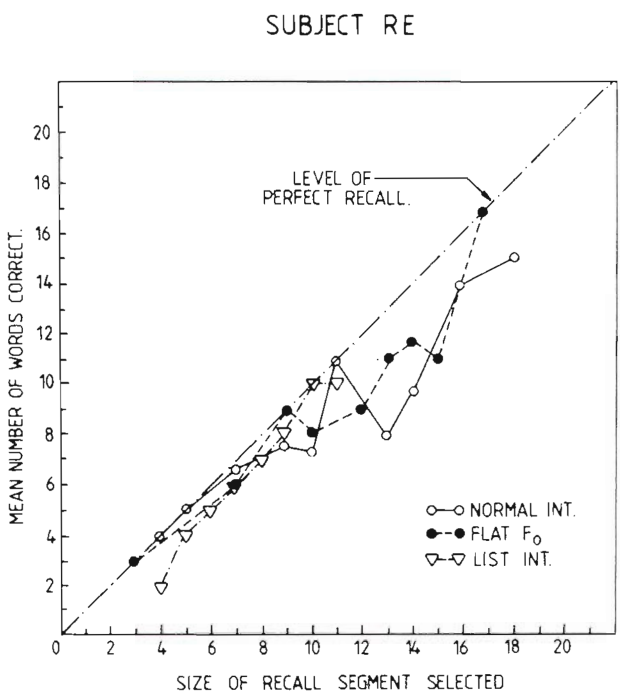

Figure 2. Mean number of words correctly reported by the subject RE as a function of the size of the segment selected for recall for Normal Flat F0 and List intonation passages. The line at 45° indicates level of perfect recall.

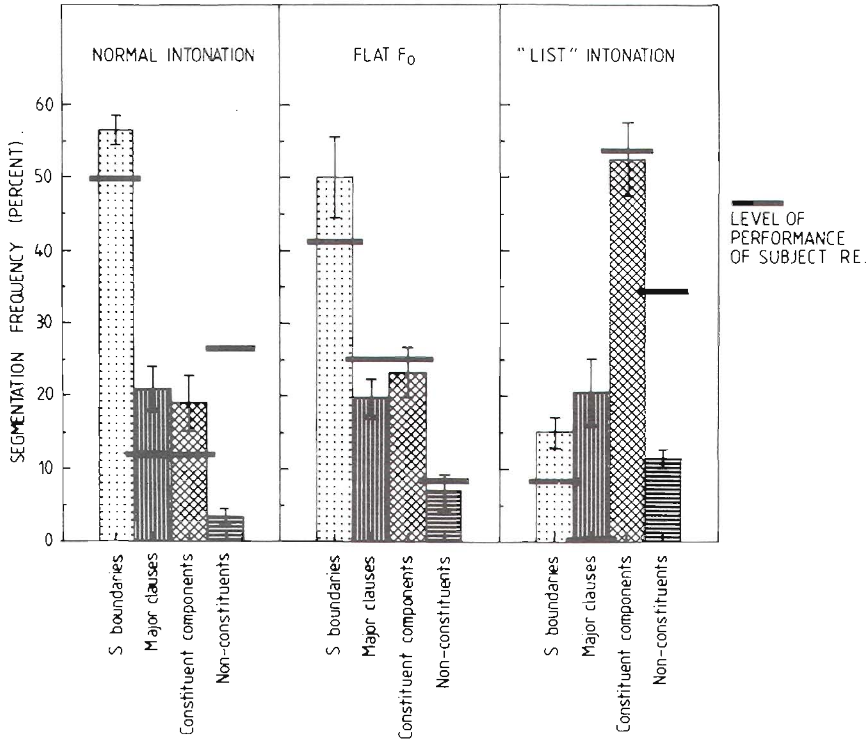

Figure 3. Relative frequency with which the subject RE segmented passages for recall (thick bars) in comparison with controls (histograms), as a function of the linguistic structure of the speech materials for Normal, Flat F0 and List intonation passages.

conditions. Segmentation was very similar, too, although with relatively small numbers of segmentations it is inappropriate to compare the individual with the pooled pattern. It is important to note that her chunks were not shorter than those of the controls, and she was willing to use long chunks—i.e. of 15 to 20 words in normal and flat conditions. As with our controls, accuracy did not fall off much with increased chunk length.

