Acta Psychologica 138 (2011) 135–142

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Retroactive interference in short-term memory and the word-length effect

Guillermo Campoy *

Universidad de Murcia, Campus de Espinardo, Facultad de Psicología, 30100 Murcia, Spain

ARTICLE INFO

Article history:
Received 8 June 2010
Received in revised form 4 February 2011
Accepted 26 May 2011
Available online 23 June 2011

PsycINFO classification:
2300
2340
2343

Keywords:
Short-term memory
Word-length effect
Retroactive interference
Proactive interference
Serial reconstruction
Recent-probes task

ABSTRACT

Two experiments investigated the possibility that the word-length effect in short-term memory (STM) is a consequence of long words generating a greater level of retroactive interference than shorter words. In Experiment 1, six-word lists were auditorily presented under articulatory suppression for immediate serial reconstruction of only the first three words. These three words were always drawn from a single set of middle-length words, whereas the last three positions were occupied by either short or long interfering words. The results showed worse memory performance when the to-be-remembered words were followed by long words. In Experiment 2, a recent-probes task was used, in which recent negative probes matched a target word in trial *n*-2. The results showed lower levels of proactive interference when trial *n*-1 involved long words instead of short words, suggesting that long words displaced previous STM content to a greater extent. By two different experimental approaches, therefore, this study shows that long words produce more retroactive interference than short words, supporting an interference-based account for the word-length effect.

1. Introduction

The word-length effect, which is the finding that immediate serial recall is better for short words than for long words (Baddeley, Thomson, & Buchanan, 1975), is one of the key phenomena in the development of short-term memory (STM) theories. Unfortunately, however, the specific mechanisms underlying this effect have not been satisfactorily determined despite a number of different explanations have been proposed over the last decades.

The word-length effect was first explained within the working memory model (Baddeley & Hitch, 1974). According to this model, verbal STM relies on the phonological loop, consisting of a phonological store where phonological traces are maintained and a control process of subvocal rehearsal that reactivates traces and counteracts decay (Baddeley, 1986; Burgess & Hitch, 1999). The time that a word takes to be rehearsed is proportional to how long the word takes to be pronounced. Rehearsal, therefore, is faster for short words than for long words, and, consequently, rehearsal is more efficient in reactivating short words before some items are lost due to decay. This would be why lists of short words are better recalled than lists of long words.

Preliminary evidence for this view emerged from the observation that the length effect could be obtained when short and long words were matched for the number of syllables and phonemes (Baddeley et al., 1975). Normally, long and short words differ in the number of syllables and phonemes, so this finding was crucial to support that notion that articulatory duration was actually the key length factor in the origin of the effect. The role of rehearsal was later supported by the finding that the effect disappeared when rehearsal was prevented by articulatory suppression, a procedure in which participants are asked to articulate an irrelevant word or phrase over and over during the memory task (Baddeley, Lewis, & Vallar, 1984). Both lines of evidence, however, have been challenged by subsequent studies. On the one hand, a number of experiments have repeatedly failed to find length effects with stimuli matched for number of syllables and phonemes (Caplan, Rochon, & Waters, 1992; Lovatt, Avons, & Masterson, 2000; Neath, Bireta, & Surprenant, 2003; Service, 1998), leading to the idea that the initial results might be a consequence of some special features of the original stimulus set (however, see Baddeley, 2007). On the other hand, some studies have shown that, under specific situations, the length effect can be found despite rehearsal prevention, for example, when lists of non-words instead of words are presented under articulatory suppression (Romani, McAlpine, Olson, Tsouknida, & Martin, 2005) or when rehearsal is prevented by temporal restrictions instead of suppression (Campoy, 2008; Coltheart & Langdon, 1998; Coltheart, Mondy, Dux, & Stephenson, 2004).

In addition to rehearsal, the word-length effect has also been explained by differences in the duration of recall, a factor usually

* Tel.: +34 868 888 769; fax: +34 868 888 161.
E-mail address: gcampoy@um.es.

0001-6918/$ – see front matter
doi:10.1016/j.actpsy.2011.05.016

referred to as output delay. Supporters of this view argue that long word lists are recalled worse because they take more time to be reported, and, therefore, stored words suffer more from decay during recall (Cowan, Day, Saults, & Keller, 1992; Dosher & Ma, 1998). The primary evidence for this hypothesis comes from a study by Cowan et al. (1992) in which word length was manipulated independently for the first and the second half of each list. According to the output delay hypothesis, worse recall was predicted when the first words to be recalled were long, given that the rest of the words would be more affected by decay in this condition. Although the results were congruent with this prediction, a subsequent study by Lovatt, Avons, and Masterson (2002) showed that this finding was hardly replicated, while, when found, the effect disappeared if errors in the recall of the first items were controlled. On the other hand, some experiments by Lewandowsky and colleges also cast doubts on the role of output delay by showing no effect of delaying retrieval on memory performance despite rehearsal prevention (Lewandowsky, Duncan, & Brown, 2004; however, see Cowan & AuBuchon, 2008). Finally, the role of output delay as a main factor in the word-length effect has been challenged by a number of experiments showing consistent effects with procedures in which short and long words do not differ in output delay, such as serial recognition (Baddeley, Chincotta, Stafford, & Turk, 2002; Campoy, 2008) and serial reconstruction (Campoy & Baddeley, 2008; Lovatt et al., 2000).

Accounts based on rehearsal or output delay are known as time-based explanations (Hulme, Surprenant, Bireta, Stuart, & Neath, 2004), given that the key factor is the duration of the stimuli. Subsequent explanations proposed that the word-length effect is actually a consequence of the phonological complexity of the items (e.g., number of phonemes or syllables) rather than their duration (Caplan et al., 1992; Hulme et al., 2004; Service, 1998). These accounts are known as complexity-based explanations (Hulme et al., 2004).

