APHASIOLOGY, 2001, *15* (2), 151–169

Improvement of oral naming by unsupervised computerised rehabilitation

Palle Møller Pedersen and Kirsten Vinter

Bispebjerg Hospital, Copenhagen, Denmark

Tom Skyhøj Olsen

Gentofte Hospital, Gentofte, Denmark

The effects of unsupervised, computerised rehabilitation of anomia in aphasia were investigated with three single-case studies. The training was carried out in the patients' own homes without any supervision. It included semantic and phonemic tasks as well as written naming tasks. The computer was programmed to progress through a preselected set of tasks according to set criteria of performance. The patient could not alter the preselected sequence, but could start and stop a session at any time. The length of the training was determined by the time it took the patient to work through the full set of tasks. Comparison was made of their progress in naming of trained and untrained words. A cross-over design was used, and baselines were established with two patients. All patients improved in their general naming performance, but to different degrees. A relatively specific effect for the trained words was found in two patients, whereas the third showed a somewhat more general effect. The chosen therapy was probably not equally appropriate for all three patients. It is concluded that oral naming can be improved with unsupervised computer rehabilitation in the home even though the patient is not required to speak during the training.

Anomia—word finding difficulties—is a core symptom of aphasia, and it is common to the different types of aphasia. The mildest type of aphasia, according to the classification of the Boston group (Goodglass & Kaplan, 1983), may have anomia as the major symptom, and is accordingly called anomic aphasia. The symptom of anomia is only weakly localised within the left hemisphere, but different localisations of brain lesions are assumed to be associated with different types of anomia (Benson & Barber, 1996). It has been proposed that there are different underlying mechanisms, and that each mechanism demands its own type of therapy, a so-called ''model-appropriate therapy'' (Nettleton & Lesser, 1991). The last point remains controversial, but it is obvious that single-case and small group studies have established the symptom of anomia as a promising target for rehabilitation (Nickels & Best, 1996a).

Cognitive neuropsychological models all assume that oral naming is at least a multi-step process that originates in a semantic system and passes through a word-form phonological output lexicon before phonetic assembly of the sounds of the word. Some

Address correspondence to: Palle Møller Pedersen, Department of Neurology, Bispebjerg Hospital, DK-2400, Copenhagen, NV, Denmark.

The study was supported by a generous grant from the Danish Health Foundation. We wish to thank the patients, their relatives, and the Departments of Neurology at Bispebjerg Hospital and Hvidovre Hospital, Copenhagen, and the Danish Stroke and Aphasia Association (Hjernesagen) for their participation in the project.

http://www.tandf.co.uk/journals/pp/02687038.html DOI:10.1080/02687040042000106

authors believe that there is both a semantic system, probably of a non-linguistic nature, and a truly linguistic semantic lexicon. Damage can occur to entities inside these systems, or there could be an access problem. It is often assumed that damage to internal entities will result in relatively stable problems with certain words, whereas access problems will be manifested in errors on different words on repeated testing. It is also often assumed that category specificity in the naming errors implies damage to the semantic system or lexicon, whereas effects of word frequency are ascribed to the phonological output lexicon. Phonological paraphasias may imply problems with the phonological output lexicon or later stages, whereas semantic paraphasias have less clear implications. Problems with the phonological output lexicon are also clearly implied when there is anomia in the oral but not in the written modality (Nickels, 1997). Cases with clear-cut dissociations among symptoms are important in order to be able to show that the semantic system, the semantic lexicon, and the phonological output lexicon are indeed psychologically real, relatively independent processing modules. In ordinary rehabilitation, however, most patients must be expected to have lesions that include the neural substrates for several modules, and thus a less clear-cut symptom picture.

A number of researchers have tried to establish the most effective rehabilitation for anomia. Howard et al. (1985a) distinguished among cueing, facilitation, and therapy. *Cueing* helps the patient to produce the word immediately after the cue. A subsequent effect is called *facilitation*. Bringing about a lasting effect is called *therapy*. Howard et al. first established that a semantic task—in which the patient was not required to say the target word at all—was more effective in *facilitation* of subsequent naming (the following day) than a phonological task. A subsequent study of semantic and phonological based *rehabilitation* did not find the same clear difference in the effects of the two strategies (Howard et al., 1985b). Later research in the rehabilitation of anomia has been rooted in the cognitive neuropsychological model of single word processing, and it has been assumed that the most effective type of training is dependent on the locus of the underlying disturbance in the production of single words (Nickels & Best, 1996a). It seems, however, that irrespective of the locus of the underlying disturbance, the most effective strategy in most cases is a combination of semantic and phonological approaches (Nickels & Best, 1996b).

It is generally believed that aphasia therapy must be intensive and of long duration to be effective (Basso, 1989), but the therapist resources are usually not sufficient. Thus, it would be very helpful if computers could supplement the traditional speech therapy. If anomia can be rehabilitated without the patient saying the target word, then it becomes a possible candidate for computer rehabilitation, in spite of current computer limitations—that they cannot evaluate the patient's spoken output in a way that would be adequate for supplying the patient with feedback on oral naming. It has already been shown that written naming exercises presented by a computer may have results that generalise to oral naming (Deloche et al., 1992).

Most non-computerised methods include a semantic component (Nickels & Best, 1996b). It is therefore important to try a combination of semantic, phonological, and written naming methods with a computer. Unsupervised training of anomia has, in fact, been used successfully with non-computerised training methods (Nickels & Best, 1996b), but non-computerised, unsupervised training has the limitation that an immediate and reliable feedback for correctness of the response cannot always be given.

Research by Loverso et al. (although in the area of agrammatism, not anomia) has suggested that training delivered by a human therapist is more effective than training delivered by a computer: the goal was reached in a shorter time with the human therapist than with the computer using the same training protocol (Loverso, Prescott, & Selinger, 1992). However, if the program allows the patient to work unsupervised with the computer, it may still be worthwhile because more rehabilitation can be offered at lower cost.