Not only was chunk length normal, the location of segmentation points with respect to syntactic boundaries was also normal, with approximately normal proportions of segmentations at sentence, clause and constituent boundaries in each condition; segmentation was affected by list intonation in much the same way as for our controls. Notice that if RE is treating list read passages as successive STM tasks, she is performing much better than for lists of unrelated words, mode chunk length being 7- or 8-word with an accuracy of around 90%.

Thus her performance on this test of sentence repetition, with sentences presented as part of a coherent text, is entirely normal.

GENERAL DISCUSSION

Two working hypotheses have been evaluated in this study:

1. Normal phonemic processing is necessary for the proper functioning of the phonological short-term store, and hence a deficit in phonological processing is sufficient for an impairment in short-term memory tasks employing this store.
2. The phonological short-term store implicated in STM tasks is necessary for the proper functioning of syntactic analysis in comprehension, and hence a deficit in STS is sufficient for an impairment in this functioning. In the models described below, syntactic analysis and comprehension depend on maintaining a verbatim record of the input sentence in STS pending the completion of syntactic and interpretative processes. Thus, it follows as a corollary to this hypothesis that intact memory for a sentence is a necessary condition for its proper interpretation, at least where analytic and interpretative processes cannot be completed online phrase by phrase as the sentence is heard. These models would, therefore, predict that, if a subject has intact comprehension for long and complex sentences, he or she would also have intact repetition for these sentences.

The tests presented here and in a previous paper (Campbell and Butterworth, 1985) show that our subject RE has an impairment of central phonological processing: she is poor at tasks that require her to segment words and to make rhyme judgements, in all presentation modes; but her auditory speech discrimination (Campbell and Butterworth, 1985), her ability to detect mispronunciations in heard words and to repeat polymorphemic words are all unaffected, demonstrating that her peripheral processing of speech is unimpaired up to and including lexical access. Concomitantly, she has a significant deficit in STM tasks normally using phonological coding and storage, and, indeed, our previous study demonstrated that she will use visual or orthographic coding and storage to mediate these tasks. This indicates that she has developed a compensatory strategy in the face of her functional disability, although this strategy still yields subnormal, as well as qualitatively different, performance on STM tasks. It is still unclear whether the deficit in central phonological processing prevents the establishment of representations appropriate for holding in STS, or whether a deficit in STS prevents RE performing the full range of phonological processing tasks.

Nevertheless, the first hypothesis is clearly supported by our results. This is broadly in line with the consensus on the involvement of phonological coding in STM tasks; however, more specific proposals, like Allport's (1984), would fail to predict a poor STM performance in the face of intact processing up to lexical access and intact word repetition.

To evaluate the second hypothesis, we used four tests involving the syntactic analysis or comprehension of auditory and written material.

Comprehension

In Parts 4 and 5 of the Token Test (De Renzi and Vignolo, 1962), which requires comprehension of complex sentences, RE performed at least as well as 10 matched controls. Moreover, unlike the controls, she was unaffected, on written presentation of Part 4 of the test, by concurrent articulation. Similarly, on the TROG (Bishop, 1982) task, she achieved almost perfect scores on both written and spoken versions of this test of complex sentence comprehension.

On a specially devised sentence–picture verification task, RE scored within the range of matched undergraduate controls. The test sentences were both long and complex, involving active and passive verb forms, and subject and object relative clauses. Accuracy thus depended on full syntactic analysis, which, in turn, required knowledge of exact word order and the establishment of coreferentiality between phrases and "gaps" separated by intervening material.

A more direct test of her ability to carry out syntactic analysis came in the syntactic judgement task. In this, RE was asked to say whether word strings, presented auditorily or visually, were wrong in some way. Correct performance on these strings required the maintenance of both morphological and order information. The d' measures showed that she was as good as matched controls at detecting ungrammatical strings.

