Memory & Cognition
https://doi.org/10.3758/s13421-019-00898-5

Verbal working memory and linguistic long-term memory: Exploring the lexical cohort effect

Benjamin Kowialiewski[1,2] · Steve Majerus[1,2]

Abstract
Numerous studies have shown that verbal working memory (vWM) performance is strongly influenced by linguistic knowledge, with items more familiar at sublexical, lexical, and/or semantic levels leading to higher vWM recall performance. Among the many different psycholinguistic variables whose impact on vWM has been studied, the lexical cohort effect is one of the few effects that has not yet been explored. The lexical cohort effect reflects the fact that words sharing their first phonemes with many other words (e.g. *al*cove, *al*ligator, *al*cohol…) are typically responded to more slowly as compared to words sharing their first phonemes with a smaller number of words. In a pilot experiment (Experiment 1), we manipulated the lexical cohort effect in an immediate serial recall task and found no effect. Experiment 2 showed that, in a lexical decision task, participants responded more quickly to items stemming from small cohorts, showing that the material used in Experiment 1 allowed for a valid manipulation of the cohort effect. Experiment 3, using stimuli from Experiment 2 associated with maximal cohort effects during lexical decision, failed again to reveal a cohort effect in an immediate serial recall task. We argue that linguistic knowledge impacts vWM performance via continuous interactive activation within the linguistic system, which is not the case for the lexical cohort variable that may influence language processing only at the initial stages of stimulus activation.

Keywords Working memory · Cohort competition · Linguistic knowledge

Introduction

Language-based models of verbal working memory (vWM) assume that temporary storage of verbal information relies on direct activation of corresponding representations within the linguistic system (Acheson & MacDonald, 2009; Gupta, 2009; Majerus, 2013; N. Martin, Saffran, & Dell, 1996; R. C. Martin, Lesch, & Bartha, 1999). This is supported by the fact that many psycholinguistic variables affecting language processing also affect vWM (Brener, 1940; Guérard & Saint-Aubin, 2012; Hulme et al., 1997; Kowialiewski & Majerus, 2018b; Majerus, Van der Linden, Mulder, Meulemans, & Peters, 2004; Poirier & Saint-Aubin, 1995, 1996; Romani, McAlpine, & Martin, 2008; Watkins & Watkins, 1977). A specific psycholinguistic variable has, however, never been investigated in vWM: the lexical cohort competition effect. This effect is characterized by the fact that words drawn from large lexical cohorts (e.g. *al*cove, *al*ligator, *al*cohol…) are usually responded to more slowly than words drawn from small cohorts in lexical decision tasks as a result of many lexical competitors getting co-activated for words from large lexical cohorts (Marslen-Wilson, 1987). The purpose of the present study was to investigate whether the lexical cohort variable can also impact vWM, as predicted by language-based models of vWM.

vWM closely interacts with phonological, lexical, and semantic linguistic variables. At the sublexical/phonological level, this is illustrated by studies showing that nonwords containing structures of high phonotactic probability (i.e., high biphone frequencies) are associated with higher vWM performance than nonwords containing structures of low phonotactic probabilities (Gathercole, Frankish, Pickering, & Peaker, 1999; Majerus et al., 2004). Likewise, the lexical levels of representation have also been shown to impact vWM performance, with higher recall performance for words than nonwords (Brener, 1940; Hulme, Maughan, & Brown, 1991; Jefferies, Frankish, & Lambon

✉ Benjamin Kowialiewski
bkowialiewski@uliege.be

[1] Psychology and Neuroscience of Cognition Research Unit (PsyNCog), University of Liège, Place des Orateurs 1, 4000 Liège, Belgium

[2] Fund for Scientific Research – F.R.S.-FNRS, Brussels, Belgium

Published online: 18 January 2019

Ralph, 2006a), and high-frequency words also leading to higher recall performance as compared to low frequency words (Hulme et al., 1997; Hulme, Stuart, Brown, & Morin, 2003; Poirier & Saint-Aubin, 1996; Watkins & Watkins, 1977). Furthermore, lists composed of words having many versus few lexical neighbors also lead to differential recall performance in vWM tasks (Clarkson, Roodenrys, Miller, & Hulme, 2016; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002; Vitevitch, Chan, & Roodenrys, 2012). It is important to distinguish here the lexical neighborhood and the lexical cohort effects: while the lexical cohort effect characterizes words sharing the same onset, the lexical neighborhood effect characterizes words differing from each other by a single phoneme substitution, deletion, or addition independently of phoneme position. Finally, semantic variables also affect vWM performance, with higher recall performance for lists composed of high versus low imageability words, and for semantically related versus unrelated words lists (Campoy, Castellà, Provencio, Hitch, & Baddeley, 2015; Poirier & Saint-Aubin, 1996; Romani et al., 2008).

These psycholinguistic effects in vWM tasks can be explained by language-based models of vWM processing, assuming fast and direct interactions between vWM and the linguistic system (Acheson & MacDonald, 2009; Gupta, 2009; Majerus, 2013; N. Martin et al., 1996; R. C. Martin et al., 1999), where to-be-remembered items are activated within the linguistic system as soon as they are presented in a vWM task. Interactive activation models of language processing are particularly suited for explaining these results (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; McClelland & Rumelhart, 1981). Contrary to purely feedforward activation models, in interactive activation models, each level of language processing (phonological, lexical, and semantic) is allowed to directly activate adjacent levels via bi-directional connexion weights. In the case of single-word repetition, initial activation at the phonological level spreads toward the lexical level. At the same time, activation at the lexical level spreads to the semantic level and reactivates the phonological level. Finally, the semantic layer reactivates the lexical level. These interactions are supposed to occur iteratively over the time-course of a language-processing task. The impact of lexical and semantic knowledge on vWM can be explained in the same manner: verbal memoranda associated with richer or more stable lexico-semantic representations will receive stronger feedback activations from adjacent layers, and hence will be less prone to decay over time. This approach has been modelled in a computational model of single-word repetition (Martin, Dell, Saffran, & Schwartz, 1994; N. Martin et al., 1996) by extending Dell's spreading activation model of picture naming. Although this computational model was constructed to explain single-word repetition performance, a conceptual attempt has been made to extend it to whole list repetition (N. Martin et al., 1996). This conceptual approach is also consistent with attention-based models assuming direct interactions between the attentional and long-term memory systems (Barrouillet & Camos, 2007; Cowan, 2001; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012), and with other language-based models assuming strong interactions between vWM and language activation (Acheson & MacDonald, 2009; Gupta, 2009).

As we have shown, many psycholinguistic variables have been assessed with regard to their impact on vWM performance. However, one specific effect has not yet been investigated: the effect of lexical cohort competition. As already noted, this effect is related to the number of lexical competitors sharing their first phonemes with a target stimulus (Marslen-Wilson, 1987; Tyler, Voice, & Moss, 2000): "alligator" shares the onset syllable /æl/ with the words "alcohol" or "alcove." Words sharing their first phonemes with many other words (i.e., from large cohorts) are usually associated with slower response times in lexical decision tasks as compared to words drawn from small cohorts (Gaskell & Marslen-Wilson, 2002; Kocagoncu, Clarke, Devereux, & Tyler, 2017; Marslen-Wilson, 1987; Tyler et al., 2000; Zhuang, Randall, Stamatakis, Marslen-Wilson, & Tyler, 2011). Words from a given cohort are co-activated and compete for selection when a given speech input is analysed, leading to larger competition effects for words stemming from large cohorts. Furthermore, these lexical competition effects interact with semantic access: in lexical decision tasks, concreteness/imageability effects are most pronounced for words stemming from larger cohorts (Tyler, Moss, Galpin, & Voice, 2002; Zhuang et al., 2011). This situation has been explained by considering that, when direct and fast mapping between phonological and lexical levels is difficult, semantic levels of processing intervene by providing additional information that will help to disambiguate the target stimulus (Evans, Lambon Ralph, & Woollams, 2012).

