The cost of learning new meanings for familiar words

Greg Maciejewski[1], Jennifer M. Rodd[2], Mark Mon-Williams[1], & Ekaterini Klepousniotou[1]

[1] School of Psychology, University of Leeds, UK

[2] Department of Experimental Psychology, University College London, UK

Corresponding author:

Dr. Ekaterini Klepousniotou

School of Psychology

University of Leeds

Leeds

LS2 9JT

Email: e.klepousniotou@leeds.ac.uk

Phone: +44 (0)113 343 5716

Abstract

Research has shown that adults are skilled at learning new words and meanings. We examined whether learning new meanings for familiar words affects the processing of their existing meanings. Adults learnt fictitious meanings for previously unambiguous words over four consecutive days. We tested comprehension of existing meanings using a semantic relatedness decision task in which the probe word was related to the existing but not the new meaning. Following the training, responses were slower to the trained, but not to the untrained, words, indicating competition between newly-acquired and well-established meanings. This effect was smaller for meanings that were semantically related to existing meanings than for the unrelated counterparts, demonstrating that meaning relatedness modulates the degree of competition. Overall, the findings confirm that new meanings can be integrated into the mental lexicon after just a few days' exposure, and provide support for current models of ambiguity processing.

Keywords: lexical/semantic ambiguity; semantic processing; language acquisition; vocabulary

Introduction

Language is perpetually in flux, such that even adults must often learn new meanings for words they already know. For example, recent advancements in computer technology have resulted in new meanings for the words “mouse”, “virus”, and “cloud”, while those using social networking websites have recently learnt new meanings for the words “follow”, “tweet”, and “post”. Adults may also encounter familiar words in new contexts when they take up a hobby or join a community. For instance, those starting a degree in statistics need to learn new, highly specific meanings for the words “variable”, “significant”, and “model”. Therefore, it appears that the ability to learn new meanings for known words continues to be important throughout adult life. Not only does this ability allow us to acquire entirely new information, but it also modifies our existing knowledge of words and the way we use them, which is evident in the ubiquity of distinct forms of semantic ambiguity in all languages.

Most of the new meanings we need to learn are somewhat related to the existing meanings of words with respect to physical properties (e.g., “mouse”), function (e.g., “virus”), or other conceptual features. This form of ambiguity between related word senses - polysemy - is very common across languages (Srinivasan & Rabagliati, 2015) as it reflects speakers’ tendency to use existing words to label novel albeit conceptually related objects, concepts, and actions (Clark & Clark, 1979; Lehrer, 1990; Nunberg, 1979). It is important to note though that polysemous words differ in how their senses are related and extended (for a recent review, see Vincente, 2018). In regular/metonymic polysemy, the multiple senses of a word are highly related and follow common and predictable patterns of extension, such as the animal for meat (e.g., “rabbit”) and instrument for action sense alternations (e.g., “shovel”). In irregular polysemy, on the other hand, the senses are loosely and often figuratively related, and the way they are extended is idiosyncratic and unique to a particular word (e.g., “drone” denoting a male bee or a type of aircraft; “eye” denoting an organ or a hole in a needle). Nevertheless, polysemy as a whole can be easily distinguished from homonymy in which a single word form is associated with multiple unrelated meanings (e.g., “bank”). This form of ambiguity, considered a historical accident, is far less common than polysemy (Rodd, Gaskell, & Marslen-Wilson, 2002) and corresponds to new meanings that are seemingly unrelated to the original meanings of words (e.g., “catfish” denoting a type of fish or an individual who has a false online identity).

While there have been multiple investigations into learning new words (for a review, see Davis & Gaskell, 2009), little is known about adults’ ability to learn new meanings for words that already exist – an important prerequisite for skilled language use. Extensions of the work on word learning into the semantic domain are clearly warranted as the questions of how and when new meanings are integrated into existing lexical-semantic representations, and how they affect access to those representations, remain largely unexplored. To date, a few studies (Clark & Gerrig, 1983; Frisson & Pickering, 2007; McElree, Frisson, & Pickering, 2006) have shown that adults can easily derive new senses of familiar words from context, provided that the interpretation follows the conventional pattern of metonymic sense extension, such as the producer for product sense alternation (e.g., “to study Darwin” or “to read Dickens”). A more recent study (Rodd et al., 2012) has also found that adults are good at learning new loosely related meanings (e.g., “sip” denoting a small amount of hacked computer data), either incidentally through reading short text or intentionally through intensive training. While it appears that learning new (related) meanings for familiar words is a relatively easy task, the question we address in the current study is whether and how it affects the processing of existing meanings.

More specifically, the present experiments examine the prediction in the semantic ambiguity literature that long-term consolidation of new meanings would slow the comprehension of existing meanings as a result of semantic competition. Although to date there is no evidence to support this prediction for newly-learnt word meanings, there are a few studies to suggest that such competition is likely to arise (Fang & Perfetti, 2017; Fang & Perfetti, 2019; Fang, Perfetti, & Stafura, 2017; Rodd et al., 2012). For example, Fang and Perfetti (2017) found that even the attempt to learn new meanings can hinder access to well-established meanings, manifesting as reduced semantic priming from existing meanings, shortly after the learning phase, before new meanings were fully integrated into the mental lexicon. In a more recent study, however, Fang and Perfetti (2019) showed that this interference was short-lived without further training and restricted to high-frequency words (e.g., “plenty”). Learning new meanings for low-frequency words (e.g., “exodus”) appeared to serve as an opportunity to reconsolidate their existing meanings instead. In yet another study, Fang et al. (2017) conversely found that it is also possible for existing meanings, especially those of high-frequency words, to hinder access to new meanings, again as early as the learning phase. Taken together, these studies suggest that the learning experience per se can produce interference in the retrieval of both new and well-known word meanings.

In contrast to Fang et al. (2017) and Fang and Perfetti (2017) who investigated meaning retrieval during the learning phase, Rodd et al. (2012) explored how consolidation of new meanings impacted on participants’ ability to recognise previously unambiguous words. Their second experiment, which involved a 6-day learning period, revealed shorter lexical decisions to trained than untrained words, suggesting that new meanings had been sufficiently consolidated to influence word processing in a task that did not even require access to semantic knowledge. Interestingly, in their third and final experiment with shorter but more semantically demanding training (e.g., writing a coherent story using new word meanings), Rodd et al. (2012) reported that the processing benefit was larger for words paired with new related than unrelated meanings, which is consistent with the view that polysemy benefits word recognition (e.g., Armstrong & Plaut, 2008; Klepousniotou & Baum, 2007; Rodd et al., 2002).

Overall, two key findings emerge from the study by Rodd et al. (2012). First, while Fang et al. (2017) and Fang and Perfetti (2017, 2019) showed that new semantic knowledge can interact with existing knowledge as soon as the learning phase, Rodd et al.'s (2012) finding of a polysemy advantage only after demanding training suggests that new meanings must be extensively trained and sufficiently consolidated in order to uncover their full impact on existing lexical-semantic representations. Second, Rodd et al. (2012) demonstrated that, once consolidated, new related and unrelated meanings influenced word-form processing in the same way as polysemy and homonymy in existing words, indicating that learning new meanings in experimental settings mirrors the impact of ambiguity in natural language. However, since none of the studies reviewed above used a task that required disambiguation or selection of the well-established meaning following extensive training, the outstanding question is how long-term consolidation of new meanings affects the ability to correctly understand words in their existing meanings. The ambiguity literature is relevant in this regard since it shows that for words that have multiple familiar meanings semantic competition arises between these meanings and results in slowed comprehension.

Evidence for semantic competition between familiar meanings comes from research on the processing of ambiguous words in isolation or neutral context. For example, eye-movement studies (Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986) found that, in late-disambiguation sentences, gaze durations are typically longer for homonyms with balanced meaning frequencies (e.g., "football/electric fan") than for non-homonyms. A similar disadvantage effect has been observed in semantic relatedness decision latencies for word pairs involving both homonyms and polysemes (Gottlob, Goldinger, Stone, & Van Orden, 1999; Hoffman & Woollams, 2015; Pexman, Hino, & Lupker, 2004; Piercey & Joordens, 2000). Overall, the literature suggests that ambiguity, particularly that between unrelated meanings, slows semantic processing due to competition between the multiple interpretations of a word. This competition should be predominantly observed when the word is encountered on its own, or when prior context is not sufficiently strong to bias a particular interpretation (e.g., Duffy et al., 1988; Simpson & Krueger, 1991).

Semantic competition in word comprehension is also a key assumption of existing models of ambiguity processing, particularly those postulating distributed lexical-semantic representation (e.g., Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd, Gaskell, & Marslen-Wilson, 2004). In short, parallel-distributed processing (PDP) models suggest that the consistency of form-to-meaning mapping determines the speed of the semantic activation process. For ambiguous words with inconsistent form-to-meaning mappings, activation of the single orthographic representation triggers initial activation of multiple semantic representations that compete for full activation of their respective semantic features, thus slowing semantic processing.

Although the idea remains somewhat controversial (for a review, see Eddington & Tokowicz, 2015), some of the PDP models (Armstrong & Plaut, 2008; Rodd et al., 2004) also suggest that the degree of semantic competition may additionally depend on the form of ambiguity, or relatedness in meaning. In particular, Rodd et al. (2004) argue that because the different senses of polysemes share at least some semantic features (e.g., “to dip a brush in paint” vs. “to take a dip in the pool”), their form-to-meaning mappings may be more consistent than those for homonyms, and therefore produce less competition in the race for semantic activation.

In summary, the ambiguity literature makes two important predictions - newly-acquired meanings should slow the comprehension of existing meanings through semantic competition, and this effect should be greater for new unrelated meanings. Two experiments were designed to test these predictions. Training materials were adapted from Rodd et al. (2012). New related meanings imitated irregular polysemy, whilst the unrelated counterparts imitated homonymy. For the former, new meanings were loosely related to original meanings through a single semantic feature and could not be derived through a rule of sense extension typical of regular polysemy/metonymy (e.g., animal-for-meat or instrument-for-action relations). Likewise, our training was largely based on that of Rodd et al. (2012, Experiment 3) who were successful in teaching adult participants a large number of new meanings and demonstrated that their intensive, 4-day learning period allowed those meanings to be sufficiently consolidated to influence online word recognition. This is also in line with studies of word learning which suggest that while a few exposures may be sufficient to learn new word forms, this knowledge is not normally integrated into the mental lexicon until after offline sleep-dependent consolidation has taken place (for a review, see Davis & Gaskell, 2009). This literature in particular motivated us to employ multi-day training that would allow new meanings to develop robust lexical-semantic representations and produce potential competition.

In order to establish the impact of such consolidation on the processing of existing meanings, a semantic relatedness decision task was used in which trained words (e.g., “sip” denoting a small amount of hacked computer data) were probed with words that related to the existing meaning (“sip-liquid”) or were unrelated (“sip-eel”). Participants’ responses to the same target-probe word pairs were compared before and after training. This task was chosen because it required selection of the existing, dominant meaning, and thus tapped into word disambiguation. Note that we did not include probe words instantiating the new meanings so that any interference in the post-training performance could be attributed to consolidating the new meaning, rather than explicit switching between the new and original meanings throughout the task.