These data are clearly consistent with the second hypothesis: RE's implied deficit in STS functioning does not lead to an impairment on tasks requiring syntactic analysis or sentence comprehension, even when the word strings are long and syntactically complex. When the evidence offered in the literature to support the hypothesis is re-examined, it turns out not to be very persuasive.

Few studies attempt to show that the comprehension deficit in patients with poor STM scores is of the specific kind that should result from an impairment to the functioning of the short-term store, and not of any other kind. The STM patients KF (Warrington, Logue and Pratt, 1971) and JB (Shallice and Butterworth, 1977) make some errors on the Token Test. The Token Test is, of course, a high-level comprehension test, and almost all patients with cortical damage will make *some* errors, for a variety of possible reasons.

Caramazza et al. (1983) show that JS has difficulty in matching pictures with written sentences involving spatial prepositions. In judging whether written sentences are correct, JS was able to reject all sentences that were *semantically* deviant (e.g. "The barber captured the razor") but accepted 9/10 *syntactically* deviant sentences (e.g. "The girl will dressing the doll"). Caramazza et al. produce *no* evidence to show that JS does *not* have a general syntactic problem. Indeed, he makes many syntactic errors in speech production, spontaneous writing and written sentence comprehension. They, however, believe that his syntactic difficulty can be attributed to a problem with phonological memory. They assert that

> to compute the syntactic relations among the major lexical items, he has to convert the graphemic representations into a phonological code, store this information in working memory, and analyse it syntactically. (1981:168)

The second patient, MC, reported by Caramazza, Basili, Koller and Berndt (1981) and Caramazza, Berndt, Basili and Koller (1981) was, like agrammatic Broca's aphasics, very poor at matching reversible sentences to pictures (e.g. "The cat is chased by the dog"), but much better with sentences where there was only one possible interpretation of the relations among major lexical items ("The bone was chewed by the dog"). Performance with written sentences was as impaired as with spoken sentences. As with JS, there is no evidence to exclude the possibility of a general syntactic deficit; and there is no systematic test of factors likely to be particularly affected by memorial, but not by syntactic, problems, like sentence length. In fact, it is clear that both patients can have difficulty with the interpretation of short and simple sentences, like the reversible one above.

Vallar and Baddeley (1984a, 1984b) argue that PV, their patient with impaired STM, has a specific problem in judging the truth of sentences only when they are long, and their correct interpretation depends on knowledge of constituent order. Thus PV performed with reasonable accuracy judging the truth of short sentences (e.g. "Bishops can be bought in shops") but had considerable difficulty with long sentences like "Many people know that often books contain pictures of various kinds which are sometimes printed in various colours;" and "One could reasonably claim that sailors are often lived on by ships of various kinds." The interpretation of this result is problematic: the long sentences are not only longer, but they are syntactically and propositionally more complex. As a result, the task of verification becomes more difficult. The first long sentence above contains several propositions: [many people know P], [books often contain pictures], [the pictures in books are of various kinds] etc. It is thus unclear whether PV's difficulty should be attributed to (1) a syntactic problem in parsing complex sentences, (2) problems in verifying propositionally complex sentences, or (3), as Vallar and Baddeley assert, a problem in coding or maintaining the *order* of surface constituents (in a specifically phonological form). It is unfortunate that they do not test their hypothesis directly by, for example, comparing the effects of constituent deletion with constituent misordering. Under their hypothesis, only deviant strings of the former kind should be detectable.

In the experiments we have reported, RE, despite a digit span of 3 to 4, shows no evidence of any deficit of sentence parsing or comprehension, even when long and syntactically complex sentences are involved.

Sentence Memory

As we have argued and will explore in more detail below, the standard models of the role of STS claim that comprehension depends on holding a verbatim record of the input sentence. Given the results above, these models yield contradictory predictions about RE's immediate sentence repetition. On the one hand, her sentence repetition should be poor because repetition of word and digit lists is poor; on the other hand, it should be intact because her sentence comprehension is intact.