With regard to vWM, as for language-processing tasks, language-based models of vWM predict that words drawn from larger cohorts will be more difficult to activate during the encoding stage due to their larger ambiguity, causing item interference effects. This should result in poorer recall performance for words drawn from large versus small cohorts, via either an increased number of omissions or an increased number of intrusions, or both. At the same time, it should be noted that the lexical cohort effect involves the early stages of lexical access and lexical selection. Once a lexical representation has been activated, the lexical cohort variable is no longer considered to exert any effect. This strongly contrasts with other psycholinguistic effects such as lexicality and word imageability effects, which are considered to have a more continuous impact on vWM maintenance as the underlying lexical and semantic variables provide stabilizing feedback all over the vWM maintenance phase – at least according to interactive activation models of language processing. Therefore, we could also expect a reduced or even an absent effect of the lexical cohort variable in vWM performance. In this study, we explored, via two experiments (Experiments 1 and 3), the impact of cohort competition on a word-list immediate serial recall task, with all words (for a given list) stemming from large or small lexical cohorts. Experiment 2 was a control experiment checking the validity of the cohort competition manipulations using a lexical decision task.

Experiment 1

Method

Participants A total of 16 (12 females, four males) participants aged between 18 and 33 years (M = 22.27, SD = 3.83) were recruited from the university community after giving their informed consent. All participants were native French speakers and reported no history of neurological disorders or learning difficulties. The study had received approval from the local ethics committee.

Materials The list of stimuli consisted of 210 words associated with large cohorts and 210 words associated with small cohorts. The cohort competition variable was computed using the following procedure. First, we selected all existing French words, including nouns, verbs, and adjectives, using "Lexique381" from the Lexique 3.0 database (Lexique 3; New, 2006). From this pool, only the canonical (lemma) form of the stimuli was retained. Hence, different words sharing the same lemma (e.g., *sister – sisters*) were considered as the same word in the cohort. When several words shared a common phonological form (i.e., homonym), only the most frequent word form was considered, since its lexical form is supposed to win the competition over the other, less frequent lexical forms. From this final pool, the number of words sharing their first phonemes with a target stimulus was computed, and this for each individual word and for increasing numbers of onset phonemes until a given word's phonological uniqueness point was reached (i.e., when the word could be identified in an unambiguous manner) (Marslen-Wilson, 1987; Zhuang, Tyler, Randall, Stamatakis, & Marslen-Wilson, 2014). We derived a first measure quantifying the number of competitors in the initial cohort, which we refer to here as the *number of competitors*, as a function of the number of shared-onset phonemes. We also computed a *cohort competition* variable, which for a given target word, corresponds to its lexical frequency, divided by the summed frequency of all its competitors. For instance, given the cohort composed of "cat," "cab," and "car," with lexical frequencies of 5, 3, and 2, respectively, the cohort competition value for "cat" will be equal to $\frac{5}{(3+2)} = 1$. Likewise, the cohort competition value for "cab" will be equal to $\frac{3}{(5+2)} = .43$. As can be seen in this example, *smaller* values represent *higher* competition because the target has less weight (in terms of lexical frequency) in the cohort. Both measures of competition have been shown to impact lexical access in linguistic tasks, but cohort competition seems to be the most reliable variable (Tyler et al., 2000; Zhuang et al., 2011), since it also takes into account the weight of each competitor within the cohort. As expected, the two measures were highly correlated for our pool of stimuli ($r = -.84$, $r^2 = .71$ after log transformation, $BF_{10} > 100$; see *Statistical analysis* below for the interpretation of the Bayes factor). Critically, we also controlled for lexical neighborhood density by measuring the number of real words in the Lexique381 pool that could be created by adding, deleting, or substituting one phoneme in the target word. To do that, we used the Levenshtein distance (Levenshtein, 1966), which computes the minimal distance between two character arrays. These changes include deletion, suppression, and substitution. The Levenshtein distance was computed between the target word and all other words of the Lexique381 database (after the word-selection process detailed above), allowing us to compute the number of lexical neighbors: a word associated with a Levenshtein distance of 1 was considered to be a neighbor. From this pool, the words containing five or six phonemes and having a high lexical frequency ($freq_{log} > -1.52$) were selected and then divided in small and large cohort stimuli using a median split. This set of stimuli was then further reduced by selecting by hand only the small and large cohort words that were matched according to the different psycholinguistic variables mentioned below.

The two lists of stimuli differed according to the number of competitors ($M_{log} = 2.65$, $SD_{log} = .26$ and $M_{log} = 1.71$, $SD_{log} = .45$ for high- and low-competition words, respectively, $BF_{10} > 100$) and cohort competition values ($M_{log} = -.75$, $SD_{log} = .31$ and $M_{log} = .76$, $SD_{log} = .57$ for high- and low-competition words, respectively, $BF_{10} > 100$) by considering the first two phonemes. Cohort competition values also differed between the two lists when considering the four initial phonemes. The two sets were matched for several other psycholinguistic variables: biphone frequency (M = 887.64, SD = 293.19 and M = 883.81, SD = 347.49 for high- and low-competition words, respectively, $BF_{01} = 9.18$, Tubach & Boë, 1990), lexical frequency ($M_{log} = .84$, $SD_{log} = .27$ and $M_{log} = .84$, $SD_{log} = .53$ for high- and low-competition words, respectively, $BF_{01} = 9.24$ taken from the "freqlemfilm2" variable in the Lexique database) and number of phonemes (M = 5.47, SD = .50 and M = 5.47, SD = .50 for high- and low-competition words, respectively, $BF_{01} = 9.25$). The two set of stimuli did not differ according to the lexical neighborhood density variable (M = 3.68, SD = 3.02 and M = 3.03, SD = 2.75 for high- and low-competition words, respectively, $BF_{10} = 1.38$). Finally, since imageability ratings are available for only a restricted set of stimuli in French, we conducted an online survey in which we invited the participants to judge the degree of imageability of our stimuli on a scale ranging from 1 to 7. Because of this very large number of stimuli to judge, the participants were free to stop the survey at any moment. Sixty-seven participants took part in the survey, and each stimulus was judged 16.15 times on average. The two sets of stimuli were equivalent in terms of imageability ratings (M = 4.60, SD = 1.42 and M = 4.64, SD = 1.42 for high- and low-competition words, respectively, BF_{01} = 8.93). All the linguistic properties for this set of stimuli are summarized in Table 1. Note that an additional analysis in which homonyms were included for computing the number of competitors and cohort competition values led to a similar pattern of results, with low- and high-cohort stimuli still reliably differing on these values.

The items were recorded by a French-native female speaker in a neutral voice. Each item was then isolated in a separate file, the length of which corresponded to its acoustic duration. A Bayesian independent samples t-test showed that high-and low-competition stimuli did not differ according to their length (M = 751 ms, SD = 100 ms and M = 751 ms, SD = 88 ms for high- and low-competition words, respectively, BF_{01} = 9.25). We removed the residual background noise via Audacity®, which uses a Fourier analysis (see http://wiki.audacityteam.org/wiki/How_Audacity_Noise_Reduction_Works for more information).

Procedure Each participant received a different version of the vWM task. We manipulated the cohort competition effect by presenting six-word lists composed of words drawn from either high- or low-cohort competition, such that a given list was composed of words drawn exclusively from one stimulus condition. The lists were pseudorandomly presented with the constraint that a given condition could not appear on more than three consecutive trials. In order to avoid phonological overlap, two adjacent words could not share the same first two or last two phonemes within each list, because phonological similarity has been shown to strongly influence vWM processing (Baddeley, 1966), both at the item and the serial order levels of processing (Gupta, Lipinski, & Aktunc, 2005). In addition, for each trial, we computed Latent Semantic Analysis (LSA) values (http://lsa.colorado.edu/, using the semantic space "Francais-Monde-Extended"). LSA measures reflect the extent to which two words co-occur in the same linguistic corpora. The higher the co-occurrence of two words, the higher their (theoretical) semantic association values. This lexical variable is important to control for because it has been shown to impact vWM performance and has previously been shown to drive, at least partially, the lexical frequency effect (Hulme et al., 1997, 2003; but see Poirier & Saint-Aubin, 2005; Tse & Altarriba, 2007). We computed LSA values for adjacent items within a given word list (Saint-Aubin, Guérard, Chamberland, & Malenfant, 2014). For each adjacent pair of stimuli, the LSA values were then averaged for each trial. We observed that the two stimulus conditions did not differ (M = .05, SD = .04 and M = .06 and SD = .04 for high- and low-cohort trials, respectively), and this was supported by strong evidence (BF_{01} = 12.21). There were 35 experimental trials in each stimulus condition. Participants could take a short break after 35 trials if they needed to. The whole experiment took approximately 35 min to be performed.