We predicted responses to otherwise unambiguous words to be slower after training, particularly when the new meanings were unrelated to the existing ones (e.g., Armstrong & Plaut, 2008; Klepousniotou, Titone, & Romeo, 2008; Rodd et al., 2004). We assumed that this training effect would indicate slower activation of response-relevant features of well-established meanings due to competition from response-irrelevant features of newly-learnt meanings. This was in line with earlier studies (Fang & Perfetti, 2017, 2019) suggesting that existing meanings become less accessible while learning new meanings. We also expected this training effect to appear on “yes” trials involving related word pairs as well as “no” trials involving unrelated word pairs. The rationale was that while the new and the existing meaning were consistent with the same response on “no” trials (e.g., “sip-eel”), they could possibly trigger response conflict on “yes” trials (e.g., “sip-liquid”) after the training had taken place (Pexman et al., 2004). The finding of a comparable training effect on both trials was, therefore, critical to explaining the effect in terms of changes to semantic activation processes, rather than changes to response-selection demands of the task. On the whole, then, the current study sought support for the prediction that, once integrated into the mental lexicon, newly-acquired meanings compete with well-established meanings.

Experiment 1

Method

Participants

Twenty students and members of staff [14 females; aged 19-48 (*M* = 30.5, *SD* = 11.1)] from the University of Bedfordshire participated in the experiment in exchange for a £20 voucher. This sample size was deemed appropriate based on Rodd et al.'s (2012) work (15-22 participants per experiment). Participants were monolingual native speakers of British English with no known history of language-/vision-related difficulties/disorders. All reported to be right-handed. The experiment received ethical approval from the Department of Psychology, University of Bedfordshire Ethics Committee.

Materials

New word meanings

Thirty-two target words and short paragraphs describing their new related meanings (e.g., “sip” denoting a small amount of hacked computer data) were taken from Rodd et al. (2012)[1]. The paragraphs used each word in its new meaning five times, such that each instance provided a different piece of information about the new word referent (e.g., one sentence explained what a sip was, whereas another mentioned that extracting data in sips prevents hackers from being caught). Most of the new meanings referred to recent inventions, colloquial and scientific terms, or social phenomena (see the definitions in Appendix 1), and they were related to the existing meanings with respect to function (e.g., “bone” as the core of a star; n = 5), physical properties (e.g., “foam” as a type of nuclear waste; n = 12), being a specific variant of a more general meaning (e.g., “crew” as a group of musicians; n = 7), or the imagery that the word elicited (e.g., “hive” as a busy household; n = 8)[2]. Thus, as in existing irregular polysemes, the new meanings were related to the original meanings through a single feature but could not be derived via a productive rule (e.g., animal-for-meat or part-for-whole relations) as the relationship between the meanings was unpredictable and unique to each word. New unrelated meanings were, on the other hand, created by swapping the paragraphs across pairs of targets

---

[1] The word “slim” in Rodd et al.’s (2012) stimulus list was changed to “hamster” (Experiment 1) or “mouse” (Experiment 2) so that all trained words had noun/noun-verb interpretations. The word “hamster” was replaced with “mouse” so that lexical and semantic properties of the trained and untrained targets in Experiment 2 were matched more rigorously.

[2] As the experiment was not explicitly designed to explore the type of the relationship between the new and the existing meaning (e.g., physical properties vs. function), future studies will need to establish whether there could be an impact on learning performance based on the way new meanings are related.

to minimise any overlap between the related and unrelated meanings for each word. Two versions of the paragraphs were created so that each contained 16 words with new related meanings and 16 words with new unrelated meanings. The related meanings in Version 1 were presented as unrelated in Version 2, and vice versa. Participants were pseudo-randomly assigned to learn new meanings from either version. The words used in these paragraphs constituted “trained” words in the experiment.

Relatedness decision task

Each trained word served as a target in the semantic relatedness decision task assessing the comprehension of existing meanings. To examine potential practice/session effects on task performance, the stimulus list also included 16 untrained control words that did not feature in any of the training materials. All the trained and untrained targets had noun or noun-verb interpretations (but were all used as nouns in the task) and only one meaning in the Wordsmyth Dictionary (Parks, Ray, & Bland, 1998). Although both trained and untrained targets had a few related word senses, neither exhibited patterns of sense extension typical of metaphorical (e.g., animal-for-human-characteristic relations) or metonymic polysemy (e.g., animal-for-meat relations). The two types of targets were also statistically comparable (all *t*s < 1.5) with respect to nine lexical and semantic variables, such as word-form frequency and the number of related word senses (see target properties in Table 1 below).

Table 1. Experiment 1: Descriptive statistics of lexical and semantic properties of the target words.

Variable	Trained targets	Untrained targets
Example	"sip"	"cash"
Letters	4.4 (1.1)	4.8 (1.0)
Raw frequency	16.5 (17.6)	19.1 (12.7)
Log frequency	1.0 (0.5)	1.2 (0.3)
Orthographic neighbours	6.8 (5.3)	6.4 (5.0)
Wordsmyth senses	4.7 (2.1)	4.2 (1.5)
WordNet senses	4.5 (2.3)	4.3 (1.0)
Concreteness	6.2 (1.4)	6.2 (2.1)
Imageability	5.5 (0.6)	5.8 (0.5)
Age of acquisition	6.2 (1.4)	6.2 (2.1)

Note. Standard deviations are given in the parentheses. Word-form frequency and the number of orthographic neighbours come from the MCWord Database (Medler & Binder, 2005). Wordsmyth and WordNet sense counts come from the Wordsmyth (Parks et al., 1998) and WordNet dictionaries (Fellbaum, 1998), respectively. Concreteness and imageability ratings come from the MRC Psycholinguistic Database (Coltheart, 1981). Age-of-acquisition ratings come from Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012).

Each target was paired with six probe words – three semantically related to the existing but not the new meaning (e.g., "sip-liquid") and three unrelated to either meaning of the target (e.g., "sip-eel"). The number of probes was tripled to increase the number of observations and to generalise training effects across different pairs of words. The pairs were presented using a within-participants design, such that each participant responded to the same target six times but only once to each of the probes. Most of the probes were related to the targets through category membership (e.g., "hive-nest"), physical properties (e.g., "beef-lamb"), or object-action relationship (e.g., "bandage-wrap"). The related and unrelated probes had only one meaning in the Wordsmyth Dictionary (Parks et al., 1998) and were matched, at the group level, on word-form frequency and length (see Table 2 below) across the pairs involving the trained and untrained targets (all Fs < 1.5).

Prior to the experiment, 15 monolingual native speakers of British English [11 females; aged 20-39 (*M* = 31.0, *SD* = 5.3)] rated target-probe relatedness on a 7-point scale (where 1 denoted "highly unrelated" and 7 denoted "highly related"). This online stimulus pre-test confirmed that the related/unrelated pairs were judged as such, and that the degree of relatedness or unrelatedness did not significantly differ (both *t*s < 1.5) between the sets of trained and untrained targets (see Table 2 below). All the word pairs used in Experiment 1 are presented in Appendix 2.

Table 2. Experiment 1: Descriptive statistics of lexical and semantic properties of the probe words.

Variable	Trained probes		Untrained probes	
	Related targets	Unrelated targets	Related targets	Unrelated targets
Example	"sip-liquid"	"sip-eel"	"cash-receipt"	"cash-moth"
Letters	5.0 (1.1)	5.3 (1.2)	5.1 (1.2)	5.4 (1.1)
Raw frequency	16.6 (12.0)	16.7 (11.6)	18.8 (13.1)	16.4 (13.1)
Log frequency	1.1 (0.4)	1.1 (0.3)	1.2 (0.4)	1.1 (0.4)
Target-probe relatedness	6.1 (0.4)	1.8 (0.3)	6.2 (0.4)	1.9 (0.3)

Note. Standard deviations are given in the parentheses. Information on the different variables can be found in the note for Table 1.

Worksheets

Participants completed four online worksheets, adapted from Rodd et al. (2012), on four consecutive days to further consolidate the new word meanings before their final testing session on Day 5. On Day 1, Worksheet 1 involved selecting the trained words from a drop-down menu and matching them to brief definitions of their new meanings. On Day 2, Worksheet 2 involved writing a new example sentence for each trained word that was compatible with its new meaning. On Day 3, Worksheet 3 involved writing a coherent story using all the trained words in their new-meaning context. On Day 4, Worksheet 4 involved answering one open-ended question about each of the new word referents. For Worksheets 2 and 3, participants were instructed to provide sufficiently detailed context that would clearly convey the new meanings. There was no word-count limit, and participants could write in any style and on any subject. However, they had to use each of the trained words at least once. The trained words were presented randomly in Worksheets 1 and 4 but alphabetically in Worksheets 2 and 3. The worksheets were designed and administered using the Qualtrics survey builder (http://qualtrics.com/).

Procedure

The experiment (for an overview, see Figure 1 below) took place over five consecutive days and lasted for four hours in total. Following Rodd et al. (2012, Experiment 3), the experiment consisted of an initial lab-based training session on Day 1, four home-based training sessions involving the online worksheets on Days 1-4, and a final lab-based testing session on Day 5. On Day 1, participants completed a pre-training relatedness decision task and then read paragraphs describing new word meanings. Later that day and over the following three days, participants completed the worksheets. On Day 5, they came back to the lab to complete the same relatedness decision task (using the same stimuli as on Day 1), followed by a recall task assessing their memory for the new meanings and a rating task assessing the semantic relationship between the new and existing meanings of the trained words. Each participant completed the two lab-based sessions at a similar time of the day (+/- 3 hours), exactly five days apart. All the lab-based tasks were programmed in SuperLab 4.5 (http://superlab.com/).

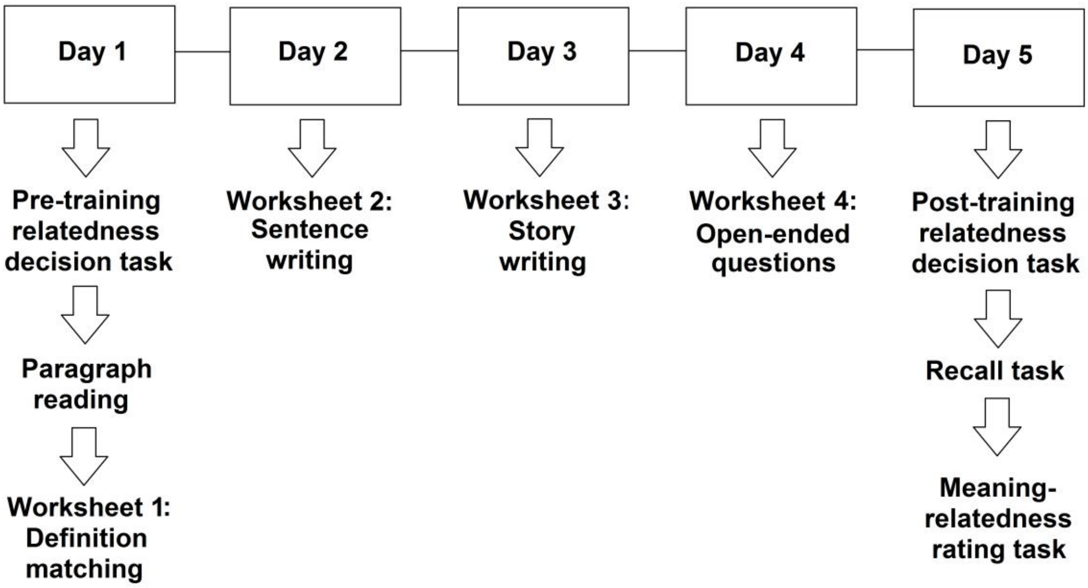

Figure 1. Overview of Experiment 1.