We carried out three tests of immediate sentence repetition. In the first, we re-presented the Token Test sentences, this time for immediate repetition. Although comprehension had been perfect in auditory presentation, she made errors on seven sentences in repetition. Thus she can understand sentences she cannot accurately repeat. Her errors preserve the syntactic form of the target.

Our second test used the correct versions of the long sentences from the grammaticality judgement task. Overall, her performance is outside the range of matched controls, but her errors show normal evidence of comprehension in the sense that her substitutions and omissions are consistent with the meaning of the target. Unlike controls, she recalls the last word in the sentence no better than earlier words; but like controls, there is a marked primacy effect for the first word.

We can assume, though we have not tested this, that the controls show normal serial position primacy and recency effects for list materials; RE, on the other hand, shows only a primacy effect in these tasks (Campbell and Butterworth, 1985). Thus, despite the differences between RE and controls on list recall, both repeat sentences in a rather list-like way and are presumably affected by the same factors that shape their respective list performances.

To recapitulate the most surprising finding: RE is unable to repeat at normal levels sentences whose grammaticality she is able to judge at normal levels.

On a test of text memory devised by Wingfield and Butterworth (1984), RE's performance was as accurate as those of our control subjects. This test requires syntactic analysis to perform normally, since controls segment the auditory text, with normal and flattened intonation, at major syntactic boundaries; and RE adopted the same strategy. Not only did she recall the segments to the same level of accuracy, but she did so even for segments of up to 15 words and hence well beyond her span for unrelated items. This held for materials in which intonational cues were absent and where it was believed that the extra syntactic processing thereby entailed should increase transient memory load. It is hard to see how normal levels of performance can be achieved without the subject comprehending the text, especially since segmentation is controlled by syntactic analysis of the input. And it is worth pointing out that normal performance can only be achieved if RE opts for the same strategies as controls and does not try to adapt to a sentence repetition difficulty by taking shorter segments.

Contrary to the corollaries of the second hypothesis, RE is quite able to parse and understand sentences she has abnormal difficulty in repeating. Details of her performance indicate that she has syntactically analysed and understood the sentences to be repeated. Her errors are largely order and meaning preserving. In general, she is best on stimuli where full semantic and pragmatic support is available from context, as in the text memory task, and worst with the least support from context, as in the Token Test sentences. These sentences are lexically repetitive and semantically arbitrary—there is no special reason why the red circle should be put between the yellow square and the green square. These data, combined with abnormally poor recall of the last items in the second sentence repetition task, suggest that sentence repetition performance is mediated largely by the representations of sentence meaning and structure and does not rely on support from a phonological representation held in STS.

Since RE *does* perform poorly on STM tasks, which is usually interpreted as evidence of a severely limited phonological short-term store, we can conclude that this store is not necessary for sentence comprehension. It may, however, have a role in sentence recall, where reconstructive processes will make use of phonological material in STS as well as more abstract representations of syntax and semantics. While it may be possible to parse and interpret many sentences "on-line" without reference back to prior phrases, this is unlikely always to be true, and additional short-term stores for syntactic and semantic material need to be postulated to hold these representations. Alternatively, RE, unlike the controls, though she can access phonological STS for the purposes of comprehension, is unable to do so for the purposes of recall, and, if STS mediates recall of the last items of a sentence, she will have difficulty retrieving them.

In the next section, we re-examine proposals that have been made as to the role of working memory stores in sentence comprehension in the light of our results with RE.

Models of the Relation of STS to Language Comprehension

In the Introduction, we reported a widespread consensus linking STM to language comprehension; however, the consensus conceals a variety of proposals as to what exact role a phonological STS plays in the comprehension process. These proposals can be divided into two broad types—one in which STS is essentially "passive", simply registering the input, and one in which the contents of STS are "controlled" by online interpretative processes.