Participants performed three unrecorded practice trials before the beginning of the main vWM task. At the beginning of each trial, an on-screen countdown display starting from 3 was first presented, followed by a blank screen and the auditory items presented at a rate of one item every 1,200 ms. Each list was directly followed by a sinusoidal tone of 440 Hz lasting for 150 ms, signaling the start of the

Table 1 Values for linguistic matching variables between high- and low-cohort word stimuli used in Experiments 1 and 2

Linguistic variables	Cohort competition		BF
	High	Low	
Number of competitors (M_{log})	2.65 (.26)	1.71 (.45)	BF_{10} = 5.32e+86
Cohort competition (M_{log})	-.75 (.31)	.76 (.57)	BF_{10} = 1.99e+118
Biphone frequency	887.64 (293.19)	883.81 (347.49)	BF_{01} = 9.18
Lexical frequency (M_{log})	.84 (.27)	.84 (.53)	BF_{01} = 9.24
Number of phonemes	5.47 (.5)	5.47 (.5)	BF_{01} = 9.25
Neighborhood density	3.68 (3.02)	3.03 (2.75)	BF_{10} = 1.38
Imageability	4.60 (1.42)	4.64 (1.42)	BF_{01} = 8.93
Acoustic length	751 (100)	751 (88)	BF_{01} = 9.25
LSA values	.05 (.04)	.06 (.04)	BF_{01} = 12.21

Note. Log transformation of mean values is signalled by "(M_{log})". Values in parenthesis represent standard deviations

Bayesian factor (BF) values are based on Bayesian independent samples t-tests recall phase. After participants had recalled the items, they were invited to initiate the next trial using the SPACEBAR of the keyboard. Participants were told that they had to recall aloud any item they could remember and in the serial order in which the items had been presented. In order to ensure accurate scoring of serial recall performance, participants were asked to use a sheet when recalling each item. The sheet was placed directly in front of them on the desk in landscape orientation, and was composed of six squares placed along the horizontal axis (see Appendix A). The participants were invited to move their finger to the right by one square each time they recalled an item. Pilot tests had shown that participants often failed to recall all six items because they struggled to count how many items they had already recalled. The pointing procedure helped participants to keep track of the number of recalled items and allowed the experimenters to accurately score serial recall performance. When participants could not remember a given item in the list, they were invited to say the word "blanc" (i.e., "blank" in French). Task presentation was controlled via OpenSesame software running on a desktop station computer (Mathôt, Schreij, & Theeuwes, 2012). The auditory stimuli were presented via headphones directly connected to the computer. The loudness was adjusted to comfortable listening levels for each participant during the practice trials. The participants' responses were recorded with a digital recorder and stored on computer disk for later transcription and scoring.

With regard to the scoring procedure, we performed different analyses. First, we used an item-recall scoring procedure in which an item was scored as correct if it was recalled regardless of its recall position. For instance, given the target sequence "Item1 – Item2 – Item3 – Item4 – Item5 – Item6" and the output sequence "Item1 – Item2 – Item4 – Item3 – blank – Item6," items 1, 2, 3, 4, and 6 were scored as correct. This scoring procedure is particularly sensitive to item recall. In addition, we also performed a strict scoring procedure in which an item was scored as correct only if it was recalled at the correct serial position. Using this scoring procedure, only items 1, 2, and 6 would be scored as correct in the previous example.

Statistical analysis We performed a Bayesian analysis instead of the traditional frequentist analyses in order to substantially reduce Type-1 false error probabilities (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). The Bayesian approach has the advantage of computing continuous values against or in favor of a given model, rather than deciding for the presence of an effect based on an arbitrary statistical threshold. Evidence in favor of a model is given by the Bayesian factor (BF), reflecting the likelihood ratio of a given model relative to other models, including the null model. Both the null model and the effect of interest can be simultaneously tested, by directly comparing the alternative hypothesis against the null hypothesis, and vice versa. The BF_{10} is used to determine the likelihood ratio for the alternative model (H_1) relative to the null model (H_0) and the BF_{01} to determine the likelihood ratio for H_0 relative to H_1. We use the classification of strength of evidence proposed in previous studies (Jeffreys, 1998; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011): A BF of 1 provides no evidence, $1 < BF < 3$ provides anecdotal evidence, $3 < BF < 10$ provides moderate evidence, $10 < BF < 30$ provides strong evidence, $30 < BF < 100$ provides very strong evidence, and $100 > BF$ provides extreme/decisive evidence. We also report $BF_{inclusion}$ values that compare the evidence of all models including a given factor relative to all other models not including this factor. All the analyses were performed using JASP (JASP Team, 2017) and we used default Cauchy prior distribution parameters as implemented in JASP (Version 0.8.5.1).

Results and discussion

We first assessed the effect of cohort competition (high, low) as a function of serial position (1–6) using a Bayesian repeated measures ANOVA. For the item-recall measure, we found moderate evidence *against* the effect of cohort competition ($BF_{01} = 7.93$), decisive evidence supporting the serial position effect ($BF_{inclusion} \rightarrow +\infty$), and strong evidence against the interaction term ($BF_{01} = 24$). Similar results were observed when using a strict recall criterion, with moderate evidence against the cohort competition effect ($BF_{01} = 9.09$), decisive evidence in favor of the serial position effect ($BF_{inclusion} = 3.002e+15$), and very strong evidence against the interaction term ($BF_{01} = 30.30$). These results are shown in Fig. 1. Hence, we observed no impact of cohort competition on recall accuracy, and this absence of difference was reliably supported, as shown by the BF_{01}.

This pilot experiment provides evidence for the *absence* of a cohort competition effect on vWM recall performance. At the same time, we cannot rule out the possibility that this absence reflects insufficient contrasted lexical cohorts, or that this effect does not characterize the French language (although the latter possibility is rather unlikely, English and French sharing many lexical properties; but see Sadat, Martin, Costa, & Alario, 2014). So far, no study has investigated the cohort competition effect in French, and therefore we need to check whether our stimulus material is appropriate for eliciting this effect in French by using a lexical decision task that has been most frequently used to investigate this effect in other languages (mostly English).

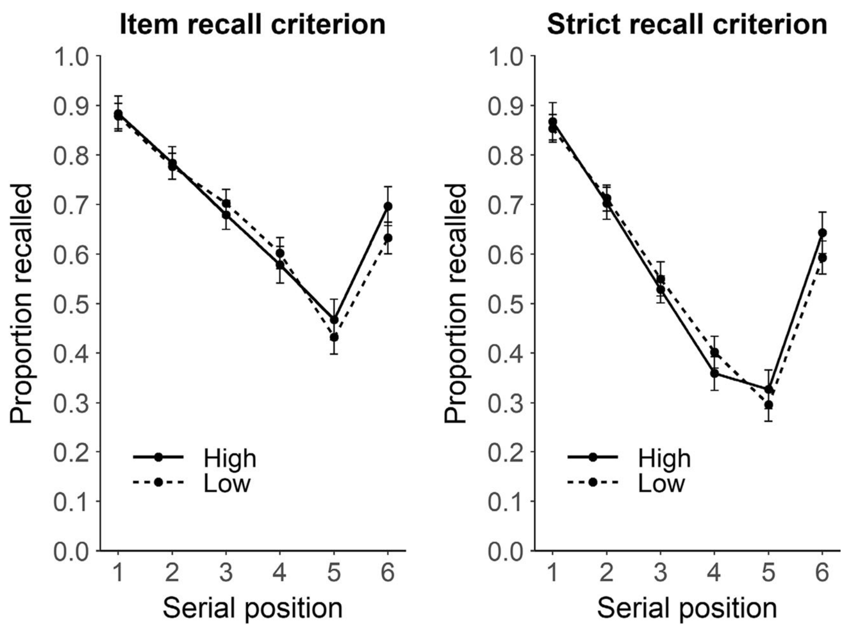


Fig. 1 Experiment 1 – Proportion of items correctly recalled (y-axis) across serial positions (x-axis) as a function of the lexical cohort variable, for item recall- (left panel) and strict recall- (right panel) scoring procedures. The solid and dashed lines represent high- and low-cohort stimuli, respectively. Error bars represent standard errors, after correction for between-subject variability (Cousineau, 2005)

Experiment 2: Cohort competition in lexical decision

Experiment 2 assessed the effect of cohort competition in a linguistic, lexical decision task where participants were invited to judge the lexical status of word and nonword stimuli. The primary goal of Experiment 2 was to assess whether the absence of the cohort competition effect observed in Experiment 1 was due to the specific set of stimuli we had created. Second, examining the occurrence of a cohort competition effect in French is important to demonstrate its generalizability across languages. Third, since we used word and nonword stimuli for lexicality judgement, the impact of cohort competition was factorially manipulated and hence was also assessed on nonword stimuli. Even though one study investigated effects of cohort competition on nonwords using correlational methods (Zhuang et al., 2014), cohort competition in nonwords has never been directly manipulated, and evidence supporting its existence is scarce.