Relatedness decision task

In this task, participants decided whether the target and the probe were related in meaning by pressing keyboard buttons (A labelled “no”, L labelled “yes”). Participants made “yes” responses with the index finger of their dominant (right) hand and “no” responses with the index finger of their left hand. On both testing sessions (Days 1 & 5), the task began with 10 randomised practice trials with feedback on both response accuracy and latency. The experimental stimuli were presented in three blocks, such that each block contained the same target with a different related and unrelated probe. There were two self-paced breaks – one after the first block and the other after the second block. Trials began with a 500 ms fixation cross, followed by a target presented for 300 ms. A probe appeared immediately after the target (0 ms inter-stimulus interval) and remained on the screen until participants made a response. There was a 500 ms delay between trials. Both response speed and accuracy were emphasised in the instructions, and participants were instructed and given examples of what constitutes semantic relatedness. The instructions on Day 5 were the same as those on Day 1 and did not mention anything about the new meanings of the words.

Paragraph reading

Following the relatedness decision task on Day 1, participants read short paragraphs describing new meanings. The paragraphs were presented on a computer screen, one at a time in randomised order. Participants pressed the spacebar to indicate when they had finished reading each paragraph. To ensure they read the text slowly and carefully, 500 ms after having pressed the spacebar each paragraph was followed by a yes-no question that was related to a specific feature of the new word referent (e.g., “Can only hackers extract sips”?). Once participants answered the question (by pressing the L button labelled “yes” or the A button labelled “no”), the next paragraph appeared after 500 ms. There was an equal number of “yes” and “no” responses in the task. Participants had as much time as they needed to read the paragraphs and answer the questions.

Worksheets

At the end of Day 1, participants received a paper booklet containing the paragraphs and were instructed to use it as a companion for all the worksheets. The order of the worksheets was the same for all participants. Participants completed Worksheet 1 by the end of Day 1 after the lab-based testing session. For the other worksheets (2-4), they received access to a given worksheet at 8 a.m. on each day and had to complete it by midnight of that day. All the participants completed the worksheets within this timescale.

Recall task

On Day 5, participants came back to the lab and first performed the same relatedness decision task as on Day 1. They then completed a recall task in which they recalled and typed a maximum of nine features/properties that were true of the new word referents only. Participants had as much time as they needed to complete this task but could not use the companion booklet. They typed in “nothing” if they could not recall any information and pressed the ALT button to move to another word which appeared after a delay of 500 ms. The words were presented one a time in randomised order.

Meaning-relatedness rating task

At the end of the experiment, participants rated the semantic relatedness between the existing and the new meaning of each trained word on a 7-point scale (where 1 denoted “highly unrelated” and 7 denoted “highly related”). The words were presented in randomised order, together with the paragraphs that participants had read on Day 1. The aim of this task was to verify that participants considered the new related/unrelated meanings as such.

Results

Meaning-relatedness rating task

Our first aim was to confirm that the experiment was successful at manipulating the semantic relatedness between the new and the existing meaning. Participants' ratings of meaning relatedness were analysed using a generalised mixed-effects model fitted with the Poisson probability distribution[3]. The model included the factors of Meaning Type (new related meaning, new unrelated meaning) and Version (1, 2). There were no effects of Version in any of the tasks. Thus, throughout the study, effects involving Version are not reported as the sole purpose of this factor was to account for potential effects of counter-balancing (Pollatsek & Well, 1995). Following Barr, Levy, Scheepers, and Tily (2013) and Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017), the optimal random-effects structure justified by the data in all our analyses was identified using forward model selection[4]. For the

[3] We first attempted to analyse the ratings using a linear mixed-effects model. However, the residuals of the model showed an inverse normal distribution that was insensitive to data transformation, violating the assumption of linear but not generalised mixed-effects modelling.

[4] We began analysis with a model that included significant random intercepts and tested all possible slopes for inclusion separately. Out of significant slopes, we first added the most influential one (based on the value of χ^2 from model-comparison tests) to the base model and then tested whether the second most influential slope further improves the model. We continued to test and include the remaining slopes until the model failed to converge.

ratings of meaning relatedness, the model included significant random intercepts for subjects and items and a random slope for Version across items. Fixed effects were tested using likelihood-ratio tests comparing full and reduced models. All modelling was conducted using the “lme4” package (Bates, Mächler, & Bolker, 2011) in R (R Development Core Team, 2004). Following Nakagawa and Schielzeth (2013) and Johnson (2014), marginal R^2 (variance explained by fixed effects) and conditional R^2 (variance explained by fixed and random effects) for all mixed-effects models were estimated using the “MuMIn” package (Bartoń, 2014).

The model (marginal R^2 = .36, conditional R^2 = .48) revealed a significant effect of Meaning Type [$\chi^2(1) = 51.9$, $p < .001$]. As expected, new meanings in the related condition (M = 4.5, SD = 0.6) were rated as more semantically related to existing meanings than new meanings in the unrelated condition (M = 1.9, SD = 0.6). We further tested the effectiveness of the relatedness manipulation using a logistic regression model that predicted item category (new related vs. new unrelated meaning) based on mean item ratings and the factor of Version. The ratings accounted for a considerable amount of variance in item category (Cox & Snell’s R^2 = .65; Nagelkerke’s R^2 = .87), and the model [$\chi^2(2) = 21.8$, p <. 001] correctly classed 30 out of the 32 words as having either new related or new unrelated meanings. This demonstrates that our manipulation of meaning relatedness was a successful one.

Worksheets

We then analysed participants’ learning performance, both on the online worksheets and the recall task. Worksheet results are summarised in Table 3 below.

For Worksheet 1 (definition matching), one mark was assigned for each trained word that was correctly matched to the definition of its new meaning. For Worksheets 2 (sentence writing) and 3 (story writing), participants received one mark for each trained word in the new-meaning context, regardless of how many times that word was used. Finally, for Worksheet 4 (open-ended questions), one mark was assigned for each correctly answered question about a new word referent. The analysis of Worksheet 2 results excluded three participants – one who provided semantic associates of the existing meanings of the trained words and two who created their own new meanings for these words. The analysis of Worksheet 3 results excluded one participant and 3.3% of the data from the other participants because these responses lacked in detail and may have instantiated existing meanings. We first attempted to analyse the responses using logit mixed-effects modelling, but this was not warranted – no random effects were significant (i.e., the number of correct responses did not substantially vary across subjects or items). A set of by-subjects (F_1) and by-items (F_2) ANOVAs with the factors of Meaning Type and Version was used instead. As expected, the analyses revealed no effects of Meaning Type on either of the four worksheets (all Fs < 2). The overall performance was at ceiling, most likely because participants were allowed to use the companion booklet with the paragraphs when completing all the worksheets. This confirms that the home-based training provided an opportunity to further consolidate both the new related and new unrelated meanings of words.

Table 3. Experiment 1: Mean subject percentages of correct responses for the online worksheets.

Meaning type	Worksheet 1	Worksheet 2	Worksheet 3	Worksheet 4
New related	99.4 (1.9)	99.3 (2.1)	93.8 (8.1)	96.3 (5.9)
New unrelated	98.4 (4.0)	98.9 (2.5)	91.8 (9.6)	92.8 (10.4)

Note. Standard deviations are given in the parentheses.

Recall task

For the recall task, participants received one mark for each of the five properties of the new word referents that were stated in the paragraphs. As in Rodd et al. (2012), we analysed the number of "correct responses" (i.e., responses to trained words for which at least one property was correctly recalled) and the number of correctly recalled properties for correct responses only (i.e., a maximum of five properties). Both analyses excluded one participant who correctly recalled only 7 out of the 32 new meanings of the trained words. Overall, participants' recall performance was good - the percentage of correct responses ranged (across participants) from 53 to 100% (M = 87.5, SD = 13.1). Most of the incorrect responses were null ("nothing") responses (78%), with the remaining responses being "transfer errors" (i.e., recalling a property of a different new word referent).

Numbers of correct responses were analysed using a logit Meaning Type × Version mixed-effects model that included a significant random intercept for subjects. The analysis [$\chi^2(1) = 35.7$, $p < .001$; marginal $R^2 = .11$, conditional $R^2 = .52$] showed that the percentages of correct responses were significantly higher for the words with new related (M = 94.7, SD = 6.7) than unrelated meanings (M = 80.3, SD = 21.2).

Numbers of correctly recalled properties for correct responses were analysed using a linear Meaning Type × Version mixed-effects model that included significant intercepts for subjects and items and a random slope for Meaning Type across items. The model [$\chi^2(1) = 0.1$, $p = .72$; marginal $R^2 = .03$, conditional $R^2 = .33$] showed that Meaning Type did not influence the number of recalled properties (new related meaning: $M = 2.8$, $SD = 0.5$; new unrelated meaning: $M = 2.8$, $SD = 0.6$).

Relatedness decision task

Our final aim was to establish the impact of learning new meanings on the processing of existing meanings. Three of the 20 participants were removed from all analyses of the relatedness decision task – one due to an exceptionally small number of correct responses in the recall task (22%) and the other two due to very slow and variable responses across all trials ($M = 1538.5$, $SD = 1217.8$; $M = 1100.9$, $SD = 638.4$). Analyses of both response accuracy and latency excluded trials involving trained targets for which participants could not recall any property of their new word referents (7.6% of all responses). This was necessary to ensure that we examined training effects for words with truly consolidated new meanings. For RTs, we also excluded errors (7.9% of the remaining responses) and outliers (two standard deviations above/below a participant's mean per condition; 5.1%). RTs were log-transformed to further minimise the impact of potential outliers and normalise the distribution of residuals.

Accuracy and latency data were analysed using mixed-effects models with the factors of Target Type (new related meaning, new unrelated meaning, untrained), Session (pre-training, post-training), Trial Type (“yes”, “no”), and Block (1, 2, 3)[5]. Block was included to account for potential variability in responses due to counter-balancing or target repetition. All models included significant random intercepts for subjects and items. The random slope for the Session × Trial Type interaction across subjects was significant and was included in the latency but not the accuracy model. For RT results, we report back-transformed means and confidence intervals that were estimated from the mixed-effects models using the “lmerTest” package (Kuznetsova, Brockhoff, & Christensen, 2015).