Regarding the question of how item complexity results in length effect, Romani et al. (2005) suggested that long words are recalled worse because complex words comprise a larger number of segments; hence, it is more likely that some of these segments are lost before recall. According to this explanation, therefore, lists of long words are recalled worse because of the lower individual level of *recallability* of each long word (for other item-based explanations, see Brown & Hulme, 1995; Neath & Nairne, 1995). This kind of explanation, however, seems to be ruled out by experiments showing that long words are not recalled worse than short words when both long and short words are presented in mixed lists (Hulme et al., 2004; see also Bireta, Neath, & Surprenant, 2006; Cowan, Baddeley, Elliott, & Norris, 2003).

A more viable complexity-based explanation has been proposed by Hulme and collaborators (Hulme et al., 2004; Hulme et al., 2006). According to this explanation, the recall of an item depends on the relative distinctiveness of its memory trace in relation to the other traces in the *retrieval set*. As far as the word-length effect is concerned, distinctiveness is a consequence of phonological complexity, with more complex items resulting in lower levels of distinctiveness among traces. This account has the advantage of including the word-length effect in a more general framework; thus, results in a range of situations are explained by a single factor (distinctiveness). However, the reasons why more complex items are less distinctive still remain to be clearly established. Regarding this question, Hulme et al. (2004, p. 103) suggested that it is more difficult to maintain distinctive traces of long (more complex) words because they comprise more phonological information, and there is a limitation regarding how much phonological information can be maintained at the same time. Unfortunately, the precise nature of this limitation has not been determined.

Finally, Tehan and colleagues (Hendry & Tehan, 2005; Tehan & Tolan, 2007) suggest that the word-length effect is a consequence of the fact that processing long words is more demanding and/or takes longer than processing shorter words. They assume that item and order information are processed separately; thus, there can be some trade-off between the processing of both kinds of information. Because long words are harder to process, item processing is accomplished at the expense of order processing, resulting in worse order encoding for long words. This would be why long words are recalled worse in standard paradigms, which typically include serial recall. This hypothesis, however, is inconsistent with a previous study by Russo and Grammatopoulou (2003) in which participants were required to recall lists of short and long items either serially or freely. According to the item/order trade-off account, the word-length effect should be eliminated or, at least, reduced in free-recall situations, given that the length effect is a consequence of poorer order information for long words. Russo and Grammatopoulou (2003) did not compare serial and free recall directly, but estimations of the size of the word-length effect calculated from the paired *t*-tests in each recall condition reveal that the effects tended to be even larger in free-recall conditions. In their Experiment 1, for example, the size of the word-length effect with immediate free recall was virtually twice the size of the effect with immediate serial recall (3.14 vs. 1.58).[1] Further evidence against the account by Tehan and colleagues emerges from error analyses (Hulme et al., 2006). According to the item/order trade-off account, it could be predicted that lists of long words would be recalled worse mainly because of order errors. In contrast to this prediction, Hulme et al. (2004, 2006) found that the number of order errors for lists of short and long words was virtually identical, with the worse recall of long words being essentially a consequence of more omissions.

From all of the above, it becomes apparent that we still lack a fully satisfactory understanding of the mechanisms underlying the word-length effect. Further investigation is necessary, therefore, to test more deeply the different proposals put forward so far and to explore new factors that could be involved in the origin of the effect. The latter was the objective of this study, which investigated the possible role of retroactive interference. Specifically, this study aimed to investigate the possibility that the longer (more complex) words generate greater levels of retroactive interference over information previously stored in STM, a possibility that would provide a novel explanation for the word-length effect. This objective was addressed by adopting two different but complementary approaches in two experiments.

2. Experiment 1

Experiment 1 was based on the general assumption that auditory verbal stimuli gain direct access to STM, even though they are not relevant for the task (e.g., Baddeley, 1986). In this experiment, lists of six words were auditorily presented for immediate memory test, but participants were previously instructed to remember only the first three items. The key aspect of this procedure was that the first three words were always drawn from a single set of middle-length words, whereas the last three positions were occupied by either short or long words. According to the retroactive interference hypothesis, it was expected that short and long words would gain access to STM and generate some degree of interference over the previously stored relevant information. The main question was whether long words would generate more interference than short words, with the consequence of worse memory performance for lists in which to-be-remembered words were followed by long words.

The procedure in Experiment 1 has some parallels to other experiments involving irrelevant verbal stimuli, like those on the irrelevant speech effect (Colle & Welsh, 1976) and the suffix effect (Dallett, 1965). In the present experiment, however, relevant and

[1] Effect sizes were calculated by the formula $d = T/\sqrt{n}$, where *T* is the paired *t*-test value and *n* is the number of participants (adapted from Morris & DeShon, 2002, Eq. 4, p. 107).

interfering stimuli comprised word lists that were presented as typical lists in standard experiments. This characteristic of the procedure contrasts with most of the irrelevant speech and suffix experiments, in which interfering stimuli usually differ in nature and/or belong to a different auditory stream. The fact that stimulus presentation mimicked that in standard word-length experiments is important to facilitate generalization. For the logic of the experiment, however, it was also important that short and long words did not have to be remembered. Otherwise, it would not have been possible to distinguish the interference effect generated by the mere presentation of long and short words from other effects that could emerge from maintenance operations. The most obvious of these sources of confusion would be rehearsal. If participants had to remember short and long words, supporters of the phonological loop model could allege that long words take more time to be rehearsal and, therefore, that the first three words would suffer more from decay when the words in the second half were long. In this regard, it is important to note that the possibility that short and long words were occasionally rehearsed could not be completely ruled out, even though participants did not have to recall these words. This fact could challenge the interpretation of the results, and, for this reason, articulatory suppression was introduced to prevent rehearsal.