Method

Participants A total of 29 (28 females, one male) participants aged between 19 and 25 years (M = 21.14, SD = 1.43) were recruited from the university community after providing informed consent. All participants were native French speakers and reported no history of neurological disorders or learning difficulties. The study had received approval from the local ethics committee.

Materials The same set of words as in Experiment 1 was used. We additionally created nonword stimuli for the lexical decision task used in Experiment 2. We first constructed a large (N > 10e+5) set of nonwords using an algorithm programmed under MATLAB® (the script and the modified Lexique381 pool was made available in the Open science framework using the following link: https://osf.io/3rkh5/) with the constraint that a given item could not match any entry within Lexique381. In addition, all the nonwords were five to six phonemes long, and were created by randomly assembling phonemes from the French language, by constraining the program to use the syllabic structures characterizing the stimuli of the word pool. Two sets of 210 nonwords were selected, with the constraint that they had to strongly differ in their number of lexical competitors. As for the word stimuli, this was made by computing the number of words in the Lexique381 database (after discarding homophones and non-lemma forms; see Experiment 1 for the details of the cleaning process) sharing their onset with the target nonword. For instance, given the nonword "caz," the words "cat," "cab," and "car" were considered as competitors. As for the word stimuli, we also computed the summed frequency of all the competitor words as an equivalent to the cohort competition variable (Zhuang et al., 2014).

The high- and low-competition nonwords differed according to their number of competitors (M_{log} = 2.61, SD_{log} = .34 and M_{log} = 1.462, SD_{log} = .43 for high- and low-competition nonwords, respectively, BF_{10} > 100) and the variable of competitors summed frequency (M_{log} = 3.29, SD_{log} = .55 and M_{log} = 1.74, SD_{log} = .95 for high- and low-competition nonwords, respectively, BF_{10} = 1.67e+61). The high- and low-competitors' nonwords were matched for biphone frequency (M = 646.76, SD = 335.29 and M = 646.206, SD = 378.22 for high- and low-competition nonwords, respectively, BF_{01} = 9.5), neighborhood density (M = .3, SD = .7 and M = .18, SD = .61 for high- and low-competition nonwords, respectively, BF_{01} = 1.75), and number of phonemes (M = 5.47, SD = .5 and M = 5.47, SD = .5 for high- and low-competition nonwords, respectively, BF_{01} = 9.25). The stimuli did not differ according to their acoustic duration (M = 688 ms, SD = 95 ms and M = 704 ms, SD = 89 ms for high- and low-competition words, respectively, BF_{01} = 2.01).

Both word and nonword stimuli were matched according to their number of phonemes (M = 5.47, SD = .5 and M = 5.47, SD = .5 for word and nonword stimuli, respectively, BF_{01} = 12.961) and syllabic structures (98.81% had a similar consonant/vowel/semivowel structure). The word and nonword stimuli could, however, not be perfectly matched for all psycholinguistic variables. This is not problematic for the purpose of the present experiment as we were not interested in directly comparing the word and nonword stimuli. More specifically, the word and nonword stimuli differed strongly in their phonotactic frequency (M = 885.73, SD = 321.11 and M = 646.48, SD = 356.972 for word and nonword stimuli, respectively, BF_{10} = 9.14e+19), acoustic duration (M = 751.038 ms, SD = 94.06 ms and M = 696.215 ms, SD = 92.423 ms for word and nonword stimuli, respectively, BF_{10} = 6.32e+13), neighborhood density (M = 3.58, SD = 2.9 and M = .24, SD = .66 for word and nonword stimuli, respectively, BF_{10} = 1.64e+78), and, to a lesser extent, their number of lexical competitors (M = 2.18, SD = .6 and M = 2.04, SD = .7 for word and nonword stimuli, respectively, BF_{10} = 8.38). All these values are summarized in Table 2.

All the nonword items were recorded by a French-native female speaker in a neutral voice, using the same voice as for the words. Each item was then isolated in a separate file, the length of which reliably corresponded to its acoustic duration. We removed the background noise via Audacity®. The word and nonword stimuli were recorded at different moments. In order to ensure that word and nonword stimuli did not differ at the level of general acoustic parameters, we checked fundamental frequency and intensity values. An analysis of the fundamental frequency (F0) using the "freqz" function implemented under MATLAB® showed evidence for a very small difference between the word and nonword stimuli, with significant overlap of values (M = 446.5, SD = 262.06 and M = 500.3, SD = 326.3 for word and nonword stimuli, respectively, BF_{10} = 2.31). Intensity values were associated with positive evidence for an absence of difference (M = 208.8, SD = 76.31 and M = 200.9, SD = 90.62 for word and nonword stimuli, respectively, BF_{01} = 5.11).

Procedure For each participant, stimuli were presented in pseudorandom order, with the constraint that a given stimulus condition (word/nonword, high/low cohort) could not be repeated on more than three consecutive trials. There were 210 trials for each of the four stimulus conditions, and it took approximatively 45 min to perform the whole experiment. Participants were allowed to take a maximum of three short breaks if they needed to. Participants performed 14 practice trials before administration of the main task. If participants made a mistake during the practice trials, they received corrective feedback on their performance.

Each trial began with an on-screen fixation cross lasting on average 1,000 ms, plus/minus a random duration sampled from a continuous uniform distribution ranging from 0 to 250 ms. The fixation cross was directly followed by a blank screen and the target item. Participants were told that they had to judge the lexical status of the item (i.e., "You will have to judge whether the presented item is a word or a nonword"), and had to press the "S" (for word) or "L" (for nonword) key on the keyboard to indicate their response. The next trial directly began after each keypress. Participants were told that they had to respond as quickly as possible, without sacrificing

Table 2 Values for linguistic matching variables between high- and low-cohort nonword stimuli used in Experiment 2

Linguistic variables	Cohort competition		BF
	High	Low	
Number of competitors (M_{log})	2.61 (.34)	1.462 (.43)	BF_{10} = 3.31e+103
Competitors summed freq. (M_{log})	3.29 (.55)	1.74 (.95)	BF_{10} = 1.67e+61
Biphone frequency	646.76 (335.29)	646.206 (378.22)	BF_{01} = 9.5
Number of phonemes	5.47 (.5)	5.47 (.5)	BF_{01} = 9.25
Neighborhood density	.3 (.7)	.18 (.61)	BF_{01} = 1.75
Acoustic length	688 (95)	704 (89)	BF_{01} = 2.01

Note. Log transformation of mean values is signalled by "(M_{log})". Values in parenthesis represent standard deviations

Bayesian factor (BF) values are based on Bayesian independent samples t-tests accuracy. In order to stress rapidity, throughout the entire experiment, an on-screen message instructed participants to respond more quickly when they failed to respond within 2,500 ms after a stimulus' onset (scored as a no-response). Stimulus presentation was controlled via OpenSesame software running on a desktop station computer (Mathôt et al., 2012). The auditory stimuli were presented via headphones directly connected to the computer. The loudness was adjusted to comfortable listening levels during the practice trials. The experiment was separated into four blocks, allowing participants to take a very short break between blocks if they needed to. Both response accuracy and time were recorded. Hits and false alarms were combined via d' prime scores (Stanislaw & Todorov, 1999).