As discussed in the Introduction, our hypotheses were mainly concerned with the effects of Session on RTs. In particular, we expected slower relatedness decisions to the trained, but not to the untrained, targets following the learning of new meanings, both on “yes” and “no” trials. We also predicted this effect to be greater for the trained words with new unrelated than related meanings. For this reason, our post hoc analyses explored only those interactions that involved the effect of Session and were relevant to the hypotheses. These tests were conducted using the “phia” package (De Rosario-Martinez, 2015), and their significance thresholds were adjusted using the Bonferroni method.

Mean error rates (%) for the trained and untrained targets are illustrated in Figure 2 below. The response-accuracy model (marginal R^2 = .04, conditional R^2 =

[5] Target Type and Block were coded using Helmert contrasts. For Target Type, Contrast 1 compared both trained targets to the untrained counterparts (Untrained = -2/3, Related = 1/3, Unrelated = 1/3), and Contrast 2 compared the two types of trained targets (Untrained = 0, Related = -1/2, Unrelated = 1/2). For Block, Contrast 1 compared Block 1 to Blocks 2 and 3 (1 = 2/3, 2 = -1/3, 3 = -1/3), and Contrast 2 compared Blocks 2 and 3 (1 = 0, 2 = 1/3, 3 = -1/3). Deviation coding was used for both Session (Pre = -1/2, Post = 1/2) and Trial Type (Yes = -1/2, No = 1/2).

.23) revealed a significant Session × Trial Type interaction [$\chi^2(1) = 6.5$, $p < .01$]. Post hoc tests indicated a small but significant increase in post-training error rates on “no” trials ($M_{pre} = 4.9$, $SD = 2.2$; $M_{post} = 6.5$, $SD = 4.7$; $p < .05$), but not on “yes” trials ($M_{pre} = 10.7$, $SD = 4.2$; $M_{post} = 9.9$, $SD = 4.1$; $p = 1$). As for results that did not involve Session, there was a significant main effect of Trial Type [$\chi^2(1) = 16.4$, $p < .001$], with less accurate responses on “yes” trials involving related word pairs ($M = 10.3$, $SD = 3.8$) than on “no” trials involving unrelated word pairs ($M = 5.7$, $SD = 3.2$). There were also significant Trial Type × Target Type [$\chi^2(2) = 7.4$, $p < .05$] and Trial Type × Target Type × Block interactions [$\chi^2(4) = 10.8$, $p < .05$]. No other effects approached the significance threshold.

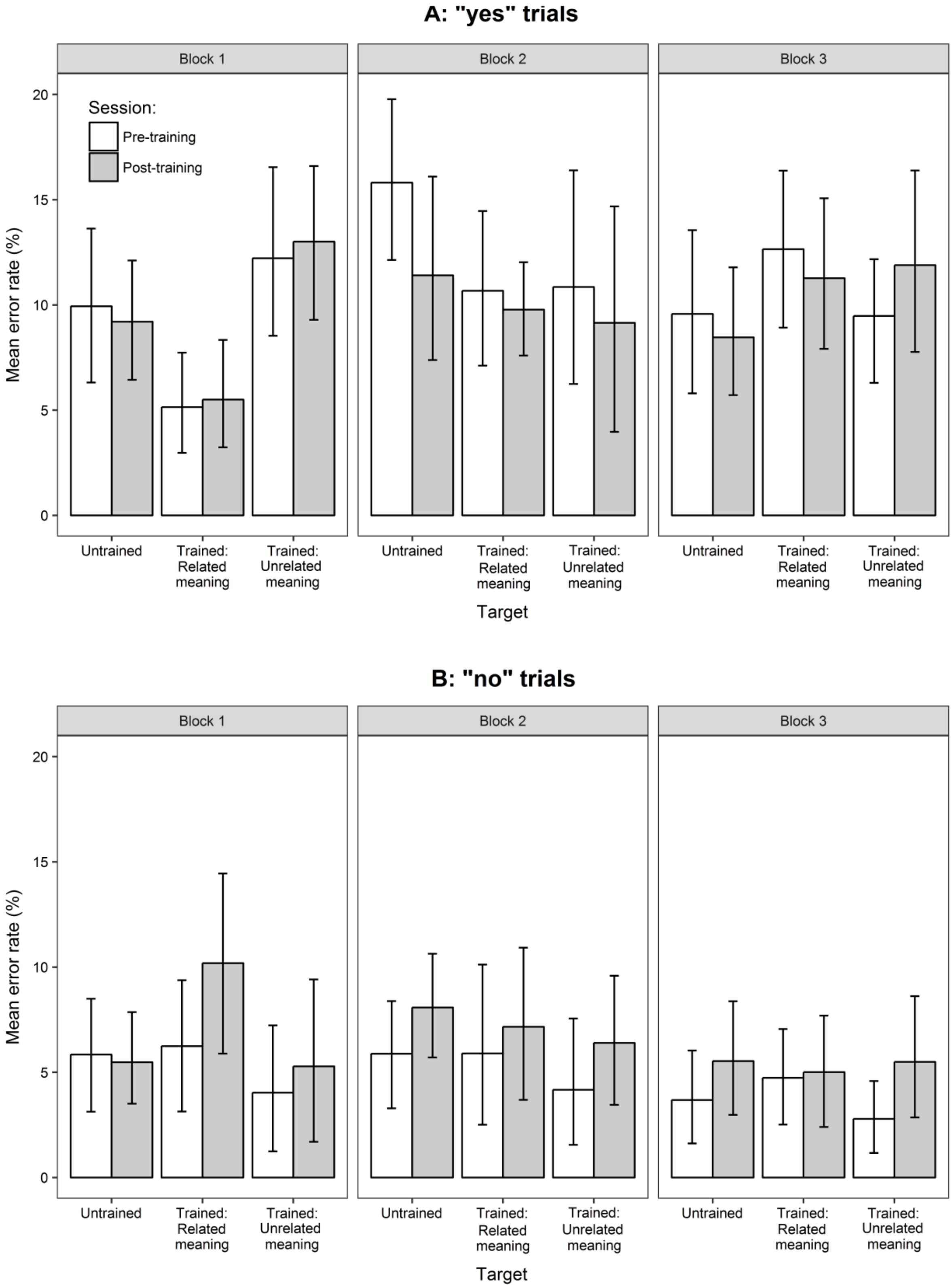

Figure 2. Experiment 1: Mean error rates across "yes" (Panel A) and "no" trials (Panel B). Error bars show 95 % confidence intervals adjusted to remove between-subjects variance (Loftus & Masson, 1994).

Mean RTs (ms) for the trained and untrained targets are illustrated in Figure 3 below. The response-latency model (marginal $R^2 = .04$, conditional $R^2 = .50$) revealed a significant Session × Block interaction [$\chi^2(2) = 17.3$, $p < .001$]. Responses were markedly slower on the post-training than the pre-training session only for Block 1 ($M_{pre} = 720.3$, 95% CIs: 663.9, 781.3; $M_{post} = 778.6$, 95% CIs: 704.4, 860.4), though this contrast was non-significant after the Bonferroni adjustment ($p = .13$)[6].

The response-latency model revealed a significant Session × Target Type interaction [$\chi^2(2) = 16.5$, $p < .001$]. We explored this result using post hoc tests that contrasted the effects of Session across pairs of target words (Related vs. Unrelated, Related vs. Untrained, Unrelated vs. Untrained). These tests showed that the slowing effect of Session was greater for the targets with new unrelated meanings ($M_{pre} = 711.4$, 95% CIs: 649, 778.6; $M_{post} = 798.5$, 95% CIs: 715.0, 891.7) than for both the targets with new related meanings ($M_{pre} = 719.1$, 95% CIs: 657.1, 787.1; $M_{post} = 769.3$, 95 %CIs: 689.0, 858.8; $p < .001$) and the untrained targets ($M_{pre} = 742.2$, 95% CIs: 677.8, 812.8; $M_{post} = 780.7$, 95% CIs: 698.9, 872.0; $p < .001$) which did not significantly differ from each other ($p = .69$). The simple effect of Session for the words with new unrelated meanings was not, however, significant after the Bonferroni adjustment ($p = .14$).

The response-latency model revealed a significant Session × Trial Type interaction [$\chi^2(1) = 8.3$, $p < .01$] that was due to an increase in post-training in RTs on “no” trials ($M_{pre} = 733.8$, 95% CIs: 672.7, 800.6; $M_{post} = 798.9$, 95% CIs: 706.0,

---

[6] Throughout this report, any results that reached the significance threshold before but not after the correction for multiple comparisons should be viewed as trends only.

904.1), though this contrast was non-significant after the Bonferroni adjustment (p = .11). There was also a significant three-way interaction between the effects of Session, Target Type, and Trial Type [$\chi^2(4) = 5.8$, $p < .05$]. Post hoc tests indicated that this was the result of an increase in post-training RTs only for the targets with new unrelated meanings on “no” trials ($p < .05$). As for results that did not involve Session, there was a significant main effect of Block [$\chi^2(1) = 16.4$, $p < .001$]. Post hoc tests showed faster responses in Block 3 (M = 720.8, 95% CIs: 665.2, 782.0) than Blocks 1 (M = 748.9, 95% CIs: 690.1, 812.5; $p < .001$) and 2 (M = 744.1, 95% CIs: 685.8, 807.4; $p < .001$), with no statistical difference between the latter (p = 1). No other effects approached the significance threshold.