A real gain in the use of therapist resources is obtained only if the patient is able to work unsupervised for long periods, that is, working alone with the computer without the presence of a therapist. Compared to traditional paper and pencil home-work, computer therapy makes it possible to control variables such as cueing, feedback, progression in tasks, numbers of trials, and time. It is also easier to supply an immediate feedback with the computer than with other materials for home-work. Immediate feedback might increase motivation, whereas a delayed correction of errors could strengthen the wrong rather than the correct response.

We assume that computer rehabilitation, like traditional speech therapy, has to be individualised in order to be effective. It is inevitable that the therapist must spend some time individualising the program, but the mechanism for this should be so user-friendly that the time devoted to it is very small compared to the amount of time the patient may subsequently work unsupervised with the program. The present study is based on such a program developed by the first author.

METHOD

Experimental design

The study was designed as three single-case designs with patients with anomia, using the same set of computer programs. Experimental control for nuisance factors like general stimulation and increased optimism was achieved by combining a within-task cross-over design and an across-functional-areas multiple-baseline design. Words were selected for the treatment of anomia from among the words the patients were unable to name on the first test. These words were split into two (or three) sets matched for word frequency and animacy (picture depicting man-made or natural objects). One set was trained first, with the other set serving as control. After completed training of the first set of words and subsequent testing, the other set was trained, thus constituting a cross-over within functional area design. A second baseline was established by testing for auditory comprehension of locative prepositions. After completion of training of anomia for all sets of words, locative prepositions were trained in the follow-up period, constituting the across-functional-areas multiple-baseline design.

Patients

Three stroke patients with aphasia were selected from the neurological departments of Bispebjerg Hospital and Hvidovre Hospital, Copenhagen, Denmark and through the Danish Stroke and Aphasia Association. They all have Danish as their mother tongue and live in Copenhagen, Denmark. Details are given later.

Assessments

Type and severity of aphasia was determined using the first part of the Western Aphasia Battery (WAB) (Kertesz, 1982) which was translated and adapted for Danish by the study group (psychometric properties of the translation will be reported elsewhere). The first part of the WAB has four sections: Spontaneous Speech, Auditory Verbal Comprehension, Repetition, and Naming. The weighted sum of sub-scores yields an aphasia quotient

(AQ) ranging from 0–100, with scores above 93.8 considered normal (non-aphasic). Type of aphasia was determined actuarially using ranges of sub-scores, and each patient was assigned to a single type of aphasia, in principle leaving no patients unclassified. Further assessments were carried out with selected tests from the Psycholinguistic Assessment of Language Processing Abilities (PALPA) (Kay, Lesser, & Coltheart 1992) which was also translated and adapted by the study group. This test is based on psycholinguistics and cognitive neuropsychology. Its primary goal is to explain the patient's performance in terms of cognitive neuropsychological models of the modules assumed to be involved in language processing, and their interconnections. It comprises 60 subtests and does not prescribe which to include in an assessment, or even where to begin. However, when deficits on a sub-test are found, there are suggestions for further assessments.

Naming performance was assessed with the Snodgrass and Vanderwart (S&V) standard set comprising 260 drawn outline-pictures (Snodgrass & Vanderwart, 1980). Word frequency was determined as the combined occurrences of all inflections of each word, as published in a Danish dictionary based on a corpus of 4,070,419 words from selected novels, short stories, newspapers and weekly magazines (Bergenholtz, 1992). Identification of lemmas was not attempted, as no human tagging of the words was carried out in this corpus. Although we refer to the concept ''word frequency'', we will use the raw occurrence counts in all tables. The authors of the present paper judged 97 of the S&V pictures to be of animate objects and 163 to be of inanimate objects. The mean frequency for the words for the animate objects is 227 (SD 652), and 220 (SD 436) for the words for the inanimate objects, so there is no interaction of frequency and animacy for the words for the S&V set of pictures (Student's *t*-test: $t = -0.7$, $p = .5$). Errors were scored as (1) no response or partial response: the patient either said nothing at all or said a small part of the word; (2) verbal paraphasia: an error in the sound of the word, where it was clear that the corrected target word was attempted; (3) literal paraphasia: a wrong word was said in place of the correct word, and it might or might not be closely related semantically; (4) mixed and neologic paraphasia: a mixture of verbal and literal aphasia that might result in a word that is not in the language. On the basis of the S&V naming responses we computed an ''instability index'' for the baseline and follow-up periods. It was computed as the number of words where the naming response changed from correct to error or error to correct divided by the number of errors on the first of the two tests in the comparison (in order to make the index comparable for patients with differing numbers of errors). The instability index must be interpreted with caution, as it has been measured only in the three patients in this study.

Rehabilitation

Rehabilitation was carried out with a set of computer programs developed for the present study. These programs can be configured with individualised content, cues, and difficulty. A number of different programs are placed in a pre-defined order for each patient, and the system automatically determines when the patient is ready to proceed from one level to another based on records of the patient's results so far. For each item in each task-set, the patient must respond correctly to the item a specified number times in a row before the item will be removed from the set. This is in order to ensure that the item has really been learned, so that accidentally correct responses over a number of trials do not allow the patient to go on. The required number of correct responses was established at three for the patients in the present study. The computer was operated with a mouse or a trackball—the keyboard was never used. When the entering of letters was required, these were shown as graphic buttons on the screen in alphabetic order (or only the letters required for the task). All words were digitally recorded on the computer in high quality (22 kHz, 16 bit, mono). The rehabilitation system's control shell program replaced Windows 95's desktop, so the computer could not be used for other purposes during the rehabilitation periods, and there was no possibility that patients could start the rehabilitation programs out of the intended order.

The design of the present study was different from most previous single-case designs in aphasia rehabilitation: we decided to train each set of words to criteria, not for a certain, predetermined amount of time. This means that a set of words could be trained for an amount of time different from its comparison set. Both methods for determining the amount of training given can be argued for and against. One advantage of training to criteria is that it takes into account the time needed to become accustomed to the types of training tasks. Another advantage is that one set is not allowed to be ''over-learned'' in comparison to the other set (or rather: we induced the same amount of ''over-learning'' for both sets of words). We also find it logical that one type of training, or cues for a given type of training, is not given up before the tasks are completely mastered, but at the same time we want all patients to go through all stages (semantic, phonologic, and graphic) of the program for each of the word-sets.