2.1. Method

2.1.1. Participants

Twenty undergraduates from the University of Murcia (Spain) took part in the experiment. All participants were native Spanish speakers and participated for course credit.

2.1.2. Stimuli and apparatus

A total of 21 Spanish words were used: seven disyllabic nouns as short words, seven trisyllabic nouns as to-be-remembered words, and seven quadrisyllabic nouns as long words. All of the stimuli were composed of consonant-vowel syllables, with the stress on the penultimate syllable. Short and long words (the interfering words) were matched for word frequency, familiarity, imageability and concreteness using the LEXESP database (Sebastián, Martí, Carreiras, & Cuetos, 2000). To-be-remembered words were selected so that they were homogeneous in these variables. The phonological similarity of to-be-remembered words was minimized by ensuring that no words shared more than one phoneme in the same position. Thus, any subset of three to-be-remembered words included highly distinguishable stimuli, avoiding the problem of the occasional inclusion of similar words in the same list affecting performance. The short words were *dato*, *filo*, *jugo*, *lema*, *rito*, *roca*, and *seda* (datum, edge, juice, motto, rite, rock, and silk); long words were *desayuno*, *diputado*, *ligereza*, *negativa*, *retirada*, *sacudida*, and *soberano* (breakfast, deputy, lightness, denial, retreat, shake, and sovereign); and to-be-remembered words were *butaca*, *ceniza*, *cometa*, *garaje*, *molino*, *paloma*, and *tejado* (armchair, ash, comet, garage, windmill, pigeon, and roof).

All of the words, spoken in a neutral tone by a female, were recorded and segmented into individual sound files using an audio recorder and editor software. A computer program generated by E-Prime (Schneider, Eschman, & Zuccolotto, 2002) was used for stimulus presentation (via headphones) and response storage.

2.1.3. Procedure

The experiment consisted of four blocks of 14 trials, with the first block being a practice one. The sequence of events in each trial was as follows. First, a row of dashes was presented, indicating that a mouse button had to be pressed to initiate the trial. One second after the mouse click, an intermittent plus sign (500 ms on, 500 ms off) appeared in the center of the screen for three seconds. Participants were instructed to begin articulatory suppression at that moment by counting aloud from one to four at the plus sign rate (four numbers per second). Next, a six-word list was presented at a rate of one item per second. The three words in the first half of the lists were always to-be-remembered words, whereas the three words in the second half were either long or short words. Participants were informed in advance that they only had to remember the first three items and would not be tested on the last three words. After the list presentation (specifically, one second after the onset of the last item of the list), the seven stimuli from the set of to-be-remembered words appeared on the computer screen, and participants employed the mouse to click on the first three words of the list in their presentation order (clicked words appeared in bold). Participants were instructed to click on a question mark in place of any word they did not remember. Once the list reconstruction was finished, participants were required to click on an unlabelled black button and stop articulating.

Each block consisted of seven trials with short words and another seven trials with long words in random order. Lists were constructed by selecting words from the specific set at random, within the constraints that every to-be-remembered word appeared in six trials per block (once in each combination of serial position and interfering word length) and that every interfering word appeared in three trials per block (once in each position). Participants were tested individually in a sound-attenuated booth, with the experimenter monitoring the articulatory suppression performance from outside the booth.

2.2. Results

Two participants failed to follow the instructions for articulatory suppression (they ceased articulating during the serial reconstruction instead of afterwards). Data from these participants were removed from the reported analysis, although the results for all the participants were qualitatively equivalent.

Raw data were scored according to a strict serial recall criterion, by which a response was counted as correct when the right item was recalled in the correct serial position. Fig. 1 shows the mean percentages of correct recall at each serial position for the two conditions of interfering word length. These percentages were submitted to a 2 (interfering word length) × 3 (serial position) within-subjects analysis of variance (ANOVA). There was a main effect of word length, $F(1, 17) = 7.265$, $MSE = 104.042$, $p = .015$, $\eta^2_p = .299$, showing that to-be-remembered words were better recalled when they were followed by short words (see also Table 1).

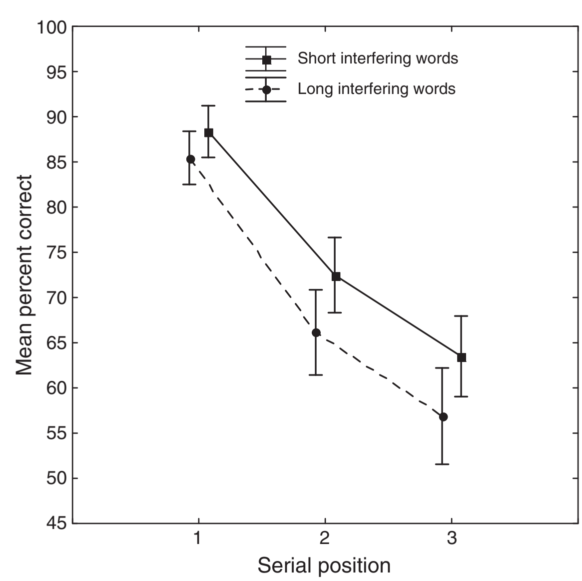

Fig. 1. Mean percentages of correct recall as a function of serial position and interfering word length in Experiment 1. Vertical bars depict standard errors of the mean.