1. *Passive STS.* In the most extreme version, proposed by Clark and Clark (1977), the listener takes in "the raw speech and retains a phonological representation of it in 'working memory'." (p. 49). Subsequent processes identify the content and function of constituents and construct a propositional representation of the whole sentence. Propositions replace the phonological representation, which is "purged" from working memory.

Baddeley's model also postulates that a phonological representation is maintained passively in an input store until interpretation is complete, but he makes it clear that representation must consist of already identified words in phonological form.

One basic problem with both models is that we know from many studies that the identification of words is not independent of and, especially, not prior to, syntactic and semantic interpretation of the sentence in which it occurs. Experiments on the time-course of word identification in shadowing (Marslen-Wilson, 1975), gating (Grosjean, 1980) and word-monitoring (Marslen-Wilson and Tyler, 1980) all demonstrate that both semantic and syntactic context have effects on the very early stages of word recognition, so that words heard in context can usually be identified some 200 msec after onset and well before offset, whereas listeners generally need to 350 msec or more to identify words heard in isolation (Marslen-Wilson, 1984, for a review of the evidence).

Moreover, these models, as can be seen, entail a dependence of sentence comprehension on sentence memory; thus our data pose a particular difficulty because sentence memory is impaired where sentence comprehension is intact.

2. *Controlled STS.* An alternative set of proposals takes into account these findings. At the most general level, listeners control the contents of STS. Wingfield and Butterworth argue from a text memory task, where control of the amount of speech material is made explicit (see above), that the data "cannot support any concept of working memory that entails only a [fixed capacity] passive storage of verbal input prior to any higher-level analyses of the speech content" (1984:362). Listeners in this task "seemed to have been forming predictive hypotheses about the structure of what they were hearing. . . . As speech proceeds, either the individual words or their analytic representations must have been saved and integrated in order to make subsequent predictions" (1984:362). In their task, subjects used syntactic and intonational information to determine appropriate grammatical juncture to stop the input and repeat what they had heard. A purely passive model would predict that input would be stopped at arbitrary intervals.

As Marslen-Wilson has argued extensively (e.g., Marslen-Wilson, 1984), listeners do not wait until the end of a clause to begin the process of semantic interpretation of speech, but begin with the first word. And, in an important series of studies, Green (1973, 1975, 1977) demonstrates that the depth to which a sentence is interpreted is under strategic control, and this, in turn, determines the current contents of the working memory system. Where relatively shallow interpretation is induced, by instructions for verbatim recall, working memory contains in readily accessible form information about the exact lexical items; whereas when deeper processing is induced, integrated semantic representations of the sentence content seem to have displaced the lexical information. Moreover, in an unpublished experiment (Green, 1973) he shows that the listener's syntactic expectation will influence the contents of working memory.

How the listener achieves an online semantic interpretation is not stated, but a proposal by Steedman and his collaborators (Ades and Steedman, 1982; Steedman, 1983) may provide a direction for further research. They propose a categorial syntax whose implementation in a model of parsing would enable each phrase, as soon as it is identified, to be associated with a semantic interpretation, and interpretation can proceed in an incremental fashion. This model would incorporate a single nested push-down stack on which sentence constituents would be held while being assigned a syntactic role. The psychological details of such a model have as yet barely been sketched (Crain and Steedman, 1985).

A fuller treatment of sentence parsing, making different syntactic and semantic assumptions, has been offered by Wanner and Maratsos (1978). In their model, lexical, syntactic and semantic analysis of input share a "common working memory." Each word in a sentence is processed one at a time, from left to right, and is assigned a syntactic role in the local phrase and in the sentence as a whole, and the word plus its assignment, and any interim analyses, are stored in working memory. Where the complete sentence role of a word or phrase cannot be determined at the point at which it occurs, it is assigned to a "HOLD" list, and its full assignment computed on the basis of the analysis of later portions of the sentence. In this way, both words and their structural function are held together in working memory. In their model, and in Steedman's, there is no simple parameter of word or constituent order in memory (as there is in, say, Baddeley's model), but, rather, labelled syntactic structures. Unfortunately they do not give a word-by-word account of the contents of working memory, but only the general idea that when many items are on HOLD this imposes an extra burden on working memory capacity.