The training progressed through the following tasks in the stated order: *Semantic tasks*—Two different programs were used in succession: (1) Choose one out of four pictures based on a word that is first both ''said'' by the computer (digitally played back through the computer's loudspeakers) and shown in writing, then only said, and finally only shown in writing. Foil pictures were chosen from the training set to be as semantically close to the target as possible. (2) Choose the correct word out of two words for one picture. The words were shown in writing and also said by the computer. Foil words were chosen from the training set to be as semantically close to the target as possible. *Phonological tasks*—(1) A target picture is shown. The patient must choose between the target word and a phonologically somewhat similar foil word (not included in either the training or the control set). (2) A word is shown with the initial letter missing. The letter must be chosen between two letters. First, both the written letter and its sound are supplied. Then only the sound, and finally only the written letter. *Written naming*—A picture is shown. The corresponding written word has to be (1) copied, (2) assembled from its letters (shown on buttons on the screen, like anagram task), (3) written without help (all letters of the alphabet shown on buttons on the screen). Control training was done during the follow-up periods with tasks to train auditory comprehension of locative relations.

Words for the training sets were selected from among the words on which the patients had naming errors (no response or paraphasias) on confrontation naming of the S&V pictures. Foil words and pictures were selected from among the other target words and pictures of the training set, except for the first type of phonological tasks. Training sets were matched for frequency and animacy.

A program for training of ''locative relations'' was used as a control task during follow up, because multiple-baselines across different functional areas was intended. It was modelled on the PALPA test 58 (Kay et al., 1992). A sentence like ''shoes between hats'' would be either displayed as text, said, or both, by the computer, and the patient had to choose among four pictures. One patient (RI) was totally unable to solve the task and a modification was introduced. In this version there were only two pictures to choose from, and all locative relations were illustrated by diagrams that were labelled by the word, which could also be said by the computer if it was clicked. For example ''on'' was illustrated by a red cross on a box and ''between'' was illustrated by a red cross between two boxes.

Statistics

The matching of word frequency across training sets of words was evaluated with a one-way anova. Changes in errors (pass–fail) on trained words on repeated testing before and after therapy were analysed with the paired MacNemar test (Willmes, 1995). Differences in trained and untrained words across training sets were first analysed with Corchrans Q, for all testing sessions for all trained words, to ascertain overall significant differences (to avoid type I errors because of multiple tests) and subsequently (post hoc) with the chi-square test to evaluate the effects of the single training periods. The difference in the proportion of errors in naming of animate and inanimate object was analysed with the chi-square test, and the difference in mean frequency for words with and without naming errors was analysed with the unpaired Student's *t*-test. The required two-tailed significance level was established at $p < 0.05$ for all statistical tests. The statistical analyses were carried out with the SPSS for Windows v. 10 (1999).

STUDY 1

KB was a 57-year-old academic who suffered a stroke on 14 September 1995 with right-sided hemi-paralysis and global aphasia. An acute CT-scan showed a left-sided media infarct. A CT-scan one month later showed a 3×7 cm large fronto-parietal left-sided infarct. Dobbler ultrasound showed a total occlusion of the left internal carotid. A speech therapist found relatively better ability to communicate by reading and writing compared to auditory comprehension and speaking. WAB, taken before baseline on 12 March 1997, showed a transcortical sensory aphasia with an AQ of 68.5 (Table 1). KB was, however, very close to being classified as having anomic aphasia, with a comprehension score only slightly below the score required for anomic aphasia. He did, in fact, fluctuate from transcortical sensory aphasia to anomic aphasia and back on the three tests taken during rehabilitation (see Table 1 for details). KB was also the patient with by far the lowest fluency score, bordering on non-fluent aphasia on all three tests (transcortical motor aphasia).

KB had a high percentage of verbal paraphasias on confrontation naming with the S&V pictures, but was also the patient with the highest percentage of literal and mixed paraphasias (Table 2). The verbal paraphasias cannot be assumed to express a semantic system dysfunction, as they might be the result of compensation attempts: it is a better communication strategy to say a word semantically close to the target than to say nothing at all (Caplan, 1992). Moreover, there was no animacy effect, and there was a clear frequency effect. Apart from his type of errors, it was noted that he had many self-corrections. Successful self-correction seemed only to happen after semantic paraphasias, not after phonological errors. On the WAB (Table 1) he had a low Fluency score in ''Spontaneous Speech'', but he was quite good in ''Word fluency'' in the ''Naming'' part of the WAB (compared to the two other patients). This pattern suggests that KB's anomic difficulties were associated with the phonological output lexicon.

From baseline to pre-therapy, the naming responses changed from correct to error on 41 words and from error to correct on 29 words (instability index: 0.65). From post-therapy to follow-up, the changes were 29 and 20 (instability index 0.84). This is higher than what was found for JI, but lower than what was found for RI (see later), suggesting that there was to some degree an access problem.

TABLE 1
Western Aphasia Battery scores for KB

	Before baseline 12 Mar 97	*Before rehabilitation* 2 Jun 97	*After follow-up* 5 Feb 98
Type of Aphasia	Transcortical sensory	Anomic	Transcortical sensory
Aphasia Quotient (AQ. max. 100)	68.5	70.1	70.6
I. Spontaneous Speech (max. 20)	13	14	13
a. Information Content (max. 10)	8	9	8
b. Fluency (max. 10)	5	5	5
II. Auditory Verbal Comprehension (max. 10)	6.75	7.35	6.9
a. Yes/No Questions (max. 60)	54	60	54
b. Auditory Word Recognition (max. 60)	50	52	55
c. Sequential Commands (max. 80)	31	35	29
III. Repetition (max. 10)	8.2	7.7	8.3
IV. Naming (max. 10)	6.3	7	7.1
a. Object Naming (max. 60)	40	50	47
b. Word Fluency (max. 20)	7	5	6
c. Sentence Completion (max. 10)	8	10	9
d. Responsive Speech (max. 10)	8	5	9