Table 1
Mean percentage of the responses for each category (standard deviations in parentheses) and *t*-tests in Experiment 1.

	Correct responses	Order errors	Omissions	Intrusions
Short interfering words	74.780 (13.761)	10.758 (9.200)	8.907 (7.488)	5.556 (4.047)
Long interfering words	69.489 (16.202)	11.905 (8.698)	11.728 (9.542)	6.878 (5.632)
t-test (two-tailed)	$t(17)=2.695$ $p=.015$	$t(17)=1.127$ $p=.275$	$t(17)=2.215$ $p=.041$	$t(17)=1.516$ $p=.148$

There was also a main effect of serial position, $F(2, 34)=29.401$, $MSE=225.856$, $p<.001$, $\eta^2_p=.634$, reflecting a primacy effect. The interaction between word length and serial position was not significant, $F(2, 34)=1.024$, $MSE=37.534$, $p=.370$, $\eta^2_p=.057$.

To perform error analyses, the individual responses in the serial recognition task were categorized into four categories: correct responses (when a participant clicked on the correct word in the right position), order errors (clicks on words presented in the current list but recalled in the wrong position), omissions (clicks on the question mark), and intrusions (clicks on words outside the target list). Table 1 shows the percentage of responses in each category, together with the results of *t*-tests for differences between the two conditions of interfering word length. The observation of this table reveals that the three kinds of error tended to be numerically larger in the condition with long interfering words, although the difference reached statistical significance only for omissions.[2]

2.3. Discussion

The results of Experiment 1 showed that to-be-remembered words presented in the first half of the lists were recalled worse when they were followed by long interfering words. This finding is congruent with the idea that longer words generate greater levels of retroactive interference over previous content in STM. The effect of length was principally a consequence of differences in the number of omissions, whereas there was no difference in terms of order errors. This fact suggests that interfering word length had an effect mainly on memory regarding item identity.

Regarding the word-length effect, the main conclusion is straightforward. As far as we assume that the interference generated by the interfering words in this experiment is equivalent to that produced by list items in standard procedures, the present results support the idea that the word-length effect could be to some extent a consequence of the greater level of retroactive interference generated by long words.

Although the results in Experiment 1 seem to fit especially well within an interference framework, there is an alternative decay-based account that deserves closer examination. This alternative account would be based on the idea that common attentional resources support both the maintenance of STM content and input processing. In this situation, the presentation of each new input would involve the temporary neglect of the maintenance activity as the new input is processed, resulting in a diminution in the quality of memory traces representing previous content. Such a view is congruent with the time-based resource-sharing model (Barrouillet, Bernardin, & Camos, 2004; Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Barrouillet & Camos, 2008) and with diverse proposals suggesting attentional refreshing mechanisms (Hudjetz & Oberauer, 2007; Raye, Johnson, Mitchell, Greene, & Johnson, 2007). On the basis of this idea, the results of Experiment 1 could be explained by alleging that interfering words were necessarily processed and that the processing of interfering words was more demanding and/or took more time when those words were long (Tehan & Tolan, 2007). The presentation of long interfering words, therefore, resulted in larger maintenance neglect, with the consequence of more decay and poorer performance.

[2] The difference in order errors was still not significant after correcting for the number of errors as suggested by Saint-Aubin and Poirier (1999), $t(17)=1.742$; $p=.100$; two-tailed.

This alternative account could be considered unlikely in light of recent studies challenging the role of decay as a forgetting mechanism in STM (Berman, Jonides, & Lewis, 2009; Lewandowsky, Geiger, & Oberauer, 2008; Lewandowsky & Oberauer, 2009; Oberauer & Lewandowsky, 2008; for a review, see Lewandowsky, Oberauer, & Brown, 2009). Nonetheless, Experiment 2 aimed to find additional evidence for the hypothesis that longer words generate greater levels of interference by using a procedure that prevented the temporary neglect from being an alternative explanation.

3. Experiment 2

The temporary neglect account is a potential source of confusion for any interference study in which interfering material is presented while participants have to maintain to-be-remembered information in STM. In these situations, supporters of the temporary-neglect account can always attribute the interference effect to decay taking place while attentional resources are allocated to interfering stimulus processing at the expense of attentional refreshing. There are at least two approaches to tackle this problem. The first potential solution is to introduce an attention-demanding task to withdraw attentional resources from trace refreshing (e. g., Lewandowsky et al., 2008, Experiment 4; Oberauer & Lewandowsky, 2008, Experiment 2). The idea is that, in this situation, interference stimuli do not generate a disruption of the attentional refreshing mechanism because this mechanism is already blocked by the secondary task. The second approach is to evaluate the interference effect on STM contents that are no longer relevant at the time when the interfering stimuli are presented (Berman et al., 2009). The assumption in this case is that to-be-evaluated memory content is not actively maintained after it ceases to be relevant (Berman et al., 2009; Lewandowsky et al., 2009); therefore, neither attentional refreshing nor rehearsal need to be prevented. Although, in principle, both approaches are valid, the first approach has the inconvenience of making the task especially complex and unnatural because the participants have to perform three simultaneous tasks (articulatory suppression, attention-demanding task, and memory task). Thus, the second approach was the one adopted in Experiment 2.