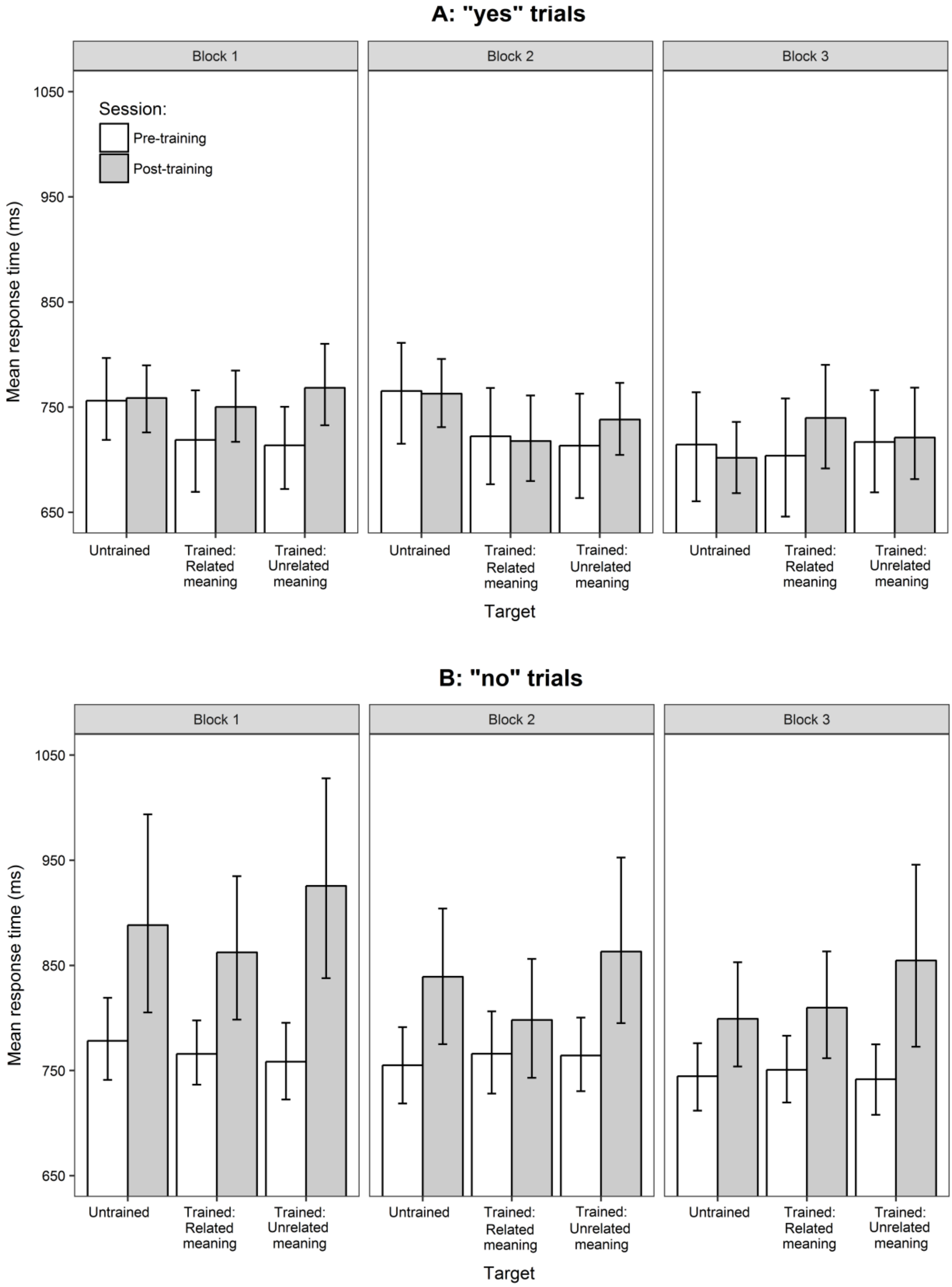

Figure 3. Experiment 1: Mean untransformed RTs across “yes” (Panel A) and “no” trials (Panel B). Error bars show 95 % confidence intervals adjusted to remove between-subjects variance.

The significant Session × Block interaction suggests that the influence of the training might have changed across the three blocks of the task. This motivated us to examine more closely participants' performance in Block 1. The rationale was that the processing of the targets in the later blocks could have been influenced by the earlier recent encounters with the words, biasing participants' interpretation towards existing meanings and reducing potential semantic competition. In contrast, the first encounter with the targets in Block 1 would represent a "purer" measure of processing speed unaffected by earlier form-to-meaning mapping. We therefore conducted another model only for RTs in Block 1. This model included the fixed effects of Session, Target Type, and Trial Type, random intercepts for subjects and items, and a random intercept for the Session × Trial Type interaction across subjects.

The model (marginal R^2 = .05, conditional R^2 = .55).revealed a Session × Trial Type interaction [$\chi^2(1) = 5.9$, $p < .05$] that was due to a significant increase in post-training RTs on "no" trials (M_{pre} = 741.0, 95% CIs: 672.8, 815.8; M_{post} = 833.9, 95% CIs: 725.8, 958.1; $p < .05$), but not on "yes" trials (M_{pre} = 707.8, 95% CIs: 649.7, 770.9; M_{post} = 734.9, 95% CIs: 672.7, 802.6; $p = .61$). There was also a significant Session × Target Type interaction [$\chi^2(2) = 16.5$, $p < .001$]. As above, we explored this result using post hoc tests that contrasted the effects of Session across pairs of target types. These analyses showed that the slowing effect of Session was greater for the targets with new unrelated meanings (M_{pre} = 711.4, 95% CIs: 649, 778.6; M_{post} = 798.5, 95% CIs: 715.0, 891.7) than for both the targets with new related meanings (M_{pre} = 719.1, 95% CIs: 657.1, 787.1; M_{post} = 769.3, 95% CIs: 689.0, 858.8; $p < .01$) and the untrained targets (M_{pre} = 742.2, 95% CIs: 677.8, 812.8; M_{post} = 780.7, 95% CIs: 698.9, 872.0; $p < .001$) which did not significantly differ from each other (p = .35). The simple effect of Session was significant only for the trained words with new unrelated meanings (p < .05). No other effects approached the significance threshold.

Discussion

Experiment 1 showed that participants consolidated many of the new meanings over the course of our intensive training. Their ability to recall the meanings was superior for meanings that were semantically related to the existing meanings than for unrelated meanings. Notably, meaning relatedness facilitated the likelihood of access to the semantic representations for the newly-learnt meanings but not the amount of information within these representations. As in Rodd et al. (2012), participants recalled as many semantic features for related word referents as they did for the unrelated counterparts, whenever they correctly recalled any information about the new meanings. Thus, it appears that the overlap in semantic features between the new and existing meanings acts as a cue during the learning and/or retrieval of new meanings, leading to better recall for related meanings. However, this overlap does not seem to determine the robustness or richness of the semantic representations as typically defined in terms of the number of semantic features (e.g., McRae, 2004; Yap, Tan, Pexman, & Hargreaves, 2011).

With regard to the impact of learning new meanings, the experiment showed that the meanings were integrated into the mental lexicon, such that they affected performance in the online speeded task. Participants' processing of existing meanings slowed after the consolidation, but only in certain conditions. The analysis involving all experimental blocks revealed that the training slowed responses to words with new unrelated meanings but not the related counterparts. There was also an indication that the overall impact of training decreased as the task progressed, such that it was mainly observed only in the first block. Further analysis focusing on responses in Block 1 revealed that the training effect was restricted to words with new unrelated meanings on “no” trials. Although this seems to suggest that newly-learnt meanings slowed the processing of existing meanings, and that this interference effect was sensitive to the semantic relatedness between the two meanings, caution should be applied when interpreting results from “no” trials on their own. Since we cannot confirm which meaning participants selected on these trials (as both would yield a correct response), the training effect could indicate difficulties in access to existing meanings due to interference from new meanings and/or difficulties in access to new meanings. We do, however, point out that there was also a numerical albeit non-significant training effect for “yes” trials and for words with new related meanings (see Figure 3 above), which addresses to some extent the issue with “no” trials. We offer some explanations as to why these trends did not reach the significance threshold below.

While we tripled the number of semantically related and unrelated probe words (i.e., “yes” and “no” responses) to compensate for typically low numbers of participants and items in studies using artificial language learning paradigms, the results clearly demonstrated that this approach did not benefit detection power. First, we found that the overall performance became faster towards the end of the task, most likely due to practice involved in making multiple relatedness decisions to the same targets. Second, the results showed a gradual decrease in the training effect over the course of the task, particularly for “yes” trials, such that participants’ processing of existing meanings on the post-training session appeared slower only during Block 1 (i.e., during the first encounter with the trained words). Thus, it appears that the repetition of the targets in the existing-meaning context modulated the training effect.

We suggest that having disambiguated a trained word towards its existing meaning on the first “yes” trial facilitated the processing of that meaning on the subsequent two trials, eliminating the otherwise slowing effect of learning. Strong support for this account comes from recent word-meaning priming studies (Rodd, Lopez Curtin, Kirsch, Millar, & Davis, 2013; Rodd et al., 2016) which have demonstrated that even a single recent encounter with a particular meaning of an ambiguous word can temporarily bias future form-to-meaning mappings in favour of that meaning. However, it is also possible that participants actively suppressed new meanings during the later encounters with the trained words after having realised that none of the probes instantiated those meanings. Such a task strategy would also bias participants’ comprehension and reduce the training effect in Blocks 2 and 3. Although we cannot establish whether it was strategic processing or more implicit word-meaning priming that was in play in the current experiment, it is clear that the results were influenced by target-word repetition. In order to address these issues, we designed and conducted Experiment 2.

Experiment 2

Experiment 2 was largely similar to Experiment 1, but it involved a few changes that were designed to address issues raised from Experiment 1. First, the target words in Experiment 2 were presented with two, rather than six, probe words – one related probe that instantiated the existing meaning and one unrelated probe.

Contrasting the effects of consolidation on “yes” and “no” trials was critical to the design of the study in understanding the locus of the effects (see General Discussion). Thus, although some (minor) repetition of the target remained, we did account for it in the analysis. Second, in order to compensate for the reduction in the number of trials per item, we created new sets of target-probe word pairs that were well-matched on 13 psycholinguistic variables, rather than word-form frequency and length alone. Third, we used a faster variant of the relatedness decision task, such that the target and the probe were presented for 200 ms and 500 ms, respectively. These changes aimed to reduce the variability in response latencies that was observed in Experiment 1, particularly for “no” trials. Finally, we tested a larger group of participants to further increase detection power.

Method

Participants

Thirty students and members of staff [23 females, aged 20-35 (*M* = 26.6, *SD* = 5.3)] from the University of Leeds participated in the experiment in exchange for a £20 voucher. As in Experiment 1, participants were monolingual native speakers of British English with no known history of language-/vision-related difficulties/disorders. All were right-handed, as confirmed using the Briggs-Nebes (1975) modified version of Annett’s (1967) handedness inventory. The experiment received ethical approval from the School of Psychology, University of Leeds Ethics Committee.

Materials

The trained words, paragraphs, and worksheets were the same as those in Experiment 1. For the relatedness decision task, we used a new set of 32 untrained targets that were matched to the trained counterparts (all *t*s < 1) with respect to 13 lexical and semantic variables (see target properties in Table 4 below). All target words had noun or noun-verb interpretations (but were used as nouns in the task) and a single meaning in the Wordsmyth Dictionary (Parks et al., 1998).

Table 4. Experiment 2: Descriptive statistics of lexical and semantic properties of the target words.

Variable	Trained targets	Untrained targets
Example	“sip”	“cod”
Letters	4.4 (1.0)	4.4 (1.1)
Phonemes	3.5 (0.8)	3.4 (0.9)
Syllables	1.1 (0.3)	1.2 (0.4)
Raw frequency	17.1 (20.0)	17.2 (15.1)
Log frequency	1.0 (0.5)	1.1 (0.4)
Orthographic neighbours	7.1 (5.4)	7.1 (6.8)
Log bigram frequency	2.8 (0.5)	2.7 (0.5)
Subjective familiarity	5.0 (0.5)	5.1 (0.5)
Word senses	4.8 (2.0)	4.7 (2.1)
Semantic diversity	1.6 (0.2)	1.5 (0.2)
Concreteness	5.5 (0.7)	5.6 (0.8)
Imageability	5.6 (0.7)	5.7 (0.7)
Age of acquisition	6.2 (1.4)	6.2 (1.9)

Note. Standard deviations are given in the parentheses. Word-form frequency, bigram frequency, and the number of orthographic neighbours come from the British National Corpus (2007). The number of word senses comes from the Wordsmyth Dictionary (Parks et al., 1998). Semantic diversity norms come from Hoffman, Lambon Ralph, and Rogers (2013). Concreteness, imageability, and subjective familiarity ratings come from the MRC Psycholinguistic Database (Coltheart, 1981). Age-of-acquisition ratings come from Kuperman et al. (2012).