Their test of this proposal uses serial visual presentation of the sentence to be parsed plus a list of five unrelated words to be recalled. They found that if the list is presented while material is in the HOLD list, both comprehension and serial recall is reduced, as compared to presentation, when no material is meant to be on HOLD. This would seem to create a direct link between processes of parsing and of list recall. However, since both words and their abstract syntactic labels are held in working memory, this store cannot be exclusively phonological; and given that no specific test was made for how the visually-presented words are coded, it is unclear whether it is phonological at all. One way of testing this would be to make the list items phonologically similar to the words in the sentence, to see whether this produces the well-known decrement in recall. Even so, an important caveat should be entered concerning the interpretation of dual task studies. It is widely known, though rarely reported (but see Shallice, McLeod and Lewis, 1985), that when a subject gets into trouble on one task, performance on the concurrent task is detrimentally affected, irrespective of whether the tasks share common resources. Thus, data that show an *average* decrement in both tasks for a certain kind of material in one of those tasks—e.g. sentence fragments with items in HOLD—may indicate only that this kind of material is more likely to lead to trouble, and hence be more likely to cause performance decrement on the other task. To demonstrate that the critical material leads to an increased dependence on a resource common to both tasks, it is necessary to show that there is a performance decrement on the second task—here, list recall—on trials with *perfect* performance on the supposedly more demanding (sentence) materials. Since Wanner and Maratsos test *either* list recall *or* sentence comprehension, but not both, on each trial, this inference cannot be made.

Nevertheless, the idea that the contents of working memory should be at least partly under the control of parsing and comprehension processes and should include intermediate parses, and presumably other materials like partial semantic interpretation in addition to the phonological representation of words, is an interesting one, but there seems no compelling theoretical reason why these different kinds of representation should compete for a single limited memorial resource.

Since our results with RE dissociate phonologically mediated short-term list recall from sentence parsing and comprehension, the hypothesis of a single, passive, phonological short-term store underlying these two tasks no longer looks plausible. An alternative approach would be to think in terms of different processes—comprehension and recall—having differential access to a phonological representation held in a STS. This would require a radical recasting of current working memory models, and theorists would at last have to confront the problem of specifying in some detail the kinds of processes that utilize memorial representations in the generation of a response. But perhaps the most fruitful direction in which to proceed is to consider the possibility of multiple independent working memories, each associated with and under the control of some well-defined and ecologically sensible process.

REFERENCES

Ades, A. E. and Steedman, M. J. (1982). On the order of words. *Linguistics and Philosophy*, **4**, 517–558.

Allport, D. A. (1984). Auditory-verbal short-term memory and aphasia. In H. Bouma and D. G. Bouwhuis (Eds.) *Attention and performance X: Control of language processes.* London: Lawrence Erlbaum Associates.

Atkinson, R. C. and Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation.* New York: Academic Press.

Baddeley, A. D. (1979). Working memory and reading. In P. A. Kolers, M. Wrolstad and H. Bouma (Eds.), *Processing visible language, Vol. 1.* New York: Plenum.

Baddeley, A. D. and Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation. Advances in theory and research. Vol. 8*. New York: Academic Press.

Bishop, D. V. M. (1982). *TROG Test for Reception of Grammar*. Available from the author at Psychology Department, University of Manchester.

Campbell, R. and Butterworth, B. (1985). Phonological dyslexia and dysgraphia in a highly literate subject: A developmental case with associated deficits of phonemic processing and awareness. *Quarterly Journal of Experimental Psychology*, **37A,** 435–475.

Caramazza, A., Basili, A. G., Koller, J. J. and Berndt, R. S. (1981). An investigation of repetition and language processing in a case of conduction aphasia. *Brain and Language*, **14,** 235–275.

Caramazza, A., Berndt, R. S. and Basili, A. G. (1983). The selective impairment of phonological processing: A case study. *Brain and Language*, **18,** 128–174.