TABLE 2
Oral naming responses for 260 S&V pictures, error-types, animacy and frequency effects for KB

	Before baseline	*Before training of word set 1*	*Before training of word set 2*	*After rehabilitation*	*Follow-up*
	23 April 97	9 June 97	22 Jul 97	5 Nov 97	29 Jan 98
No. correct	152	140	171	211	202
Total no. errors	108	120	89	49	58
No response, partial response	33 (31%)	50 (42%)	22 (25%)	11 (22%)	13 (22%)
Verbal paraphasia	52 (48%)	55 (46%)	53 (60%)	34 (69%)	34 (59%)
Literal paraphasia	11 (10%)	11 (9%)	13 (15%)	2 (4%)	6 (10%)
Mixed and neologic paraphasia	12 (11%)	4 (3%)	1 (1%)	2 (4%)	5 (9%)
Errors, animate pictures	32 (33%)	44 (45%)	32 (33%)	15 (15%)	23 (24%)
Errors, inanimate pictures	76 (47%)	76 (47%)	57 (35%)	34 (21%)	35 (22%)
Statistical difference in animacy	$\chi^2 = 5.8, p = .016$	$\chi^2 = 0.04$, *NS*	$\chi^2 = 0.1$, *NS*	$\chi^2 = 0.6$, *NS*	$\chi^2 = 0.2$, *NS*
Mean frequency, correct words (SD)	307 (625)	353 (664)	267 (555)	238 (478)	273 (581)
Mean frequency, words with errors (SD)	140 (319)	102 (230)	181 (464)	237 (704)	114 (221)
Statistical difference in mean frequency	$t = 2.8, p < .005$	$t = 4.2, p < .001$	$t = 1.3$, *NS*	$t < 0.01$, *NS*	$t = 3.2, p = .002$
Errors in set 1 (51 words)	51	40	**11**	17	19
Errors in set 2 (50 words)	50	35	35	**5**	12
Errors in locative relations (24 items)	13			10	6

Rehabilitation

The study of KB was performed with both cross-over and multiple baselines. A baseline (Test 1) was established for naming (S&V) and for Auditory Comprehension of Locative Relations (PALPA 58) on 23 April 1997. Two matched sets of 51 and 50 pictures with naming errors on this test were composed. Test 2 was performed on 9 June 1997. Training was initiated with the first sets of 51 words and was completed after 39 hours of training. Test 3 was done on 22 July 1997. The second set of 50 words was trained until 5 November 1997 (the amount of training was not registered due to a computer time error). After Test 4, training of locative prepositions was carried out, and the final Test 5 was performed on 29 January 1998. The two sets of S&V pictures were matched for animacy and frequency of the corresponding words so that there were 15 pictures of animate objects in each set. There, the difference in mean word frequencies between the two sets of words (122.6 SD 321.1 and 136.1 SD 315.5) was not significant (unpaired Student's *t*-test, $t = -0.21$, $p = .85$).

Results

Table 2 describes KB's performance on naming of the 260 S&V pictures before baseline, before training of the first set of pictures, after this training, after training of the second set of pictures (after cross-over) and after follow-up. Looking at all 260 words, KB improved his naming from 108 errors to 49 errors from before to after training of both sets ($\chi^2 = 49.5$, $p < .001$), and there was only a small and insignificant regression to 58 errors at follow-up, where the improvement from pre-training was still significant ($\chi^2 = 38.7$, $p < .001$). Looking only at the 101 words selected for training, there were 75 (74%) errors before training. After training of both sets of words, there were 21 (21%) errors, and after follow-up there were 31 (31%) errors. The reduction in errors from pre- to post-therapy testing was significant ($\chi^2 = 44{,}3$, $p < .001$), the increase from post-therapy to follow-up was insignificant, and the decrease from pre-therapy to follow-up remained significant ($\chi^2 = 33{,}0$, $p < .001$). Looking at the specificity of training, a general difference in improvement between the sets can be seen, in that the overall error-distribution among the 101 trained words differs significantly over the five testing points (Cochran's Q = 166.5, $p < .001$), allowing for multiple tests for significant differences within the distribution. With training of the first set of words, 31 improved, 18 were unchanged, and 2 deteriorated. In the untrained second set, 6 improved, 38 were unchanged, and 6 deteriorated ($\chi^2 = 26.0$, $p < .001$). With training of the second set, 31 improved, 18 were unchanged, and 1 deteriorated. In the now untrained first set, 7 improved, 31 were unchanged, and 13 deteriorated ($\chi^2 = 28.9$, $p < .001$). The WAB AQ (Table 1) did not improve from before (70.1) to after rehabilitation (70.6).

KB had a significantly higher error frequency for inanimate pictures before baseline, but not at any other testing point, and the result is probably spurious. The mean frequency was higher for correct words before baseline, before training, and at follow-up, but not before training of word set 2 and after rehabilitation (Table 2).

DISCUSSION

The therapy produced a significant overall effect that was only insignificantly reduced after follow-up. It cannot be ruled out that the reduction would continue, but it was so small that a continued and even less frequent use of the computer at convenient times in the patient's own home should be sufficient to maintain the gains. The training effect was relatively specific for the trained words as the improvements were significantly greater for trained compared to untrained words. Thus, in this study there was both a clinically and statistically significant effect on the naming ability of a patient with chronic aphasia. The specificity of the rehabilitation effect is convenient for the single-case research design. A generalisation to untrained words would have been clinically preferable. The rehabilitation is, however, worthwhile even without such a generalisation: the number of words the patient is able to name is increased with minimal investment of therapist time. The WAB AQ did not improve in this patient, a fact that underlines the specificity of the effect.

The results of the training support the hypothesis that the source of KB's anomia is associated with the phonological output lexicon. The training seems to attenuate the effect of frequency and is relatively specific for the trained words, which would be expected with a disturbance of the phonological output lexicon. The moderate level of the instability index suggests that there could be a combination of access problems and internal representation problems. The high weight on phonological training that is assumed to be introduced by the extensive written naming part seems to have been an appropriate choice for this patient; it seems to have been ''model-appropriate therapy''.