New, well-matched sets of target-probe word pairs were created. Each target was paired with a single related and unrelated probe. As in Experiment 1, the related probes instantiated the existing but not the new meaning. All the probe words were nouns with only one meaning in the Wordsmyth Dictionary (Parks et al., 1998), and their numerous word properties (see Table 5 below) were closely matched between the word pairs involving the trained and untrained targets (all *F*s < 1). Prior to the experiment, 30 monolingual native speakers of British English [15 females; aged 18-38 (*M* = 29.9, *SD* = 5.7)] rated target-probe relatedness on a 7-point scale (where 1 denoted “highly unrelated” and 7 denoted “highly related”). This pre-test confirmed that the related and unrelated target-target pairs were considered as such, and that the trained (related pairs: *M* = 6.2, *SD* = 0.3; unrelated pairs: *M* = 1.9, *SD* = 0.4) and untrained targets (related pairs: *M* = 6.2, *SD* = 0.3; unrelated pairs: *M* = 1.9, *SD* = 0.4) did not significantly differ with respect to the degree of semantic relatedness/unrelatedness (both *t*s < 1). All the target-probe word pairs used in Experiment 2 are presented in Appendix 3.

Table 5. Experiment 2: Descriptive statistics of lexical and semantic properties of the probe words.

Variable	Trained targets		Untrained targets	
	Related probes	Unrelated probes	Related probes	Unrelated probes
Example	“sip-juice”	“sip-golf”	“cod-eel”	“cod-toy”
Letters	5.0 (1.2)	5.0 (1.1)	4.9 (1.2)	5.0 (1.2)
Phonemes	4.1 (1.3)	4.2 (1.2)	4.3 (1.3)	4.2 (1.3)
Syllables	1.6 (0.7)	1.5 (0.6)	1.6 (0.7)	1.5 (0.7)
Raw frequency	21.2 (19.8)	21.4 (16.9)	21.1 (21.9)	21.4 (20.3)
Log frequency	1.1 (0.5)	1.2 (0.4)	1.2 (0.4)	1.2 (0.4)
Orthographic neighbours	3.5 (3.8)	3.6 (3.7)	3.7 (4.7)	3.6 (4.2)
Log bigram frequency	2.5 (0.4)	2.6 (0.5)	2.6 (0.6)	2.7 (0.5)
Subjective familiarity	5.3 (0.6)	5.2 (0.6)	5.1 (0.6)	5.2 (0.6)
Word senses	3.8 (1.7)	3.8 (2.0)	3.7 (2.4)	3.8 (2.2)
Semantic diversity	1.5 (0.2)	1.5 (0.2)	1.5 (0.2)	1.5 (0.2)
Concreteness	5.7 (0.5)	5.7 (0.8)	5.8 (0.6)	5.8 (0.5)
Imageability	5.6 (0.6)	5.7 (0.8)	5.8 (0.4)	5.7 (0.7)
Age of acquisition	6.1 (1.8)	6.1 (1.9)	6.0 (1.7)	6.0 (1.9)
Target-probe relatedness	6.2 (0.3)	1.9 (0.4)	6.2 (0.3)	1.9 (0.4)

Note. Standard deviations are given in the parentheses. Information on the different variables can be found in the note for Table 1.

Procedure

The general procedure for the worksheets, paragraph reading, and recall was largely the same as in Experiment 1, with the following changes. First, all worksheets in Experiment 2 were completed during the home-based sessions on Days 2-4 (for an overview, see Figure 4). Second, we removed the meaning-relatedness rating task as there was no need to examine the meaning-relatedness manipulation for the same items again. Third, the inter-trial interval in the paragraph reading and recall tasks was shortened to 100 ms (as opposed to 500 ms in Experiment 1) as there was no need for participants to rest between the trials of these non-speeded tasks. For the paragraph reading task, we added 1000 ms feedback on participants’ answers to the reading comprehension questions. Finally, all the lab-based tasks were programmed in EPrime 2.0 (Schneider, Eschman, & Zuccolotto, 2010).

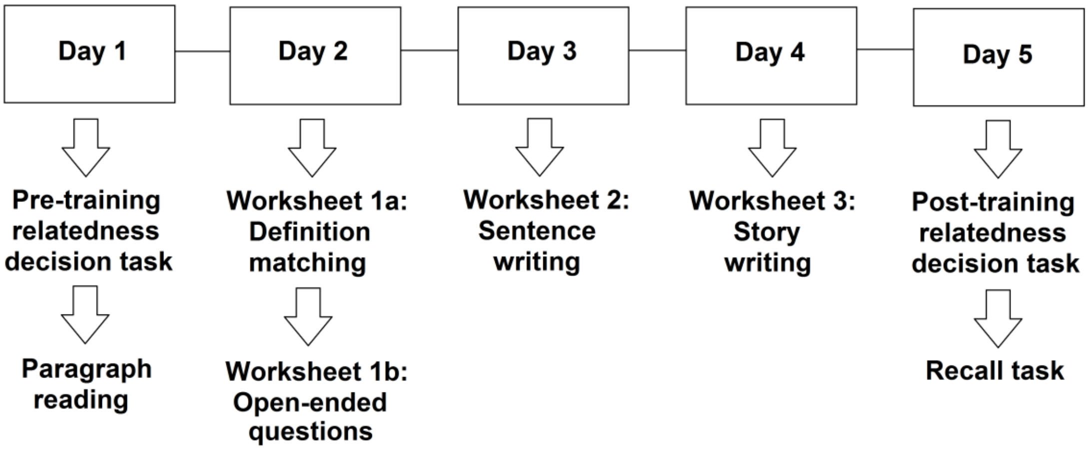

Figure 4. Overview of Experiment 2.

We also made some changes to the relatedness decision task. The new stimulus list was divided into two blocks whose order was counterbalanced across participants. One block included 64 related pairs involving 32 trained and 32 untrained targets and 64 unrelated pairs serving as fillers (which were excluded from analyses). The other block included 64 unrelated pairs involving 32 trained and 32 untrained targets and 64 related fillers. This blocked design allowed for control over target repetition, which seems to have obscured the training effect in Experiment 1, so that we were able to determine whether responses to a target word on related trials had an impact on subsequent responses on unrelated trials, and vice versa. None of the targets appeared more than once within the same block, and the fillers did not include any of the words used in the experimental stimulus list. The order of trials in each block was pseudo-randomised, such that no more than three “yes”/”no” trials appeared consecutively. A practice block, preceding the experimental blocks, included 20, as opposed to 10, trials. There were two one-minute breaks – one after the practice block and one after the first experimental block. Each experimental block began with eight fillers (excluded from analyses) to help participants get back to the habit of quick responding following a break. Trials began with a 500 ms fixation cross. After a delay of 100 ms, targets were presented for 200 ms followed by probes presented for 500 ms, with a delay of 50 ms in between. Participants were allowed an additional 1500 ms to respond. As soon as a response was made or at the end of the 1500 ms, there was a 100 ms delay before the next trial began. Participants could make relatedness decisions as soon as the probe appeared, but they had to respond within 1500 ms. All other procedures were the same as in Experiment 1.

Results

Worksheets

Performance on the worksheets and the recall task was analysed similarly to Experiment 1. For Worksheet 2 (sentence writing), we excluded 10 participants who provided definitions of the new word referents, rather than their own example sentences. For Worksheet 3 (story writing), we excluded 3.2% of responses that lacked detail and may have instantiated the existing meanings. As in Experiment 1, the analyses revealed no effects of Meaning Type (related vs. unrelated) on either of the four worksheets [all *F*s < 1, see Table 6 below].

Table 6. Experiment 2: Mean percentages of correct responses for the online worksheets.

Condition	Worksheet 1a	Worksheet 1b	Worksheet 2	Worksheet 3
New related meaning	99.4 (2.5)	99.1 (2.2)	98.7 (3.5)	98.7 (3.5)
New unrelated meaning	98.3 (3.6)	99.6 (2.3)	98.3 (3.7)	98.6 (2.8)

Note. Standard deviations are given in the parentheses.

Recall task

Overall, participants' recall performance was good - the percentage of correct responses ranged (across participants) from 50 to 100% (*M* = 89.9, *SD* = 15.1). Most of the incorrect responses were null responses (64%), with the remaining responses being transfer errors (i.e., recalling a property of a different new word referent). The percentage of correct responses was significantly higher for the words with new related (*M* = 94.4, *SD* = 12.3) than unrelated meanings [*M* = 84.4, *SD* = 19.0; $\chi^2(1) = 33.1$, $p < .001$; marginal $R^2 = .07$, conditional $R^2 = .54$]. As in Experiment 1, Meaning Type did not have a significant effect on the numbers of correctly recalled properties [related meaning: *M* = 3.7, *SD* = 0.6; unrelated meaning: *M* = 3.8, *SD* = 0.6; $\chi^2(1) = 0.8$, $p = .37$; marginal $R^2 = .01$, conditional $R^2 = .38$]. This provides further evidence that although the overlap in semantic features between the new and existing meanings acts as a cue during the learning and/or retrieval of new meanings, it does not determine the robustness or richness of their semantic representations.

Relatedness decision task

Two of the 30 participants were removed from all analyses of the relatedness decision task – one due to a small number of correct responses in the recall task (50 %) and the other due to relatively slow responses across all trials (*M* = 870.0 ms, *SD* = 129.0). As in Experiment 1, we excluded trials involving the trained targets for which participants could not recall any property of their new word referents (4.5% of all responses). For RTs, analyses also excluded errors (4.3% of the remaining response) and outliers (two standard deviations above/ below a participant's mean per condition; 4.1%). RTs were log-transformed to normalise the residual distribution.

The first set of analyses combined the trained targets across the levels of Meaning Type (new related/unrelated meaning) and compared them to the untrained targets. The rationale was that, unlike Experiment 1, Experiment 2 involved unequal numbers of targets (16 trained words with new related/unrelated meanings and 32 untrained words), thus biasing direct comparisons across the three target types. Accuracy and latency data were analysed using mixed-effects models with the factors of Session (pre-training, post-training), Target Type (trained, untrained), Trial Type ("yes", "no"), and Block (1, 2)[7]. All models included random intercepts for subjects and items. The random slope for the Session × Trial Type interaction across subjects and the random slope for Session across items were significant and included in the response-latency but not the response-accuracy model.

Mean error rates (%) for the trained and untrained targets are illustrated in Figure 5 below. The response-accuracy model (marginal R^2 = .02, conditional R^2 =

[7] There were not any effects of Block in Experiment 2, neither in the latency nor the accuracy data.