Results and discussion

Participants were on average very accurate, both for words (M = .95, SD = .04) and for nonwords (M = .97, SD = .03). A Bayesian repeated measures ANOVA on d' scores with the factors lexicality (word, nonword) and cohort competition (high, low) showed that discrimination scores did not differ as a function of lexicality (BF_{01} = 5.71) or cohort competition (BF_{01} = 2.14), and strong evidence supported the absence of an interaction (BF_{01} = 15.63).

After removing incorrect trials, response times of less than or greater than 2.5 absolute deviations from the median on an individual basis (Leys, Ley, Klein, Bernard, & Licata, 2013) were discarded from the analysis, leading to discarding a total of 1,161 observations (4.77%) from the entire set of data. The vast majority (97.42%) of these extreme values were located in the upper part of the distribution, and comprised response times between min = 970 and max = 2,493. In the lower part of the distribution, these data comprised response times between min = 164 and max = 684. For each participant, average response times were computed across all four stimulus conditions (word-high competition; word-low competition; nonword-high competition; nonword-low competition). The presence of a cohort competition effect (high, low) as a function of lexicality (word, nonword) was assessed using a Bayesian repeated measures ANOVA. We found decisive evidence supporting both effects of lexicality ($BF_{inclusion}$ = 458447.174) and cohort competition ($BF_{inclusion}$ = 5342.209). The interaction term was ambiguous ($BF_{inclusion}$ = 1.108). Hence, the cohort competition effect was reliably observed across the word and nonword conditions. The same analysis using median response times instead of average response times led to similar results. These results are shown in Fig. 2. Note that a mixed-model analysis using frequentist statistics was also conducted, and led to very similar conclusions. This latter analysis is available in Appendix B.

The observation of slower response times for words from large lexical cohorts is consistent with previous psycholinguistic studies. Interestingly, the present study also revealed a cohort competition effect for nonwords. This finding suggests that lexical activation also occurs for these stimuli, possibly in the form of lexical search processes (e.g., Vitevitch & Luce, 1999). Critically, the presence of a lexical cohort effect in the lexical decision task of Experiment 2 shows that the absence of a cohort competition effect for the vWM task in Experiment 1 cannot be imputed to the specific characteristics of our stimulus material. It should be noted that cohort competition effects, when occurring, are typically very small, with a mean difference of ~15 ms for word stimuli in the present study (see also Tyler et al., 2000 and Zhuang et al., 2011 for similar values). It could therefore be argued that the impact of the lexical cohort variable may have been too subtle to influence performance in vWM paradigms. To assess this possibility, Experiment 3 used the same vWM task setup as Experiment 1, but selecting only those word items that had led to the largest cohort competition effects in Experiment 2.

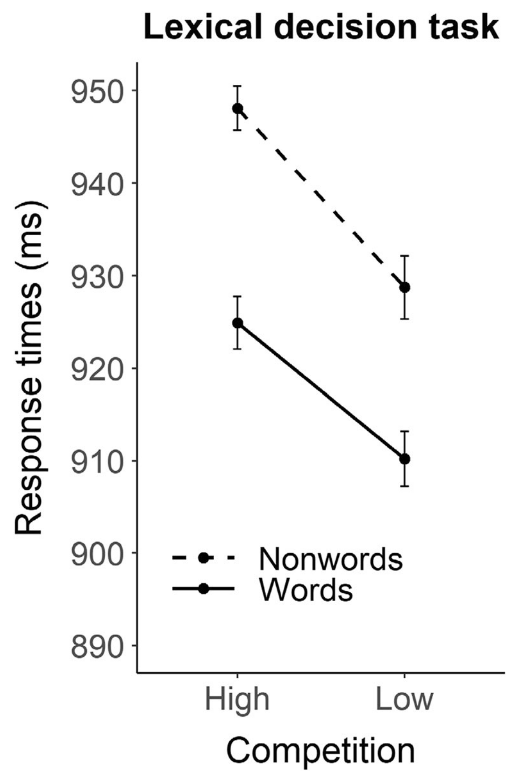


Fig. 2 Experiment 2 – Mean response times averaged across participants (y-axis) for high- and low-cohort stimuli (x-axis) separately for words (solid line) and nonwords (dashed line). Error bars represent standard errors, after correction for between-subject variability (Cousineau, 2005)

Experiment 3

This third experiment assessed the cohort competition effect in a vWM task by using only the word stimuli that had been shown to be the most responsive to the cohort manipulation variable in Experiment 2. We retained, from Experiment 2, those word stimuli associated with the slowest response times (for the large cohort category), and with the fastest response times (for the small cohort category).

Method

Participants A total of 30 (23 females, seven males) participants aged between 18 and 28 years (M = 21.37, SD = 2.30) were recruited from the university community after giving their informed consent. All participants were native French speakers and reported no history of neurological disorders or learning difficulties. The study had received approval from the local ethics committee.

Materials The word stimuli were identical to those used in Experiment 1, except that we selected the items with the largest response-time differences in the lexical decision task of Experiment 2. We were able to select 150 word items for each stimulus condition. More specifically, we first considered the median response times obtained for each word in the lexical decision task in Experiment 2. Next, we removed one by one items from the large- and small-cohort word sets so that for the remaining words the difference in terms of response times was maximized, while also ensuring that the words were still matched at the level of psycholinguistic variables between both sets. Although there was still a slight overlap between the two sets in terms of response times (M = 930.8, SD = 67.76 and M = 885.65, SD = 58.82 for high- and low-cohort stimuli, respectively, BF_{10} = 4.02e+6), the gap was now larger as compared to the initial set: ~45 ms. The two stimulus sets were matched for biphone frequency (M = 907.84, SD = 283.30 and M = 905.35, SD = 353.23 for high- and low-cohort stimuli, respectively, BF_{01} = 7.85), imageability (M = 4.76, SD = 1.34 and M = 4.76, SD = 1.45 for high- and low-cohort stimuli, respectively, BF_{01} = 7.87), lexical frequency (M_{log} = .84, SD_{log} = .25 and M_{log} = .85, SD_{log} = .42 for high- and low-cohort stimuli, respectively, BF_{01} = 7.7), number of phonemes (M = 5.47, SD = .50 and M = 5.46, SD = 50 for high- and low-cohort stimuli, respectively, BF_{01} = 7.67), neighborhood density (M = 3.71, SD = 2.97 and M = 3.2, SD = 2.75 for high- and low-cohort stimuli, respectively, BF_{01} = 2.5), and acoustic length (M = 750, SD = 97 and M = 749, SD = 83 for high- and low-cohort stimuli, respectively, BF_{01} = 7.81). The high- and low-cohort stimuli still differed at the level of number of competitors (M_{log} = 2.64, SD_{log} = .26 and M_{log} = 1.69, SD_{log} = .42 for high- and low-cohort stimuli, respectively, BF_{10} = 9.08e+66) and cohort competition values (M_{log} = -.74, SD_{log} = .29 and M_{log} = .76, SD_{log} = .56 for high- and low-cohort stimuli, respectively, BF_{10} = 5.54e+85). A summary of matching variables is provided in Table 3.

Procedure Ten different experimental lists were created, counterbalanced across subjects. As for Experiment 1, we computed LSA values for each trials. The two stimulus conditions did not differ at the level of LSA values (M = .06, SD = 0.04 and M = .05, SD = .04 for high- and low-cohort stimuli, respectively), and this absence of difference was supported by moderate Bayesian evidence (BF_{01} = 5.21). Due to a smaller number of items available for each word condition relative to the previous experiments (150 instead of 210), we decided to reduce list length by creating lists of five items (instead of six in Experiment 1). This enabled us to create 30 trials (instead of 35 in Experiment 1) per experimental conditions and to present each word only once (as in previous experiments). In addition, presentation rate was set to 1,000 ms per item to further shorten task duration. The response sheet for immediate serial recall was the same as in Experiment 1 (see Appendix A), except that it included five squares. All other aspects of the experimental procedure, including statistical analysis and scoring procedures, were identical to Experiment 1.