The procedure employed in this experiment is an adaptation of the recent probes task introduced by Monsell (1978), which, in turn, is a modification of the item recognition paradigm of Sternberg (1966). Each trial began with the presentation of three words (the target word set). After an interval, participants were presented with a probe item, and they had to respond as fast and accurately as possible as to whether the probe matched a word in the target set. Our interest was focused on negative probe trials, that is to say, trials in which the probe does not match any item in the current target set. In some cases, negative probes did not match any item in both the current and previous target sets (in other words, they were non-recent negative probes). In other cases, however, negative probes matched a target word in trial *n*-2 (they were recent negative probes), and, therefore, participants had to provide a negative response in the face of some degree of proactive interference (see Fig. 2). This proactive interference was expected to materialize in the time required to give a negative response being longer for recent negative probes than for non-recent negative probes. The key aspect of the procedure was that, while trial *n* and *n*-2 always involved middle-length words, trial *n*-1

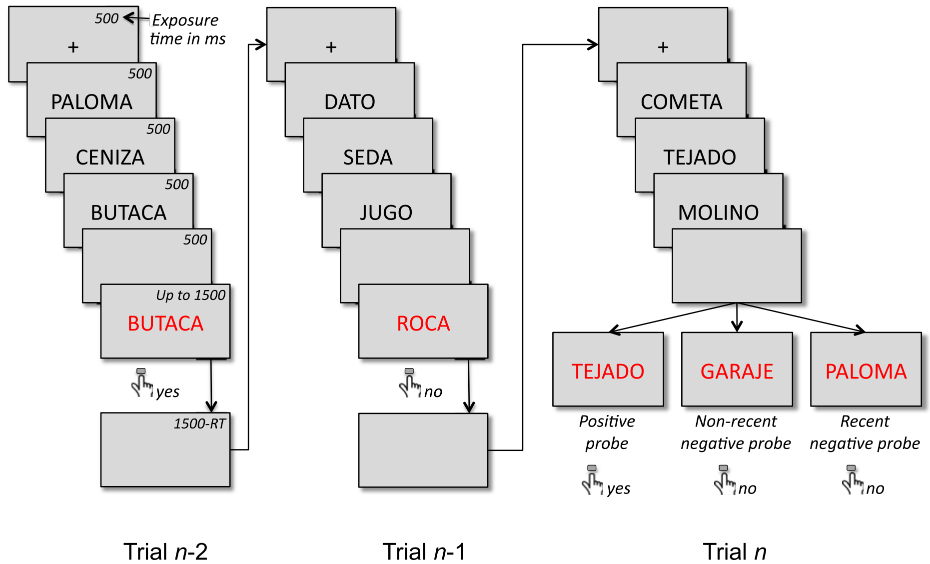

Fig. 2. Outlined representation of the experimental procedure in Experiment 2.

comprised either short or long words. If long words generated more retroactive interference over previous content (as suggested by Experiment 1), it was predicted that proactive interference coming from trial *n*-2 would be smaller when long words were presented in trial *n*-1 because long words displaced previous memory content to a greater extent than short words.

3.1. Method

3.1.1. Participants

Participants were 40 undergraduates from the University of Murcia who were native Spanish speakers and participated for course credit.

3.1.2. Stimuli and apparatus

Three sets of Spanish nouns were used: one comprising 468 middle-length words and the other two consisting of 252 short words and the same number of long words. These three sets of words were globally matched for word frequency, familiarity, imageability and concreteness according to the LEXESP database (Sebastián et al., 2000). A computer program generated by E-Prime (Schneider et al., 2002) controlled the experiment.

3.1.3. Procedure

The experiment comprised two blocks of trials, with each block including 144 pairs of trials (from the participants' perspective, there were 288 consecutive trials in a block, with no separation between pairs). In each pair of trials, the first trial (trial *n-1*) involved either short or long words, whereas target words in the second trial (trial *n*) were always middle-length words (trials *n* became trials *n*-2 for the following pair of trials). The 144 pairs within a block represented all the combinations of word length in trial *n*-1 (short or long), type of probe in trial *n*-1 (positive or negative), serial position of the positive probe in trial *n*-1 (1, 2, or 3; this was meaningless for negative-probe trials), type of probe in trial *n* (positive or negative), serial position of the positive probe in trial *n* (1, 2, or 3; this was meaningless for negative-probe trials), and type of negative probe in trial *n* (recent or non-recent; this was meaningless for positive-probe trials). Positive probes matched a word in the target set, whereas negative probes did not. Negative probes in trial *n*-1 were always non-recent negative probes, whereas negative probes in trial *n* could be either non-recent or recent. Non-recent negative probes did not match any word previously presented in the current block of trials, whereas recent negative probes were drawn at random from among the target words in trial *n*-2 (excluding words used as probes).

Each trial consisted of the following sequence (see also Fig. 2). First, a fixation cross was presented in the center of the screen for 500 ms. Next, the three words composing the target set were visually presented at a rate of one word every 500 ms (467 on, 33 off). Then, after a blank interval of 500 ms, a probe word was presented in red font. At this moment, participants indicated if the probe matched one word in the target set or not by pressing the *z* key or the *m* key on the computer keyboard as quickly and accurately as possible (the specific response-key mapping was counterbalanced across participants). The probe stimuli remained visible until a response was made or until a maximum period of 1500 ms had elapsed. Finally, the screen went blank until 1500 ms had passed from the onset of the probe presentation, and a new trial began.

Each experimental word appeared once per block, with the obvious exception of words used as positive and recent negative probes (which appeared twice). Experimental trials were preceded by 30 practice trials in which different sets of words were used.