STUDY 2

JI was 71 years old and retired from an office supervisory position when he was admitted to hospital on 15 January 1997. He had a paraplegia of his legs since childhood. In 1993 he had a stroke with some difficulty of speech and right-sided paresis and homonymous hemianopsia. He now presented with severe expressive aphasia. The next day he was assessed by the first author with the first part of the WAB. An anomic aphasia and an AQ of 76 was found (see Table 3 for details). JI was retested with WAB on 6 November 97 and now had an anomic aphasia with an AQ of 86.

Before baseline and before therapy, JI only had no-response and semantic paraphasic errors on S&V (Table 4). There was both a word frequency effect and an animacy effect.

TABLE 3
Western Aphasia Battery (WAB) scores for JI

	On acute onset 16 Jan 97	*Before rehabilitation* *6 Nov 97*
Type of Aphasia	Anomic	Anomic
Aphasia Quotient (AQ. max. 100)	76	86.4
I. Spontaneous Speech (max. 20)	16	19
a. Information Content (max. 10)	7	10
b. Fluency (max. 10)	9	9
II. Auditory Verbal Comprehension (max. 10)	8	9.9
a. Yes/No Questions (max. 60)	60	60
b. Auditory Word Recognition (max. 60)	48	58
c. Sequential Commands (max. 80)	52	80
III. Repetition (max. 10)	8.6	8.6
IV. Naming (max. 10)	5.4	5.7
a. Object Naming (max. 60)	38	39
b. Word Fluency (max. 20)	2	2
c. Sentence Completion (max. 10)	6	8
d. Responsive Speech (max. 10)	8	8

TABLE 4
Oral naming responses for 260 S&V pictures, error-types, animacy and frequency effects for JI

	Before baseline 13 Nov 97	*Before training of word set 1* 4 Dec 97	*Before training of word set 2* 25 Feb 98	*Before training of word set 3* 6 May 98	*After rehabilitation* 8 Jun 98	*Follow-up* 17 Aug 98
No. correct	76	86	86	109	113	100
Total no. errors	184	174	174	151	147	160
No response, partial response	141 (77%)	123 (71%)	141 (81%)	119 (79%)	111 (76%)	109 (68%)
Verbal paraphasia	43 (23%)	51 (29%)	33 (19%)	32 (21%)	36 (24%)	51 (32%)
Literal paraphasia	0	0	0	0	0	0
Mixed and neologic paraphasia	0	0	0	0	0	0
Errors, animate pictures	78 (80%)	71 (73%)	73 (75%)	67 (69%)	70 (72%)	73 (75%)
Errors, inanimate pictures	106 (65%)	103 (63%)	101 (62%)	84 (52%)	77 (47%)	87 (53%)
Statistical difference in animacy	$\chi^2 = 7.0$, $p = .008$	$\chi^2 = 2.8$, *NS*	$\chi^2 = 4.9$, $p = .028$	$\chi^2 = 7.7$, $p = .006$	$\chi^2 = 14.7$, $p < .001$	$\chi^2 = 12.3$, $p < .001$
Mean frequency, correct words (SD)	495 (757)	456 (724)	412 (723)	373 (673)	370 (670)	373 (687)
Mean frequency, words with errors (SD)	131 (344)	129 (348)	151 (367)	140 (359)	137 (354)	153 (372)
Statistical difference in mean frequency	$t = 4.0, p < .001$	$t = 4.0$, $p < .001$	$t = 3.1$, $p = .002$	$t = 3.3$, $p = .001$	$t = 3.3$, $p = .001$	$t = 2.9$, $p = .004$
Errors in set 1 (62 words)	62	51	**45**	48	42	46
Errors in set 2 (61 words)	61	46	53	**30**	41	43
Errors in set 3 (61 words)	61	50	48	52	**37**	43
Errors in locative relations (24 times)	9			8	8	7

Cueing with the start sound of the word was effective in bringing about a correct naming in 73% of the attempts. Qualitatively it was noted that he often tried to explain when he could not find the word, but the explanations were very vague. Thus, both ''mushroom'' (''svamp'') and ''kangaroo'' (''kænguru'') were explained as ''such one out in the forest'' (''sådan en ude i skoven''). It should be noted that JI had relatively higher Spontaneous Speech and relatively lower Word Fluency scores on the WAB than KB, which also points towards a qualitative difference in the impairments in the two patients. JI might therefore have a special kind of impairment either within the semantic system or in the connection to the phonological output lexicon. From baseline to pre-therapy the naming response changed from correct to error on 27 words and from error to correct on 37 words (instability index: 0.35). From post-therapy to follow-up the changes were 48 and 36 (instability index 0.57). These are the lowest instability indices of the three patients in the study and suggest problems with internal representations rather than access problems. JI had a better WAB Auditory Comprehension score (9.9) than the two other patients (6.8 and 7.5), which makes it unlikely that there was a disturbance in the semantic system. The anomia could, however, result from a deficient semantic feature specification of the concepts, resulting in too broad a conceptual field being supplied to the phonological output lexicon (as in Luria's 1972 theory of amnesic aphasia). The impaired representations would then be internal to the semantic lexicon (if that is assumed to be separate from the semantic system).

Rehabilitation

A baseline was established (Test 1) on 13 November 1997. Due to the large number of errors, pictures with errors were split into three matched sets consisting of 62, 61, and 61 words, with the same number of pictures (26) depicting animate objects in each. The mean frequency of the words in the three sets was not significantly different (one-way anova, 171.2 SD 495.8, 103.4 SD 207.6, and 118.2 SD 255.0, $F = 0.66$, $p = 0.52$). Naming of these three sets of pictures was subsequently trained in separate periods in a multiple cross-over design. Training of the first set was initiated on 4 December 1997 when Test 2 was performed. It was trained to criteria for 66 hours until 25 February 1998 when Test 3 was performed. The second set was trained to criteria for 103 hours until 6 May when Test 4 was performed. The third set was trained to criteria for 23 hours until 8 June 1998 when Test 5 was performed. In the follow-up period locative relations was trained for 32 hours until 17 August 1998 when Test 6 was performed.