Results and discussion

We observed very similar results to those reported in Experiment 1. Using a Bayesian repeated measure ANOVA on item-recall performance, we found moderate evidence against the cohort competition effect (BF_{01} = 9.52), decisive evidence supporting the effect of serial position ($BF_{inclusion} \to +\infty$), and very strong evidence supporting the absence of interaction (BF_{01} = 55.56) (see Fig. 3). When conducting the same analyses using the strict recall criterion, there was again strong evidence supporting the absence of a cohort competition effect (BF_{01} = 10.87), decisive evidence supporting the serial position effect ($BF_{inclusion} \to +\infty$), and very strong evidence supporting the absence of interaction (BF_{01} = 62.5).

Experiment 3 confirms the absence of a cohort competition effect in vWM, even when using stimuli that had been shown to lead to maximal cohort effects in a lexical decision task. In addition, this absence appears to be reliable given that we included a higher number of participants as compared to Experiment 1.

General discussion

This study explored the impact of the cohort competition effect on vWM. We observed in two experiments that this linguistic variable did not influence immediate serial recall performance for lists composed of words stemming from large versus small lexical cohorts. This result cannot be attributed to a problem at the level of stimulus material as, with the same stimulus set, a reliable cohort effect was observed in a lexical decision task, the task most typically used in previous studies for studying cohort competition effects.

Table 3 Values for linguistic matching variables between high- and low-cohort word stimuli used in Experiment 3

Linguistic variables	Cohort competition		BF
	High	Low	
Number of competitors (M_{log})	2.64 (.26)	1.69 (.42)	BF_{10} = 9.08e+66
Cohort competition (M_{log})	-.74 (.29)	.76 (.56)	BF_{10} = 5.54e+85
Biphone frequency	907.84 (283.3)	905.35 (353.23)	BF_{01} = 7.85
Lexical frequency (M_{log})	.84 (.25)	.85 (.42)	BF_{01} = 7.7
Number of phonemes	5.47 (.5)	5.46 (.5)	BF_{01} = 7.67
Neighborhood density	3.71 (2.97)	3.2 (2.75)	BF_{01} = 2.5
Imageability	4.76 (1.34)	4.76 (1.45)	BF_{01} = 7.87
Acoustic length	750 (97)	749 (83)	BF_{01} = 7.81
LSA values	.06 (.04)	.05 (.04)	BF_{01} = 5.21

Note. Log transformation of mean values is signalled by "(M_{log})". Values in parenthesis represent standard deviations

Bayesian factor (BF) values are based on Bayesian independent samples t-tests

Why is the lexical cohort variable associated with a null effect in vWM tasks?

For cohort models of language processing, stimuli drawn from large cohorts are considered to be more ambiguous during lexical selection because a greater number of lexical competitors are activated simultaneously during the initial stages of speech perception, and have to be inhibited during the lexical selection process (Gaskell & Marslen-Wilson, 2002; Kocagoncu et al., 2017; Marslen-Wilson, 1987; Tyler et al., 2000; Zhuang et al., 2011). Computational implementations of cohort effects consider that words drawn from large cohorts initially receive less activation and hence need more time to reach their activity threshold (Chen & Mirman, 2012). A possible explanation for the observed lack of a cohort competition effect in an immediate serial recall task is that the rapidity of lexical activation (selection) at the encoding stage is not a strong contributor to vWM performance. Contrary to other psycholinguistic effects, the cohort competition variable influences language processing only during the initial stages of lexical selection process, which may not be sufficient to produce measurable differences in terms of recall performance in vWM tasks. For other psycholinguistic effects, such as the imageability

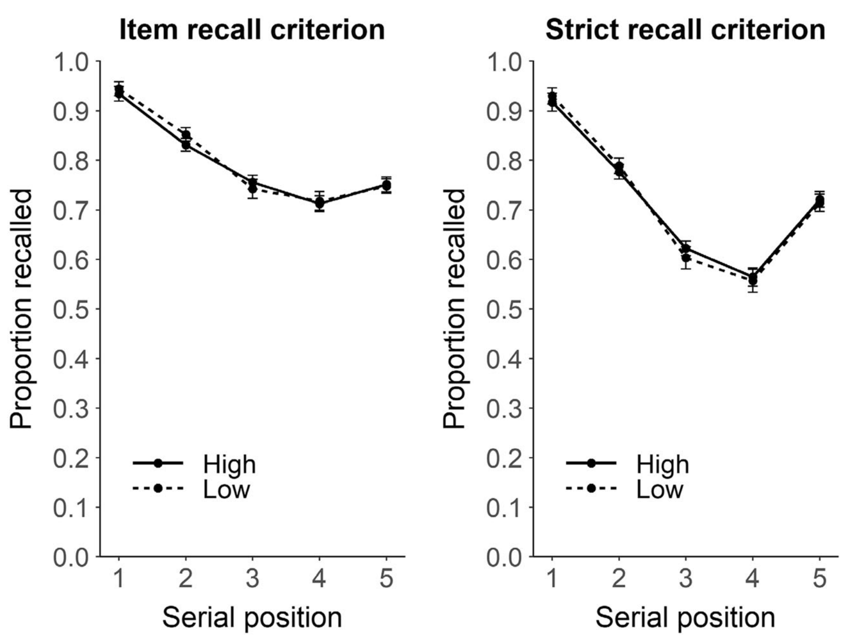


Fig. 3 Experiment 3 – Proportion of items correctly recalled (y-axis) across serial positions (x-axis) as a function of the lexical cohort variable, for item recall- (left panel) and strict recall- (right panel) scoring procedures. The solid and dashed lines represent high- and low-cohort stimuli, respectively. Error bars represent standard errors, after correction for between-subject variability (Cousineau, 2005)

effect, for example, items associated with richer and stable semantic features are considered to be more highly activated due to continuous interactive activations between lexical and semantic representations (Pexman, Lupker, & Hino, 2002; Yap, Lim, & Pexman, 2015), and during all stages (encoding, maintenance and recall) of vWM processing, leading to higher vWM recall performance. Similarly, the semantic similarity effect has been explained by assuming that semantically-related words will continuously reactivate each other through interactive activations via their shared semantic features (Dell et al., 1997), leading to overall higher activation levels.

It could be argued that other psycholinguistic effects can also be explained in terms of speed of lexical activation while still producing measurable effects in vWM tasks such as the lexical frequency effect. The lexical frequency effect has indeed been explained by assuming that high-frequency words have a higher resting activation level and hence can be activated more easily and rapidly (McClelland & Elman, 1986; McClelland & Rumelhart, 1981). At the same time, the lexical frequency effect can also be explained in terms of connection strength between phonological and lexical levels of representations (Besner & Risko, 2016), with higher connection strength for high-frequency words. The result is that for the same amount of activation at the phonological level, high-frequency words will receive more activation and will be more strongly activated as compared to low-frequency words. Finally, it must be noted that the frequency effect is also partly driven by inter-item associations in immediate serial-recall tasks, with high-frequency words co-occuring more frequently than low-frequency words (Hulme et al., 2003; Stuart & Hulme, 2000; Tse & Altarriba, 2007). Due to these higher inter-item associations, high-frequency words may also activate and support each other during WM encoding and maintenance, similar to the semantic similarity effect described above.

More generally, items associated with faster response times in linguistic tasks do not necessarily have a positive effect on WM maintenance and recall performance, as further illustrated by the lexical neighborhood density effect. Indeed, while slower response times have been observed in auditory comprehension tasks for dense neighborhood stimuli, a reverse effect is observed in vWM tasks, with words from dense neighborhoods facilitating recall performance. If vWM performance was to be explained exclusively by the rapidity of lexical activation, reduced performance for dense neighborhood stimuli should be expected. Note that, for now, the null effect observed here for the lexical cohort variable in vWM only holds for the type of task that was used in the different experiments. In immediate serial recall tasks, after encoding, memoranda are maintained via internal mechanisms such as refreshing and rehearsal, and hence are no longer externally driven, contrary to lexical decision tasks. The null effect observed in this study could be re-examined using running span tasks relying on very rapid presentation of memoranda, and unexpected, immediate output diminishing the role of internally generated representations. It should, however, be noted that, for those psycholinguistic effects that have been examined with this type of task, the effects are very similar to those observed in immediate serial recall tasks (see Kowialiewski & Majerus, 2018b).