3.2. Results and discussion

The analyses of interest were those focused on determining if the length of the words in trial *n*-1 modulated the level of proactive interference coming from trial *n*-2. Thus, only analyses involving participant's performance on recent and non-recent negative probes are reported here (other results will be provided on request). Reaction times (RTs) above or below three standard deviations from the participant's mean were not included in the analyses. Fig. 3 shows the mean RTs for correct responses as a function of the type of negative probe (recent or non-recent) and the word length in trial *n*-1 (short or long). These means were submitted to a 2×2 repeated-measures ANOVA with type of probe and word length as factors.[3] There was a main effect of type of probe, $F(1, 39) = 58.872$, $MSE = 832.923$,

[3] An equivalent analysis in terms of the number of errors was also performed. This analysis only revealed a general effect of proactive interference but failed to show a significant interaction between the type of probe and the word length. This fact was not surprising because accuracy scores are usually less sensitive than RTs in Sternberg-like tasks. Importantly, however, the proactive interference effect was numerically greater in the short-word condition, showing that there was no trade-off between speed and accuracy.

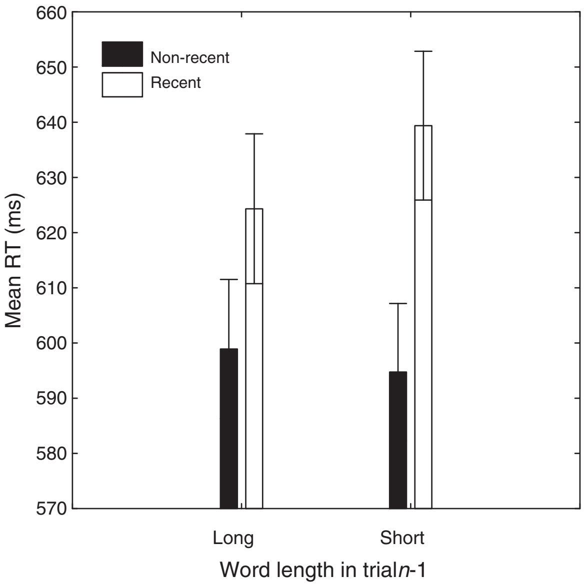

Fig. 3. Mean RTs as a function of the type of negative probe (recent, non-recent) and the word length in trial n-1 (short, long). Vertical bars represent mean standard errors.

$p<.001$, $\eta^2_p = .602$, showing the expected proactive interference effect (consisting of the correct responses being faster for non-recent negative probes than for recent negative probes, mean difference = 35 ms). The main effect of word length was not significant, $F(1, 39) = 1.872$, $MSE = 634.423$, $p = .179$, $\eta^2_p = .046$, but there was an interaction between type of probe and word length, $F(1, 39) = 5.580$, $MSE = 6631.702$, $p = .023$, $\eta^2_p = .125$. As shown in Fig. 3, this interaction was due to a greater proactive interference effect in the short-word condition, while post hoc comparison using the Fisher LSD test ($MSE = 661.70$, $df = 39$) revealed a significant proactive interference effect in the two conditions of word length (both p's < .001; mean difference = 45 ms and 25 ms for short- and long-word conditions respectively). Fig. 3 also suggests that differences between word-length conditions occurred on recent negative probes. Effectively, post hoc tests confirmed that responses for recent negative probes were significantly slower in the short-word condition ($p = .013$, mean difference = 15 ms), whereas RTs for non-recent negative probes were equivalent for the two levels of word length ($p = .447$, mean difference = −4 ms).

Confirming the prediction, Experiment 2 showed that the presentation of longer words in an interpolated trial led to a significant reduction in the level of proactive interference coming from previous memory content. As in Experiment 1, this finding suggests that longer words resulted in more displacement of items previously stored in STM.

4. General discussion

The past few years have seen a renewed debate on the question of whether forgetting from verbal STM is a consequence of temporal decay or interference (Barrouillet et al., 2007; Berman et al., 2009; Cowan & AuBuchon, 2008; Lewandowsky et al., 2008; Lewandowsky et al., 2009; Oberauer & Lewandowsky, 2008). The word-length effect is especially relevant for this debate (e.g., Lewandowsky & Oberauer, 2008) because this effect has been traditionally considered key evidence for decay models (e.g., Baddeley, 1986). In this context, this study might be of interest because it aimed to evaluate a possible interference-based account for the word-length effect, an account according to which the effect would be due to longer (more complex) words generating a greater level of retroactive interference over previous memory content.

Two experiments investigated this issue. In Experiment 1, six-word lists were auditorily presented for immediate serial reconstruction of the first three items. Lists were constructed so that to-be-remembered words in the first half of the lists were middle-length words, whereas interfering words in the second half were either short or long words. Articulatory suppression was used to prevent the situation in which the results were affected by the occasional rehearsal of long and short words. The results showed better memory performance when to-be-remembered words were followed by short words. In Experiment 2, short and long words were presented in a recent probes task. The analysis of the reaction times showed that the level of proactive interference arising from the two-back trial was lower when long words instead of short words were presented in the interpolated trial, suggesting that long words displaced previous STM content to a greater magnitude. Thus, the results of the two experiments were consistent with the idea that short and long words differ in the level of retroactive interference that they generate, with longer words resulting in more interference. Importantly, the two experiments yielded congruent results despite adopting radically different methodologies, suggesting that they reflect a fairly general phenomenon. As far as the word-length effect is concerned, this study supports the idea that the effect could be a consequence of the greater level of retroactive interference generated by long words.