Results

Table 4 describes JI's performance on naming of the 260 S&V pictures before baseline, before and after training of three sets of words (two times cross-over), and after follow-up. Looking at all 260 words, JI improved from 174 errors before training to 147 errors after training ($\chi^2 = 9.2$. $p = .002$), but deteriorated to 160 errors after follow-up, the difference from pre-therapy now being insignificant. Looking at the 184 trained words alone, JI improved from 147 errors before therapy to 120 after ($\chi^2 = 13.0$, $p < .001$), but deteriorated to 132 errors after follow-up, although still significantly better than before therapy ($\chi^2 = 4.6$, $p = .03$).

The overall distribution of errors among the three sets of trained words differs over the six testing points (Cochran's Q = 99.4, $p < .001$) allowing for multiple tests for significant differences within the distribution. With training of set one, 9 words improved (14%), 50 were unchanged (81%), and 3 deteriorated (5%). Among the control words (sets two and three), 10 improved (8%), 97 were unchanged (80%), and 15 deteriorated (12%), an insignificant difference. With training of set two, 24 words improved (39%), 36 were unchanged (59%), and 1 deteriorated (2%). Among the control words (sets one and three), 13 improved (11%), 90 were unchanged (73%), and 20 deteriorated (16%), a significant difference in favour of the treated set (χ^2 = 25.6, p < ,001). With training of set three, 17 words improved (28%), 41 were unchanged (67%), and 3 deteriorated (5%). Among the control words (sets one and two), 13 improved (11%), 92 were unchanged (75%), and 18 deteriorated (15%), a significant difference in favour of the treated set (χ^2 = 11.2, p = .004).

Table 4 shows that there was a significant effect of both word frequency and animacy on all testing points except one, and neither seemed to have changed after therapy.

Discussion

The improvements found were statistically significant, but clinically only modest. The fact that the patient trained a considerable number of hours underlines the modesty of the effect. Moreover, there was a significant regression during follow-up, putting the lasting effect of the therapy into doubt.

There are several possible reasons for the somewhat disappointing result. This patient had very severe anomia and was trained on a large number of words. Even though the words were divided into three sets, there still were 61 words in each set. This might very well be too many words to train at the same time. Seron et al. (1979) have argued that using a small number of items will be the most effective in rehabilitation of anomia (but we are not aware of any empirical test of this hypothesis). The less convincing result in this patient might be caused by this factor, although it could also be due to a mismatch of the type of training with the nature of this patient's anomia.

JI needed a significantly shorter time to reach the criteria for the third set of words, which could be interpreted as some kind of generalisation from one set of words to another—still another indication that the changes were taking place within the semantic lexicon. The training was probably not entirely appropriate for this patient. It is possible that he would have achieved more improvement with a higher percentage of semantic tasks, which also had closer semantic foils (more true semantic tasks), and also with a different type of semantic task that focused more on semantic feature specification, as this type of training supposedly targets the semantic lexicon rather than access/connection problems.

STUDY 3

RI was 68 years old and a retired skilled worker in a supervisory position when he was admitted to hospital on 6 June 1997 with anomia, some difficulty following instructions, and a right-sided homonymous hemianopsia. Further assessments the next day suggested some apraxia, agraphia, and stereognosis was found to be absent on the right side. A CT-scan on the same day showed a large and clearly demarcated infarct in the left parieto-occipital region including the cortex. Speech therapy was instituted on 10 June. The patient was discharged to his own home on 26 June. He was assessed with WAB on 22 December. An anomic aphasia was found with an AQ of 76.3 (see Table 5 for details). Qualitatively, it is noted in the S&V naming test (Table 6) that RI answered a few times with a correct English word when he did not find the Danish word. He had a substantial number of partial word answers, and sometimes he slowly arrived at the correct word after attempts with partial words and phonological distortions. He also had a high number

TABLE 5
Western Aphasia Battery scores for RI

	Before rehabilitation 22 Dec 97	*After rehabilitation* 24 Apr 98
Type of Aphasia	Anomic	Anomic
Aphasia Quotient (AQ. max. 100)	76.3	78.4
I. Spontaneous Speech (max. 20)	15	16
a. Information Content (max. 10)	7	8
b. Fluency (max. 10)	8	8
II. Auditory Verbal Comprehension (max. 10)	7.45	7.6
a. Yes/No Questions (max. 60)	60	58
b. Auditory Word Recognition (max. 60)	51	50
c. Sequential Commands (max. 80)	38	44
III. Repetition (max. 10)	8	8.3
IV. Naming (max. 10)	7.7	7.3
a. Object Naming (max. 60)	56	53
b. Word Fluency (max. 20)	1	4
c. Sentence Completion (max. 10)	10	8
d. Responsive Speech (max. 10)	10	8

of semantic errors on S&V, but this does not by itself indicate a semantic system impairment. Some of the semantic paraphasic errors may be attempts at compensation, but other cases are clearly genuine semantic errors undetected by the patient. During testing of this patient, the first author got the impression that his naming difficulties in spontaneous speech were disproportionally large compared to his confrontation naming abilities. This is also reflected in the test profile before therapy: although RI had a considerably better confrontation naming score with the S&V pictures than JI, his ability to convey information on the WAB (Informational Content) was no better than that found for JI. However, like JI, RI had quite a low Word Fluency score before therapy. On the PALPA, RI had 40/60 correct on Auditory Sentence Comprehension (PALPA 55) and 13/24 correct in Auditory Comprehension of Locative Relations (PALPA 58). On the basis of this, it seems that this patient needed a very high level of activation from the semantic system to the phonological output lexicon in order to find the word, which might primarily be an access problem. This hypothesis is supported by a high instability index: From post-therapy to follow-up the naming response changed from correct to error on 27 words and from error to correct on 15 words (instability index: 0.95; we unfortunately have no baseline). That is by far the highest instability index among the three patients in this study.