Consequences for theoretical frameworks

The absence of a cohort competition effect on vWM performance suggests that language-based models of vWM need to distinguish between the speed of lexical activation and the stability of lexical activation. To our knowledge, this is the first study investigating the effect of the speed of lexical activation on vWM performance as reflected by the cohort competition effect. For language-based and, more broadly, activation-based models of vWM (Acheson & MacDonald, 2009; Cowan, 1995, 2001; Majerus, 2013; N. Martin et al., 1996), verbal items are supposed to be activated in vWM using the same mechanisms as those used in language processing more generally. Hence, mechanisms related to the speed of lexical activation should also operate in these models, although they do not (yet) explicitly include them. In these models, items need to be constantly refreshed using the focus of attention and/or rehearsal, otherwise they will be rapidly forgotten due to decay/interference. It logically follows that items that are more strongly activated are less likely to decay up to the point of being forgotten, leading to higher vWM span. In contrast, we may consider that the speed of initial activation is supposed to have a more negligible impact, because it has only a limited influence on the overall activation level and/or decay.

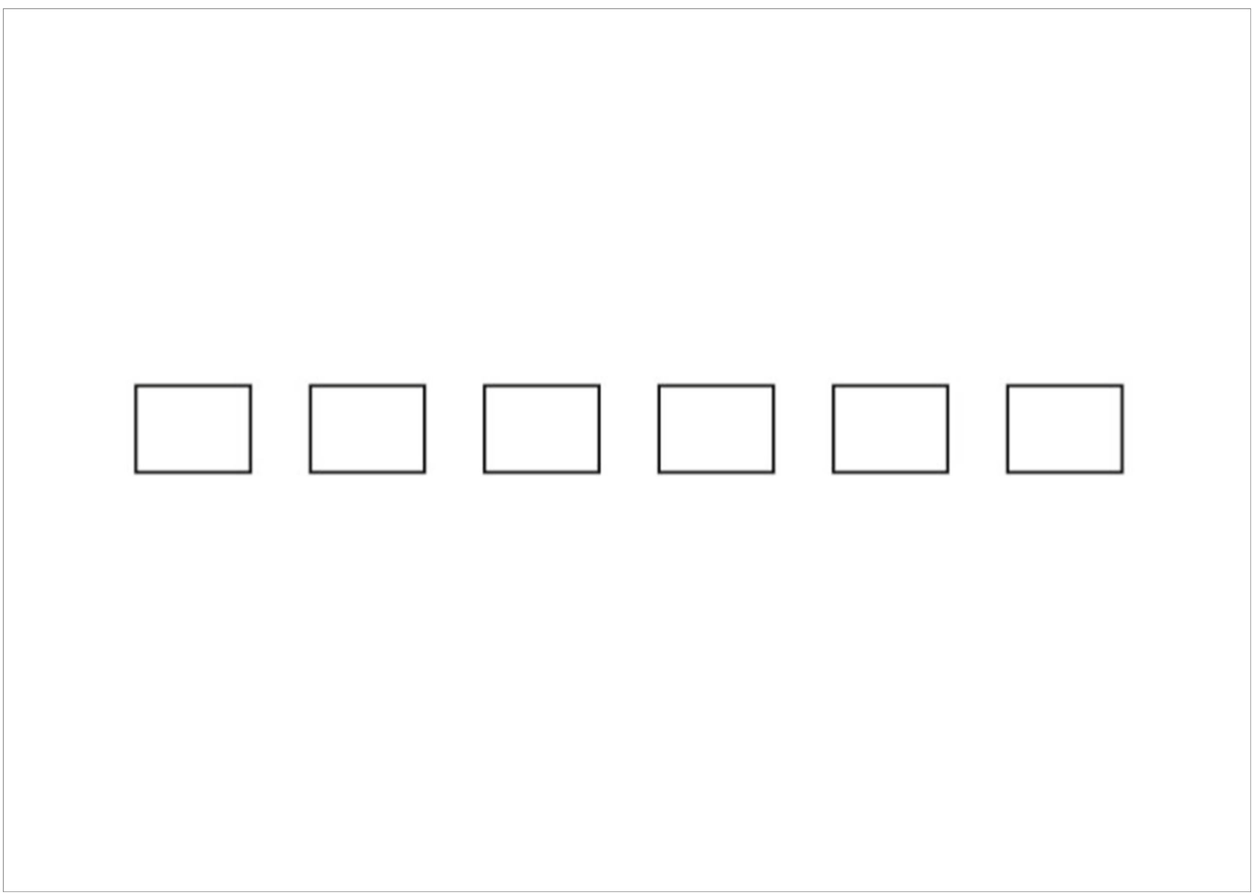
It could also be argued that the results of the present experiments support the redintegration framework, which assumes that, during the recall phase of vWM processing, a reconstruction process occurs to "redintegrate" the degraded traces that have been maintained in a phonological buffer. In a strong version of this theoretical framework (Hulme et al., 1991; Lewandowsky, 1999; Schweickert, 1993), lexical and/or semantic knowledge affect vWM processing only during the recall phase, while the encoding stage is characterized by the maintenance of only phonological codes. The recall advantage for words over nonwords, for instance, is explained by assuming that the redintegration process will use the stored lexical representations to clean up the degraded phonological traces of word stimuli. This model predicts an absence of cohort competition effect because this variable affects only the speed of activation of items during the encoding stage and during the redintegration process, but not the quality of their activation in terms of availability, strength, or robustness. At the same time, it should be noted that other evidence is not in favor of a redintegration mechanism as the exclusive account of psycholinguistic effects in vWM. For instance, strong lexicality effects have been observed in vWM tasks that do not require overt recall and redintegration (Jefferies, Frankish, & Lambon Ralph, 2006b; Kowialiewski & Majerus, 2018a; Savill, Ellis, & Jefferies, 2016). The neighborhood density effect is also of interest here. As explained above, words from dense neighborhood structures are better recalled in vWM tasks, while the redintegration hypothesis would predict the reverse: when reconstructing degraded phonological traces for words from dense neighborhoods, recall performance should decrease due to the many competing words (neighbors) that can potentially be selected for reconstructing the target word. In contrast, interactive activation models predict that high-neighborhood items should be better recalled, because they reactive each other via their shared phonological features (Gordon & Dell, 2001), resulting in a greater amount of activation, and will consequently be less affected by decay/interference. In sum, language-based models of vWM assuming interactive activation within the linguistic system during encoding, maintenance, and recall provide a theoretical framework that is able to deal with a wider range of empirical data, including those observed in the present study.

Conclusions

The absence of a cohort competition effect observed in vWM suggests that the speed of lexical activation is not a critical factor for vWM performance. Instead, the psycholinguistic effects that have a robust impact on vWM performance are driven by the strength and robustness of lexical activation.

Acknowledgements We thank S. Moes and C. Tonon for their help in data acquisition and all the participants for their time devoted to this study.

Appendix A. Recall sheet used in Experiment 1



Appendix B. Mixed model analysis for the effect of cohort competition in Experiment 2

The mixed model analysis was launched using the lme4 and lmerTest (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017) packages under R (R Development Core Team, 2008). We ran the model on response times as dependent variable, with lexicality (word, nonword), cohort competition (high, low), and the interaction term as fixed effects. The participants and items intercepts were set as random effects. Because the full model failed to converge with maximum random parameters, by-item and by-participant were set as random slopes for the effect of cohort competition, while only by-participant was set as random slope for the effect of lexicality. We found an effect of lexicality (t = -3.067, p = .00274), cohort competition (t = -3.3, p = .00103) but no interaction (t = .664, p = 0.50672), suggesting that the effect of cohort competition was equally observed for both words and nonwords. The analysis was launched using the following R code: lexical_decision.model = lmer(RT ~ competition + lexicality + (competition:lexicality) + (competition | items) + (competition + lexicality | participants), data = lexical_decision) summary(lexical_decision.model)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50–68. doi: https://doi.org/10.1037/a0014411

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology*, *18*(4), 362–365. doi:https://doi.org/10.1080/14640746608400055

Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. In *The cognitive neuroscience of working memory* (pp. 59–80).