.36) revealed a Session × Trial Type interaction [$\chi^2(1) = 6.7$, $p < .01$] that was due to a significant increase in post-training error rates on “no” trials ($M_{pre} = 3.3$, $SD = 3.0$; $M_{post} = 4.7$, $SD = 4.7$; $p < .05$), but not on “yes” trials ($M_{pre} = 5.3$, $SD = 5.0$; $M_{post} = 4.2$, $SD = 3.8$; $p = .27$). There was also a significant Session × Trial Type × Target Type interaction [$\chi^2(1) = 3.9$, $p < .05$]. Post hoc tests indicated that the interaction concerned the trained targets only. Following the training, error rates for these words were lower on “yes” trials ($M_{pre} = 6.8$, $SD = 7.1$; $M_{post} = 3.9$, $SD = 3.8$; $p < .05$), but not on “no” trials ($M_{pre} = 3.4$, $SD = 4.7$; $M_{post} = 5.2$, $SD = 6.2$; $p = .16$). No other effects approached the significance threshold.

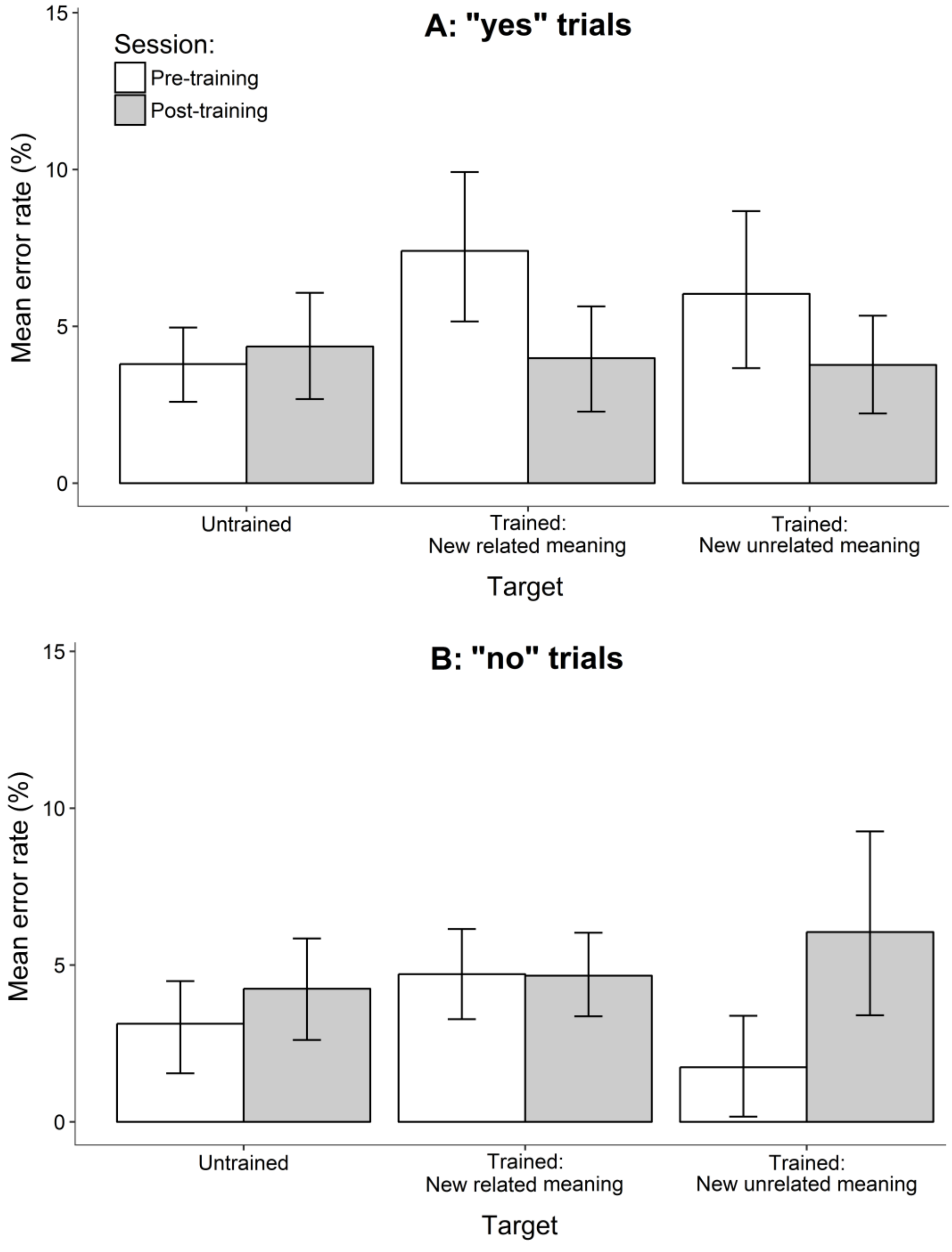

Figure 5. Experiment 2: Mean error rates across “yes” (Panel A) and “no” trials (Panel B). Error bars show 95 % confidence intervals adjusted to remove between-subjects variance.

Mean RTs (ms) for the trained and untrained targets are illustrated in Figure 6 below. The response-latency model (marginal R^2 = .09, conditional R^2 = .54) revealed a significant Session × Target Type interaction [$\chi^2(1) = 31.6$, $p < .001$]. Post hoc tests showed a significant increase in post-training RTs for the trained (M_{pre} = 598.4, 95% CIs: 570.0, 628.4; M_{post} = 639.7, 95% CIs: 605.6, 675.6; $p < .001$) but not untrained targets (M_{pre} = 581.3, 95% CIs: 553.7, 610.4; M_{post} = 587.5, 95% CIs: 556.3, 620.6; $p = 1$). There was a significant main effect of Trial Type [$\chi^2(1) = 25.3$, $p < .001$], with slower responses on “no” (M = 632.9, 95% CIs: 598.8, 668.7) than “yes” trials (M = 571.5, 95% CIs: 545.9, 598.3). Responses were also slower on the post-training (M = 613.1, 95% CIs: 581.0, 646.8) than the pre-training session (M = 589.8, 95% CIs: 562.6, 618.4), although this effect of Session only approached the significance threshold [$\chi^2(1) = 3.3$, $p = .07$]. Finally, there was a significant main effect of Target Type [$\chi^2(1) = 27.3$, $p < .001$], with slower responses to the trained (M = 618.7, 95% CIs: 589.4, 649.5) than untrained targets (M = 584.4, 95% CIs: 556.7, 613.6). No other effects approached the significance threshold.

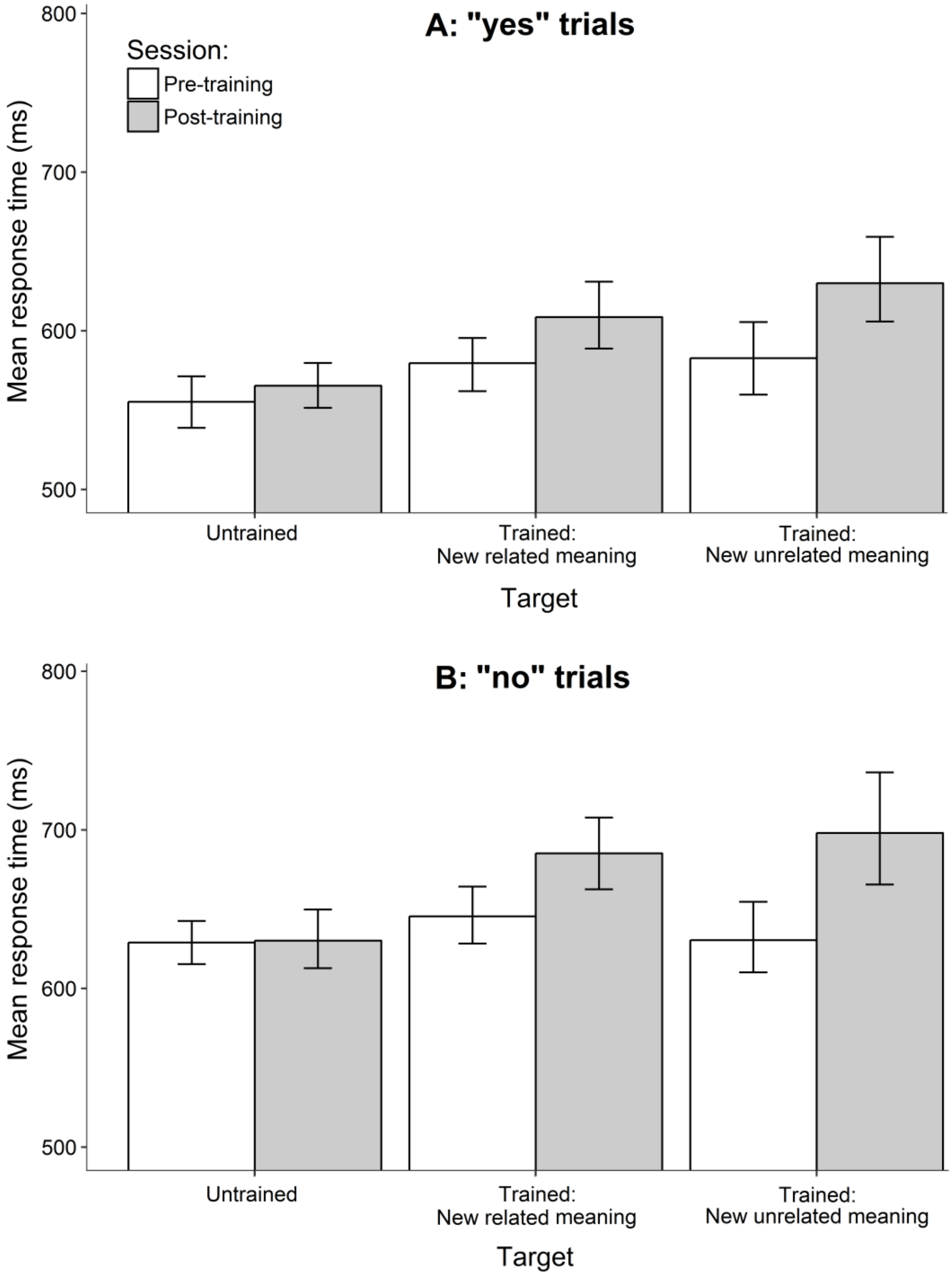

Figure 6. Experiment 2: Mean untransformed RTs across “yes” (Panel A) and “no” trials (Panel B). Error bars show 95 % confidence intervals adjusted to remove between-subjects variance.

These analyses showed that having learnt new meanings slowed participants’ responses to previously unambiguous words. To examine the role of the semantic relatedness between the existing and the new meaning, the second set of analyses excluded the untrained targets and directly compared the two types of trained targets. These response-accuracy and response-latency models included the same fixed effects as those in the models above, except that Target Type was replaced with Meaning Type (related vs. unrelated). With respect to random effects, both models included random intercepts for subjects and items. The response-latency model additionally included random slopes for the Session × Trial Type and Meaning Type × Trial Type interactions across subjects and a random slope for Session across items.

The response-accuracy model (marginal R^2 = .06, conditional R^2 = .45) revealed only a significant Session × Trial Type interaction [$\chi^2(1) = 11.4$, $p < .001$]. Post hoc tests indicated that following the training, error rates decreased on “yes” trials (M_{pre} = 6.8, SD = 7.1; M_{post} = 3.9, SD = 3.8; $p < .05$) but increased on “no” trials (M_{pre} = 3.4, SD = 4.7; M_{post} = 5.2, SD = 6.2; $p < .05$).