Rehabilitation

Pre-training testing was started on 16 December 97. No baseline was established, but a cross-over design was used. Two matched sets of 34 pictures each were made using the pictures that RI was unable to name on the S&V. The first set was trained to criteria in 36 hours, and the next set was trained to criteria in 23 hours. A follow-up testing was also performed after a period of control training with locative relations tasks. This therapy proved to be extremely frustrating for RI. He had very severe difficulty carrying out the tasks, and thus had little opportunity to gain from them. The wrong kind of therapy might have been chosen, or it could be that the impairment is resistant to therapy. It could be

TABLE 6

Oral naming responses for 260 S&V pictures, error-types, animacy and frequency effects for RI

	Before training of word set 1 16 Dec 97	*Before training of word set 2* 20 Jan 98	*After training of word set 2* 20 Feb 98	*After follow-up* 3 Apr 98
No. correct	191	214	216	204
Total no. errors	69	46	44	56
No response, partial response	38 (55%)	27 (59%)	20 (45%)	36 (64%)
Verbal paraphasia	22 (32%)	16 (35%)	18 (41%)	14 (25%)
Literal paraphasia	2 (3%)	0	0	0
Mixed and neologic paraphasia	7 (10%)	3 (7%)	6 (14%)	6 (11%)
Errors, animate pictures	29 (30%)	21 (22%)	21 (22%)	26 (27%)
Errors, inanimate pictures	40 (25%)	25 (15%)	23 (14%)	30 (18%)
Statistical difference in animacy	$\chi^2 = 0.9$, *NS*	$\chi^2 = 1.7$, *NS*	$\chi^2 = 2.5$, *NS*	$\chi^2 = 2.5$, *NS*
Mean frequency, correct words (SD)	292 (597)	260 (560)	274 (568)	276 (572)
Mean frequency, words with errors (SD)	88 (164)	132 (308)	57 (119)	98 (264)
Statistical difference in mean frequency	$t = 4.3$, $p < .001$	$t = 1.5$, *NS*	$t = 5.1$, $p = .001$	$t = 3.3$, $p = .001$
Errors in set 1, 34 words	34	**9**	12	17
Errors in set 2, 34 words	34	16	**10**	15
Errors in locative relations (24 items)	11	12	12	14

argued that the impairment has a non-verbal basis, and that language therapy tasks are therefore not appropriate.

Results

Table 6 describes RI's performance on naming of the 260 S&V pictures before training of set one, before training of set two (cross-over), after training of set two, and after follow-up (no baseline measure was taken). Looking at all 260, words RI improved from 69 errors before training to 44 errors after training ($\chi^2 = 8.3$. $p = .004$), but deteriorated to 56 errors after follow-up, the difference from pre-therapy now being insignificant. Looking only at the 68 trained words, RI improved from 68 errors before therapy to 22 after ($\chi^2 = 44.0$, $p < .001$), and deteriorated to 32 errors after follow-up, which is still significantly better than before therapy ($\chi^2 = 34.0$, $p < .001$). However, as no baseline was established, this way of analysis is questionable because of the factor of regression towards the mean. The overall distribution of errors was significantly different across the two sets at four testing points (Cochran's Q = 48.1, $p < .001$) allowing for multiple tests for significant differences within the distribution. With training of set one there was, as mentioned, no possibility of deterioration, as there were initially errors on all trained words. 25 words improved (74%) and 9 were unchanged (26%). Among the control words (set two), 18 improved (53%) and 16 were unchanged (47%), an insignificant difference. With training of set two, 9 words improved (26%), 22 were unchanged (65%), and 3 deteriorated (9%). Among the control words (set one), 3 improved (9%), 25 were unchanged (73%), and 6 deteriorated (18%), an insignificant difference. The WAB AQ score (Table 5) improved from 76.3 before rehabilitation to 78.4 after rehabilitation. As can be seen from Table 6, there was no significant difference in the proportion of errors in animate compared to inanimate words at any testing point, whereas words with errors had a significantly lower frequency in three out of four testing points.

Discussion

RI had only modest improvement in his naming after rehabilitation, and the specificity of the improvement of trained words over control words was insignificant. That there was some general improvement in naming is supported by the observation that a much shorter time was needed to reach criteria for the second set (23 hours) compared to the first set (36 hours). The small improvement in WAB score after rehabilitation could be due to chance error, but would otherwise support the more general effect of the training in this patient. As the training comprised both semantic and phonological components, we would expect it to be appropriate for a patient with problems in the semantic system–phonological output lexicon connection. It is unfortunate that we do not have a baseline for this patient. We only have the cross-over as a control, which is not sufficient when the improvements are more general.

GENERAL DISCUSSION

The expected, specific improvements for the trained words were found in two out of three patients. The third patient improved for one set of words, but not for the second set. As the studies had a cross-over design, and one of these a double cross-over, the results can also be summarised for all controlled training attempts, seven in all. Six out of seven training attempts had the expected, specific effect. Unfortunately, the chosen design does not allow us to determine whether the effect was due to some specific type of training tasks (e.g., semantic or phonologic) or all the tasks in conjunction.

A fourth patient, not reported here, was trained in the subacute period (beginning one month after the stroke). LK was 45 years old and had two infarcts in the left hemisphere. On the pre-therapy there was no frequency effect on the S&V naming test, and half the errors were verbal paraphasias. There was a large improvement in both trained and untrained words—probably due to spontaneous remission—but the improvement was significantly larger in the trained words. Due to the large general improvement a planned cross-over became irrelevant.

All three patients were tested with oral naming. Oral naming was never trained directly: the patients were not asked to say anything aloud during the training. There was thus a general transfer or generalisation from the computer tasks to a quite different test task. The computer task that was most similar to the test task was the written naming task. This confirms the results reported by Deloche et al. (1992). It should be noted that our replication was done with a different set of computer programs.

Not all patients had the same amount of improvement in naming after the training. Several explanations for these differences can be suggested. In one case (JI), it may be that too many words were trained at the same time. In another case (RI), the patient was not very motivated, possibly because his naming difficulties were disproportionally more severe in connected, spontaneous speech compared to confrontation naming. In all training sets in the present study, the foil pictures were taken from the training set. This left very little possibility to find close semantic foils. Thus, the semantic part of the training might not have been optimally effective, which is also indicated by the lack of generalisation to untrained words.