Though the present study seems to provide strong evidence for the possible role of retroactive interference in the word-length effect, one might object that the present findings are at odds with previous experiments in which word length was varied independently for the first and the second half of the lists, similar to Experiment 1 (Cowan et al., 1992; Cowan, Wood, & Borne, 1994; Cowan, Wood, Nugent, & Treisman, 1997; Lovatt et al., 2002). These experiments were designed to test the output delay account of the word-length effect (described above) and differed from Experiment 1 in that participants had to recall the whole lists and not only the first items. Despite this difference, the first impression is that retroactive interference should still result in worse memory performance for lists with longer words in the second half. However, the results were not consistent with this, with some experiments showing this pattern (Cowan et al., 1994; 1997) and others showing the opposite pattern (Cowan et al., 1992; Lovatt et al., 2002). It is important to note, however, that the effect of retroactive interference in these studies could be obscured by other unrelated phenomena. Previous studies, for example, compared SS lists (lists having short words in both the first and the second half) with SL lists (lists having short words in the first half and long words in the second half). In this situation, retroactive interference could result in worse recall for SL lists. However, SL lists probably benefited from the larger level of distinctiveness, which is known to improve immediate recall (Hulme et al., 2004; 2006). These previous studies also compared SL lists with LS lists (lists having long words in the first half and short words in the second half), which were equivalent in terms of distinctiveness. In this situation, LS lists were presumably less affected by retroactive interference during stimulus presentation because the last words presented were short. However, supporters of the phonological loop model could argue that rehearsal was less efficient for those lists because the rehearsal loop includes longer words for longer periods of time. Moreover, it seems reasonable to argue that interference could also operate at the output (see below), with this output interference being larger when the first items to be recalled are long. In short, the results of previous experiments probably represent the combination of a number of effects of different signs and do not provide useful information about the role of interference.

A different issue concerns previous experiments showing that the level of retroactive interference grows as the number of different interference items increases (Lewandowsky et al., 2008). These previous results seem congruent with the present finding that longer (more complex) words generate more interference over previous STM content. An interesting possibility is that both findings represent different expressions of the same underlying principle, namely, that the level of interference depends on the amount of interfering material (e.g., the number of phonemes). From this viewpoint, more interfering material would result in more interference, regardless of whether this larger amount of interfering material comes from the greater number of words or from longer (more complex) words. Let us assume, as a simplification, that phonological storage in STM relies on a limited pool of *representation units*. We could speculate that interference is a consequence of *overwriting*, a phenomenon that would take place when representation units participating in the representation of previous items in STM are reused or overwritten as a response to new inputs (Lange & Oberauer, 2005; Nairne, 1990, 2002; Neath & Nairne, 1995; Oberauer, 2009; Oberauer & Kliegl, 2006). In this situation, it seems plausible that the presentation of more interfering material would result in a larger number of representation units being reused, with the consequence of a greater decline in the quality of the previous representations. A similar argument could be formulated by supporters of interference by *superposition* (Farrell & Lewandowsky, 2002), which is based on the assumption that interference is a consequence of distributed representations of the items being superimposed on each other. Whereas superposition is supposed to have a global effect, overwriting is generally assumed to involve shared features (Lewandowsky et al., 2009; Oberauer, 2009). This difference, however, is not relevant for our purposes here. Finally, another potential interference mechanism is *inhibitory competition* (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005), a mechanism that seems to be biologically plausible (e.g., O'Reilly, 1998). The basic idea is that representation units are prepared to compete against each other by inhibitory connections; thus, new inputs displace STM content because the activation of units representing previous items declines as they accumulate inhibition from units reacting to new inputs. Again, it could be hypothesized that the number of units reacting to interfering stimuli is proportional to the amount of information contained in these stimuli; thus, more information results in greater levels of inhibition over previous information.

Regardless of what the interference mechanism might be, it is important to consider that a plausible function of rehearsal (and probably other maintaining mechanisms) may be to reconstruct the original representation of an item, reverting the interference generated by subsequent stimuli (Lewandowsky et al., 2009; Oberauer, 2009). When the representation of an item is reconstructed, this reconstruction presumably generates interference over other traces in STM by the same mechanism that operates when new items are presented. An equivalent situation could occur at the output, given that reconstruction most likely precedes item reporting. A complete interference-based account for the word-length effect, therefore, should take into consideration not only the retroactive interference generated at the input but also the interference that occurs every time a trace is reconstructed.

A final consideration concerns the well-known interaction between word-length effect and articulatory suppression. As mentioned before, the finding that the length effect disappears under articulatory suppression has been interpreted as supporting the rehearsal-based account. From this view, the word-length effect is a consequence of the different rate in which short and long words are reactivated by rehearsal, so that the effect vanishes when rehearsal is prevented by articulatory suppression. A different explanation is required, however, if we assume that rehearsal does not play the main role in the word-length effect. In this regard, Romani et al. (2005) suggested that suppression removes the length effect because participants in articulatory suppression conditions rely to a greater extent on lexical-semantic representations and trace reconstruction mechanisms. These mechanisms are more effective for long words than for short words, so that they compensate or even reverse the otherwise poorer recall of long words. This would explain why articulatory suppression does not remove the word-length effect when stimuli are non-words, which have no pre-existing lexical-semantic representations in long-term memory (Romani et al., 2005). This interpretation would also explain why suppression did not eliminate the length effect in Experiment 1, in which all the to-be-remembered words were same length and, therefore, there were not differences between conditions in the efficiency of the reconstruction mechanisms.

To summarize, the word-length effect seems to be a complex phenomenon that could be determined by a range of factors. The present study shows for the first time that the effect could be to some extent a consequence of differences in the level of retroactive interference generated by the presentation of short and long words. This finding is congruent with recent studies showing the central role of interference as a forgetting mechanism in verbal STM (Berman et al., 2009; Lewandowsky et al., 2008; Oberauer, 2009).

Acknowledgments

This study was supported by the Spanish Ministry of Science and Innovation within the project PSI2009-07374. I thank Stephan Lewandowsky and Gerald Tehan for their helpful comments on earlier versions of this article.