Caramazza, A., Berndt, R. S., Basili, A. G. and Koller, J. J. (1981). Syntactic processing deficits in aphasia. *Cortex*, **17,** 333–348.

Clark, H. H. and Clark, E. V. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology*, **55,** 75–84.

Conrad, R. (1972) The developmental role of vocalising in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, **11,** 521–533.

Crain, S. and Steedman, M. J. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Kartunnen and A. Zwicky (Eds.), *Natural language processing*. Cambridge: Cambridge University Press.

De Renzi, E. and Vignolo, L. A. (1962). The Token test: A sensitive test to detect receptive disturbances in aphasics. *Brain*, **85,** 665–678.

Garrett, M. F., Bever, T. and Fodor, J. A. (1966). The active use of grammar in speech perception. *Perception and Psychophysics*, **1,** 30–32.

Green, D. W. (1973). A psychological investigation into the memory and comprehension of sentences. Unpublished Ph. D. Thesis, University of London.

Green, D. W. (1975). The effects of task on the representation of sentences. *Journal of Verbal Learning and Verbal Behavior*, **14,** 275–285.

Green, D. W. (1977). The immediate processing of sentences. *Quarterly Journal of Experimental Psychology*, **29,** 1–12.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, **28,** 267–283.

Johnson-Laird, P. N. and Stevenson, R. (1970). Memory for syntax. *Nature*, **227,** 412.

Lashley, K. (1951). The problem of serial order in behaviour. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley.

Linebarger, M. C., Schwartz, M. F. and Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, **13,** 361–392.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, **189**, 226–227.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition: A tutorial review. In H. Bouma and D. G. Bouwhuis (Eds.) *Attention and performance X: Control of language processes.* London: Lawrence Erlbaum Associates.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, **189**, 226–227.

Morton, J. (1970). A functional model of memory. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press.

Norman, D. A. and Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, **7**, 44–64.

Saffran, E. and Marin, O. S. M. (1975). Immediate memory for word lists and sentences in a patient with deficient auditory short-term memory. *Brain and Language*, **2**, 420–433.

Salamé, P. and Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, **21**, 150–164.

Savin, H. B. and Perchonock, E. (1965). Grammatical structure and immediate recall of sentences. *Journal of Verbal Learning and Verbal Behavior*, **9**, 348–353.

Shallice, T. (1979). Neuropsychological research and the fractionation of memory systems. In L. J. Nillson (Ed.), *Perspectives on memory research.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Shallice, T. and Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia*, **15**, 729–735.

Shallice, T., McLeod, P. and Lewis, K. (1985). Isolating cognitive modules in the dual task paradigm: Are speech perception and production separate processes? *Quarterly Journal of Experimental Psychology*, **37A**, 507–532.

Steedman, M. J. (1983). On the generality of the nested-dependency constraint and the reason for an exception in Dutch. In B. Butterworth, B. Comrie and O. Dahl (Eds.), *Explanations for language universals.* Berlin: Mouton.

Vallar, G. and Baddeley, A. D. (1984a). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology*, **1**, 121–141.

Vallar, G. and Baddeley, A. D. (1984b). Fractionation of working memory. Neuropsychological evidence for a phonological short-term store. *Journal of Verbal Learning and Verbal Behavior*, **23**, 151–162.

Wanner, E. and Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. Bresnan and G. A. Miller (Eds.), *Linguistic theory and psychological reality.* Cambridge, MA: MIT Press.

Warrington, E. K., Logue, V. and Pratt, R. C. T. (1971). The anatomical localisation of selective impairment of auditory-verbal short-term memory. *Neuropsychologia*, **9**, 377–387.

Waugh, N. C. and Norman, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89–104.

Wingfield, A. and Butterworth, B. (1984). Running memory for sentences and parts of sentences: Syntactic parsing as a control function in working memory. In H. Bouma and D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes*. London: Lawrence Erlbaum Associates.

Revised manuscript received 13 May 1986