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:https://doi.org/10.18637/jss.v067.i01

Besner, D., & Risko, E. F. (2016). Thinking outside the box when reading aloud: Between (localist) module connection strength as a source of word frequency effects. *Psychological Review*, *123*(5), 592–599.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, *26*(5), 467.

Campoy, G., Castellà, J., Provencio, V., Hitch, G. J., & Baddeley, A. D. (2015). Automatic semantic encoding in verbal short-term memory: Evidence from the concreteness effect. *The Quarterly Journal of Experimental Psychology*, *68*(4), 759–778. doi:https://doi.org/10.1080/17470218.2014.966248

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *199*(2), 417–430. doi:https://doi.org/10.1037/a0027175.Competition

Clarkson, L., Roodenrys, S., Miller, L. M., & Hulme, C. (2016). The phonological neighbourhood effect on short-term memory for order. *Memory*, *25*(3), 391–402. doi:https://doi.org/10.1080/09658211.2016.1179330

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45.

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–185. doi:https://doi.org/10.1017/S0140525X01003922

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838.

Evans, G. A. L., Lambon Ralph, M. A., & Woollams, A. M. (2012). What's in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin & Review*, *19*(2), 325–331. doi:https://doi.org/10.3758/s13423-011-0213-7

Gaskell, M. G., & Marslen-Wilson, W. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 84–95. doi:https://doi.org/10.1037/0278-7393.25.3.562

Gordon, J. K., & Dell, G. S. (2001). Phonological neighborhood effects: Evidence from aphasia and connectionist modeling. *Brain and Language*, *79*(1), 21–31. doi:10.1006/brln.2001.2574

Guérard, K., & Saint-Aubin, J. (2012). assessing the effect of lexical variables in backward recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 312–324. doi:https://doi.org/10.1037/a0025481

Gupta, P. (2009). A computational model of nonword repetition, immediate serial recall, and nonword learning. In A. Thorn & M. Page (Eds.), Interactions between short-term and long-term memory in the verbal domain, (pp. 108–135). Hove, UK: Psychology Press.

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, *33*(6), 1001–1016.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*(6), 685–701.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1217–1232.

Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*(4), 500–518. doi:https://doi.org/10.1016/S0749-596X(03)00096-2

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006a). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language*, *54*(1), 81–98. doi:https://doi.org/10.1016/j.jml.2005.08.001

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006b). Lexical and semantic influences on item and order memory in immediate serial recognition : Evidence from a novel task. *The Quarterly*

Journal of Experimental Psychology, *59*(5), 949–964. doi:https://doi.org/10.1080/02724980543000141

Jeffreys, H. (1998). The theory of probability. Oxford, UK*:* Oxford University Press.

Kocagoncu, E., Clarke, A., Devereux, B. J., & Tyler, L. K. (2017). Decoding the Cortical Dynamics of Sound-Meaning Mapping. *The Journal of Neuroscience*, *37*(5), 1312–1319. doi:https://doi.org/10.1523/JNEUROSCI.2858-16.2016

Kowialiewski, B., & Majerus, S. (2018a). Testing the redintegration hypothesis by a single probe recognition paradigm. *Memory*, *0*(0), 1–9. doi:https://doi.org/10.1080/09658211.2018.1448420

Kowialiewski, B., & Majerus, S. (2018b). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language*, *101*, 64–83. doi:https://doi.org/10.1016/j.jml.2018.03.005

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package : Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. doi:https://doi.org/10.18637/jss.v082.i13

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Lewandowsky, S. (1999). Redintegration and Response Suppression in Serial Recall: A Dynamic Network Model. *International Journal of Psychology*, *34*(5/6), 434–446.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. doi:https://doi.org/10.1016/j.jesp.2013.03.013

Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience*, *7*, 357. doi:https://doi.org/10.3389/fnhum.2013.00357

Majerus, S., Van der Linden, M., Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language*, *51*(2), 297–306. doi:https://doi.org/10.1016/j.jml.2004.05.002

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. doi:https://doi.org/10.1016/0010-0277(87)90005-9

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language*, *47*(4), 609–660.

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory–verbal STM capacity and lexical errors in repetition. *Brain and Language*, *52*(1), 83–113.

Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*(1), 3–29.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame : An open-source , graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. doi:https://doi.org/10.3758/s13428-011-0168-7

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*(5), 375.

New, B. (2006). Lexique 3: Une nouvelle base de données lexicales. *Actes de La Conférence Traitement Automatique Des Langues Naturelles*.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779–819. doi:https://doi.org/10.3758/s13423-012-0272-4

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number of features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*, *9*(3), 542–549.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology*, *48*(2), 384–404. doi:https://doi.org/10.1080/14640749508401396

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology*, *50*(4), 408–412.

Poirier, M., & Saint-Aubin, J. (2005). Word frequency effects in immediate serial recall: Item familiarity and item co-occurrence have the same effect. *Memory*, *13*(3–4), 325–332. doi:https://doi.org/10.1080/09658210344000369

R Development Core Team. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org.

Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology*, *61*(2), 292–323. doi:10.1080/17470210601147747

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019–1034. doi:https://doi.org/10.1037//0278-7393.28.6.1019

Sadat, J., Martin, C. D., Costa, A., & Alario, F. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58. doi:https://doi.org/10.1016/j.cogpsych.2013.10.001

Saint-Aubin, J., Guérard, K., Chamberland, C., & Malenfant, A. (2014). Delineating the contribution of long-term associations to immediate recall. *Memory*, *22*(4), 360–373. doi:https://doi.org/10.1080/09658211.2013.794242

Savill, N., Ellis, A. W., & Jefferies, E. (2016). Newly-acquired words are more phonologically robust in verbal short-term memory when they have associated semantic representations. *Neuropsychologia*, *98*, 1–13. doi:https://doi.org/10.1016/j.neuropsychologia.2016.03.006

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322. doi:https://doi.org/10.1037/met0000061

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*(2), 168–175.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149.

Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 796–802. doi:https://doi.org/10.1037//0278-7393.26.3.796

Team, J. (2017). JASP (Version 0.8.1.1)[Computer software].

Tse, C.-S., & Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects. *Memory*, *15*(6), 675–690. doi:https://doi.org/10.1080/09658210701467186

Tubach, J. P., & Boë, L. J. (1990). Un corpus de transcription phonétique(300 000 phones). Constitution et exploitation statistique.

Tyler, L. K., Moss, H. E., Galpin, A., & Voice, J. K. (2002). Activating meaning in time: The role of imageability and form-class. *Language and Cognitive Processes*, *17*(5), 471–502. doi:https://doi.org/10.1080/01690960143000290

Tyler, L. K., Voice, J. K., & Moss, H. E. (2000). The interaction of meaning and sound in spoken word recognition. *Psychonomic Bulletin & Review*, *7*(2), 320–326. doi:https://doi.org/10.3758/BF03212988

Vitevitch, M. S., Chan, K. Y., & Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language*, *67*(1), 30–44. doi:https://doi.org/10.1016/j.jml.2012.02.008

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). doi:https://doi.org/10.1037/a0022790

Watkins, O. C., & Watkins, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(6), 712–718.

Yap, M. J., Lim, G. Y., & Pexman, P. M. (2015). Semantic richness effects in lexical decision: The role of feedback. *Memory & Cognition*, *43*(8), 1148–1167. doi:https://doi.org/10.3758/s13421-015-0536-0

Zhuang, J., Randall, B., Stamatakis, E. a., Marslen-Wilson, W. D., & Tyler, L. K. (2011). The Interaction of Lexical Semantics and Cohort Competition in Spoken Word Recognition: An fMRI Study. *Journal of Cognitive Neuroscience*, *23*(12), 3778–3790. doi:https://doi.org/10.1162/jocn_a_00046

Zhuang, J., Tyler, L. K., Randall, B., Stamatakis, E. A., & Marslen-Wilson, W. D. (2014). Optimally efficient neural systems for processing spoken language. *Cerebral Cortex*, *24*(4), 908–918. doi:https://doi.org/10.1093/cercor/bhs366