References

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Oxford University Press.
Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford, England: Oxford University Press.
Baddeley, A. D., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Quarterly Journal of Experimental Psychology*, *55A*, 353–369.
Baddeley, A. D., & Hitch, G. J. (1974). In G. H. Bower (Ed.), *Working memory. The psychology of learning and motivation: Advances in research and theory*, *8*. (pp. 47–90) New York: Academic.
Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, *36A*, 233–252.
Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589.
Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology. General*, *133*, 83–100.
Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*, 570–585.
Barrouillet, P., & Camos, V. (2008). The time-based resource-sharing model of working memory. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 59–80). Oxford: Oxford University Press.
Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*, 317–333.
Bireta, T. J., Neath, I., & Surprenant, A. M. (2006). The syllable-based word length effect and stimulus set specificity. *Psychonomic Bulletin and Review*, *13*, 434–438.
Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language*, *34*, 594–621.
Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551–581.
Campoy, G. (2008). The effect of word length in short-term memory: Is rehearsal necessary? *Quarterly Journal of Experimental Psychology*, *61*, 724–734.
Campoy, G., & Baddeley, A. D. (2008). Phonological and semantic strategies in immediate serial recall. *Memory*, *16*, 329–340.
Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *Quarterly Journal of Experimental Psychology*, *45A*, 177–192.
Colle, H. A., & Welsh, A. (1976). Acoustic making in primary memory. *Journal of Verbal Learning and Verbal Behavior*, *15*, 17–32.
Coltheart, V., & Langdon, R. (1998). Recall of short word lists presented visually at fast rates: Effects of phonological similarity and word length. *Memory and Cognition*, *26*, 330–342.
Coltheart, V., Mondy, S., Dux, P. E., & Stephenson, L. (2004). Effects of orthographic and phonological word length on memory for lists shown RSVP and STM rates. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*, 815–826.
Cowan, N., & AuBuchon, A. M. (2008). Short-term memory loss over time without retroactive stimulus interference. *Psychonomic Bulletin and Review*, *15*, 230–235.
Cowan, N., Baddeley, A. D., Elliott, E. M., & Norris, J. (2003). List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin and Review*, *10*, 74–79.

Cowan, N., Day, L., Saults, J. S., & Keller, T. A. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, *31*, 1–17.
Cowan, N., Wood, N. L., & Borne, D. N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science*, *5*, 103–106.
Cowan, N., Wood, N. L., Nugent, L. D., & Treisman, M. (1997). There are two word-length effects in verbal short-term memory: Opposing effects of duration and complexity. *Psychological Science*, *8*, 290–295.
Dallett, K. M. (1965). 'Primary memory': The effect of redundancy upon digit repetition. *Psychonomic Science*, *3*, 237–238.
Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, *112*, 3–42.
Dosher, B. A., & Ma, J. J. (1998). Output loss or rehearsal loop? Output-time versus pronunciation-time limits in immediate recall for forgetting-matched materials. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *24*, 316–335.
Farrell, S., & Lewandowsky, S. (2002). An endogenous model of ordering in serial recall. *Psychonomic Bulletin and Review*, *9*, 59–79.
Hendry, L. A., & Tehan, G. (2005). An item and order analysis of word length and generation effects in the short-term domain. *Memory*, *13*, 364–371.
Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory and Cognition*, *35*, 1675–1684.
Hulme, C., Neath, I., Stuart, G., Shostak, L., Surprenant, A. M., & Brown, G. D. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*, 586–594.
Hulme, C., Surprenant, A. M., Bireta, T. J., Stuart, G., & Neath, I. (2004). Abolishing the word-length effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*, 98–106.
Lange, E., & Oberauer, K. (2005). Overwriting of phonemic features in serial recall. *Memory*, *13*, 333–339.
Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin and Review*, *11*, 771–790.
Lewandowsky, S., Geiger, S. M., & Oberauer, K. (2008). Interference-based forgetting in verbal short-term memory. *Journal of Memory and Language*, *59*, 200–222.
Lewandowsky, S., & Oberauer, K. (2008). The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin and Review*, *15*, 875–888.
Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*, 1545–1551.
Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, *13*, 120–126.
Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word length effect and disyllabic words. *Quarterly Journal of Experimental Psychology*, *53A*, 1–22.
Lovatt, P., Avons, S. E., & Masterson, J. (2002). Output decay in immediate serial recall: Speech time revisited. *Journal of Memory and Language*, *46*, 227–243.
Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*, 465–501.
Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105–125.
Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.
Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81.
Neath, I., Bireta, T. J., & Surprenant, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin and Review*, *10*, 430–434.
Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin and Review*, *2*, 429–441.
O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in Cognitive Sciences*, *2*, 455–462.
Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory and Cognition*, *37*, 346–357.
Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*, 601–626.
Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, *3*, 544–576.
Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex*, *43*, 135–145.
Romani, C., McAlpine, S., Olson, A., Tsouknida, E., & Martin, R. (2005). Length, lexicality, and articulatory suppression in immediate recall: Evidence against the articulatory loop. *Journal of Memory and Language*, *52*, 406–423.
Russo, R., & Grammatopoulou, N. (2003). Word length and articulatory suppression affect short-term and long-term recall tasks. *Memory and Cognition*, *31*, 728–737.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
Sebastián, N., Martí, M. A., Carreiras, M., & Cuetos, F. (2000). *LEXESP. Léxico informatizado del español*. Barcelona: Ediciones de la Universidad de Barcelona.
Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology*, *51A*, 283–304.
Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.
Tehan, G., & Tolan, G. A. (2007). Word length effect in long-term memory. *Journal of Memory and Language*, *56*, 35–48.