The training tasks were probably most appropriate for KB, who also demonstrated the most clear effect. For JI it would probably have been more appropriate to focus on semantic training including semantic feature specification. The combination of semantic and phonological training might have been appropriate for RI, but the less than optimal semantic part could explain the small effect. Perhaps this patient's primary deficit was in the access to the semantic lexicon, thus making the therapy even less optimal. Although the patients' anomias were to some degree different, there was also an important common feature: All three had relatively well-preserved comprehension, supposedly precluding any severe deficit within the semantic system.

It is ironic that we started out believing that computer rehabilitation might be feasible in anomia rehabilitation because it is relatively easy to transfer the semantic part to computer tasks—and then ended up with the best results in the patient with the most clear-cut phonological deficits. It implies that it is feasible to train phonological deficits by the written approach, and that computer-training could be effective both in the phonological and the semantic part if more appropriate semantic tasks and foils were chosen.

We probably made an unfortunate choice of the area for control therapy. As noted in the discussion of RI, the basic impairment may really be non-verbal in nature, in which case the area is not suitable as a control area for language therapy, and some kind of non-verbal spatial therapy would be more appropriate than language-oriented therapy. Whatever the reason, the locative relations area did not work well as a control area: KB showed some improvement in the entire period, JI showed very minimal improvement over the period, and RI, in fact, deteriorated in his performance. Thus, there was no specific improvement in locative relations after the training of this area in the follow-up period. Locative relations may simply be unsuitable for language rehabilitation, and then it is not a very good control task in a multiple baseline design.

All testings were carried out by the first author, who also selected and designed the training tasks and, thus, was neither blinded nor disinterested in the results. This could influence the total scores but probably not the relative scores between trained words and control words, as there was no risk that the selected words for the different training periods for the different patients could be remembered by the tester during testing.

The follow-up period was not very long, and some regression in scores were seen. It may be that a longer follow-up period would show a total elimination of the gains. As the training was carried out with therapist involvement only in the planning and set-up phase, it would be possible to counter the regression by continued training in the patient's own home (and possibly even to foster continued improvement).

The oldest patient in the study was aged 71 years. He was highly motivated and had no difficulty in using the computer. This indicates that computer-rehabilitation is possible even in elderly stroke patients. However, this might not be true for all computer rehabilitation programs. The programs in this project were designed to be very easy to use, based on previous experience with computer rehabilitation of stroke patients.

As noted in the introduction, this study was motivated by the assumption that the use of specialised computer programs makes it possible to deliver theoretically well-motivated rehabilitation of aphasia with very little therapist intervention. The results do not, in our opinion, contradict this assumption. The same system of computer programs for rehabilitation is at present being tested with a larger set of patients in a randomised, placebo-controlled study.

Manuscript received 8 February 1999
Manuscript accepted 11 August 2000

REFERENCES

Basso, A. (1989). Therapy of aphasia. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology*, *Vol. 2*. Amsterdam: Elsevier.

Benson, D.F., & Barber, J. (1996). *Aphasia. A clinical perspective*, New York: Oxford University Press.

Bergenholtz, H. (1992). *Dansk frekvens ordbog [Danish dictionary of word frequencies].* Copenhagen: Gad.

Caplan, D. (1992). *Language. Structure, processing, and disorders*. Cambridge, MA: Bradford.

Deloche, G., Ferrand, I., Metz-Lutz, M.N., Dordain, M., Kremin, H., Hannequin, D., Perrier, D., Pichard, B., Quint, S., Larroque, C., Cardebat, D., Naud, E., Bergego, C., Pradat-Diehl, P., Tessier, C. (1992). Confrontation naming rehabilitation in aphasics: A computerised written technique. *Neuropsychological Rehabilitation*, *2*, 117–124.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd Edn.). Philadelphia: Lea & Febiger.

Howard, D., Patterson, K., Franklin, S., Orchard Lisle, V., & Morton, J. (1985a). The facilitation of picture naming in aphasia. *Cognitive Neuropsychology*, *2*, 49–80.

Howard, D., Patterson, K., Franklin, S., Orchard Lisle, V., & Morton, J. (1985b). Treatment of word retrieval deficits in aphasia. A comparison of two therapy methods. *Brain*, *108*, 817–829.

Kay, J., Lesser, R., & Coltheart, M. (1992). *PALPA. Psycholinguistic Assessments of Language Processing in Aphasia*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Kertesz, A. (1982). *Western Aphasia Battery.* New York: Harcourt Brace Jovanovich (The Psychological Corporation).

Loverso, F.L., Prescott, T.E., & Selinger, M. (1992). Microcomputer treatment applications in aphasiology. *Aphasiology*, *6*, 155–163.

Luria, A.R. (1972). Aphasia reconsidered. In A.R. Luria (Ed.,) *Cortex*, *8*, 34–40.

Nettleton, J., & Lesser, R. (1991). Therapy for naming difficulties in aphasia: Application of a cognitive neuropsychological model. *Journal of Neurolinguistics*, *6*, 139–157.

Nickels, L. (1997). *Spoken word production and its breakdown in aphasia*, Hove, UK: Psychology Press.

Nickels, L., & Best, W. (1996a). Therapy for naming disorders (Part 1): Principles, puzzles and progress. *Aphasiology*, *10*, 21–47.

Nickels, L., & Best, W. (1996b). Therapy for naming disorders (Part II): Specifics, surprises and suggestions. *Aphasiology*, *10*, 109–136.

Seron, X., Deloche, G., Bastard, V., Chassin, G., & Hermand, N. (1979). Word-finding difficulties and learning transfer in aphasic patients. *Cortex*, *15*, 149–155.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardised set of 260 pictures: Norms for naming agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

SPSS Inc. (1999). *SPSS for Windows. 10.0* [computer program].

Willmes, K. (1995). Aphasia therapy research: Some psychometric considerations and statistical methods of the single-case study approach. In C. Code & D. Muller (Eds.), *Treatment of aphasia: From theory to practice*. London: Whurr Publishers.