In contrast, the response-latency model (marginal R^2 = .07, conditional R^2 = .54) revealed a significant Session × Meaning Type interaction [$\chi^2(1) = 5.6$, $p < .05$]. Post hoc tests showed that the simple effect of Session was significant for both the words with new unrelated (M_{pre} = 595.0, 95% CIs: 565.3, 626.3; M_{post} = 645.1, 95% CIs: 609.1, 683.3; $p < .001$) and new related meanings (M_{pre} = 602.4, 95% CIs: 573.2, 633.3; M_{post} = 635.9, 95% CIs: 600.9, 672.8; $p < .01$), but was significantly greater for the former (as indicated by the interaction). There was a significant main effect of Trial Type [$\chi^2(1) = 15.0$, $p < .001$], with faster relatedness decisions on “yes” (M = 591.0, 95% CIs: 562.5, 620.9) than “no” trials (M = 648.9, 95% CIs: 610.8, 689.5). Responses were also slower on the post-training (M = 640.5, 95% CIs: 605.3, 677.6) than the pre-training session (M = 598.7, 95% CIs: 569.6, 629.4), and this main effect of Session was significant [$\chi^2(1) = 8.5$, $p < .01$] All other effects did not approach the significance threshold.

Discussion

Experiment 2 showed that consolidation of new meanings slowed participants' comprehension of existing meanings. This effect, which was observed on both "yes" and "no" trials, was greater for meanings that were unrelated to the existing meanings of the words than the related counterparts. Critically, there was no indication that the training effect extended to the untrained words, or that it was modulated by the minimal target-word repetition employed in the current experiment. Overall, the results of Experiment 2 strengthen the trends observed for Block 1 in Experiment 1, indicating that relatedness in meaning affects both the consolidation and processing of new meanings for familiar words.

Note, however, that Experiment 2 showed a speed-accuracy trade-off for the trained targets on "yes" trials. There was a 3% decrease in error rates and a 36 ms increase in RTs in that condition on the post-training session, which could reflect a shifted response criterion for related target-probe word pairs after the training. Although this trade-off may have contributed to some extent to our results, we do not think that it alone constitutes an explanation for the observed training effect (i.e., slower comprehension after learning a new word meaning). If we assumed that the slowing on "yes" trials was primarily driven by the trade-off, it would be difficult to explain why the same degree of slowing was observed on "no" trials where no trade-off occurred. It would also be difficult to explain why the slowing was greater for new unrelated than related meanings, both on "yes" and "no" trials. Thus, on the whole, the results indicate that the training effect was semantic in nature; it was sensitive to the semantic relationship between the new and the old meaning, and arose across all the conditions, regardless of whether there may have been some degree of speed-accuracy trade-off or not.

General Discussion

Recent studies have shown that the ability to learn new linguistic information continues to be important throughout adult life, hence research into learning artificial vocabulary has great potential to complement our understanding of both memory and language processes (for a review, see Davis & Gaskell, 2009). The current study focused on learning new meanings for familiar words - a frequent and natural language process that has resulted in the ubiquity of semantic ambiguity in many languages. While previous studies have shown that adults are skilled at learning new meanings (Fang et al., 2017; Hulme, Barsky, & Rodd, 2018; Rodd et al., 2012) or working out new senses of words (Clark & Gerrig, 1983; Frisson & Pickering, 2007; Murphy, 2006), little is known as to how successful consolidation of new meanings affects the comprehension of existing meanings. The present study addressed this novel question by training adults on new, fictitious meanings for known words and examining the impact of such training on their ability to understand the words in their original meanings.

Experiments 1 and 2 showed that learning new meanings influenced the processing of previously unambiguous words in a semantically engaging online task, indicating that the meanings had been successfully “lexicalised” (Gaskell & Dumay, 2003) or “engaged” within the mental lexicon (Leach & Samuel, 2007). As expected, consolidation of new meanings slowed the comprehension of existing meanings, mirroring the ambiguity disadvantage effect observed in studies using existing ambiguous words (e.g., Duffy et al., 1988; Gottlob et al., 1999; Hoffman & Woollams, 2015). We interpret this finding in line with the semantic competition account that comes from connectionist models of ambiguity processing (Armstrong & Plaut, 2008; Kawamoto, 1993; Rodd et al., 2004). Slower responses on the post-training session indicate competition from the features of the newly-learnt meaning when trying to access the features of the existing meaning. This is because the trained targets had acquired inconsistent form-to-meaning mappings over the course of the study, such that both meanings were initially activated (to some extent) upon reading the words in the relatedness decision task. It appears that new meanings (once integrated in the mental lexicon through extensive training and offline consolidation) can give rise to competition during the semantic activation process, just like words with multiple familiar meanings. Here, we show that this competition hinders participants' comprehension of well-established, dominant meanings, or their ability to swiftly access and select those meanings in the absence of contextual bias.

The current study, and in particular Experiment 2, further delineated this interference effect by demonstrating that it is modulated by the degree of semantic relatedness between the new and the existing meaning. Although having learnt a new meaning generally slowed the processing of the existing, dominant meaning, this effect was smaller when the two meanings were semantically related. In other words, our results show that the greater the relatedness between word meanings, the smaller the competition. Interestingly, we also observed a robust relatedness effect in the recall performance. As in Rodd et al. (2012), both Experiments 1 and 2 showed that participants' ability to recall new meanings was significantly better for meanings that were semantically related to well-established meanings. Overall, then, the current study shows that meaning relatedness is an important property of ambiguous words that has a pervasive impact on both learning and processing meanings of words. This finding is particularly relevant to the ambiguity literature that has to date produced mixed evidence for the relatedness effect (for a recent review, see Eddington & Tokowicz, 2015). Our study demonstrates the effect in an artificial language learning paradigm in which the same previously unambiguous words were paired (across participants) with new related or unrelated meanings. The advantage of this approach is that it allows for accurate manipulation of the polysemous or the homonymous status of words while controlling their other properties that may act as confounds in between-items studies using existing ambiguous words.

The finding that meaning relatedness modulates the degree of semantic competition has also important implications for PDP models that recognised the role of that property in ambiguity representation and processing, such as the ones proposed by Armstrong and Plaut (2008) and Rodd et al. (2004). While both models suggest that consolidation of new unrelated meanings should slow the comprehension of existing meanings, they make different predictions regarding the effect for new related meanings/senses. Consistent with our results, the model by Rodd et al. (2004) predicts that competition produced by new related meanings should be smaller than that produced by new unrelated meanings because the semantic features of the former overlap with those of existing meanings. Rodd et al. (2004) suggest that polysemes have separate but overlapping semantic representations, and that this results in reduced competition that involves only those features that are unique to the different word referents (see also Brocher, Foraker, & Koenig, 2016).

In contrast, the model by Armstrong and Plaut (2008) predicts that learning new related meanings would not slow the comprehension of existing meanings at all. According to their model, polysemes also have separate overlapping semantic representations, but any competition between the representations is cancelled out by a processing benefit at the earlier stages of word processing. Studies of ambiguity processing have shown that polysemy facilitates word recognition (e.g., Armstrong & Plaut, 2016; Klepousniotou & Baum, 2007; Rodd et al., 2002). It is on this basis that Armstrong and Plaut (2008) predict that the polysemy advantage during orthographic processing is equal to the polysemy disadvantage during semantic processing, such that the former eliminates the latter in tasks that require both processing stages to be completed (e.g., the relatedness decision task). However, while Rodd et al.'s (2012) lexical decision task showed that the learning of new related meanings can indeed benefit word recognition, our findings, from a semantically engaging task involving the same stimulus words, show that the learning still slows comprehension (i.e., access and selection of a particular word meaning). It appears that the polysemy advantage during orthographic processing does not entirely cancel out the polysemy disadvantage during semantic processing. Thus, even at the relatively early stages of meaning consolidation, new related meanings of irregular polysemes can still produce some degree of competition when the task requires meaning selection.

It should be noted that the implications of our work on the role of meaning relatedness are restricted to representational and processing differences between homonymy and irregular polysemy. The new related meanings in the current study were designed to imitate sense extension typical of irregular rather than regular polysemy. The meanings were loosely related to the existing meanings through a single semantic feature (e.g., physical property, function), and the relation between them was unpredictable and idiosyncratic, such that participants could not derive the new meanings from the existing ones based on their knowledge of words and their meanings. Thus, while our findings contrasting homonymy with irregular polysemy contribute to the literature on the relatedness effect, they make no prediction with respect to learning new word senses that follow the rules of sense extension characteristic of metonymic/regular polysemy, such as the instrument for action (e.g., “shovel”) and container for contents alternations (e.g., “pot”). Studies have shown that both adults (Clark & Gerrig, 1983; Frisson & Pickering, 2007; Murphy, 2006) and four-year old children (Srinivasan & Snedeker, 2011, 2014) have little difficulty understanding these senses in context. Furthermore, there is notable evidence that metonyms, whose senses share a large number of semantic features, have a single semantic representation, and may therefore escape competition at the semantic level (Frazier & Rayner, 1990; Frisson & Pickering, 1999; Klepousniotou, 2002; Klepousniotou et al., 2008). It is therefore reasonable to assume that new metonymic senses do not require explicit learning or integration into the mental lexicon but can be derived online via a rule of sense extension.

Alternative interpretations of the present findings, such as proposals that the effect of consolidation may not exclusively lie in semantic processing, do not seem plausible. For example, Pexman et al. (2004) argue that relatedness decisions to ambiguous words (e.g., “electric/football fan”) may be slower than those to unambiguous counterparts because the former trigger conflicting responses on “yes” trials (e.g., “sport”), making participants take additional time to decide which meaning of an ambiguous word should serve as response input. However, our results showed that not only did the training slow relatedness decisions on “yes” trials (e.g., “sip-juice”) that may involve such response-conflict resolution, but also on “no” trials (e.g., "sip-golf") where the new and the existing meaning triggered a single ("no") response. If the effect of learning new meanings were due to decision making during the response-selection phase, we would not expect to find it on "no" trials that are free of response conflict. Thus, Pexman et al.'s (2004) account fails to explain why consolidation of newly-acquired meanings would slow the processing of well-established meanings.

We also do not think that the slower performance on the post-training session was due to a task strategy whereby participants took additional time to ensure that the probe words were not related to new meanings (on both "yes" and "no" trials). Although this interpretation would be in line with Hino et al.'s proposal (2006) that ambiguity slows processing only when a task-relevant response requires analysis of the multiple word meanings, there are three issues with the idea that some "checking" process constitutes a complete explanation of the current findings. First, the results demonstrate that the slowing effect of learning was smaller for new related meanings, consistent with the evidence that competition between familiar word meanings is modulated by the degree of overlap in their semantic features (Armstrong & Plaut, 2008; Brocher et al., 2016; Klepousniotou et al., 2008; Rodd et al., 2004). The fact that the training effect, like the ambiguity effect in natural language, is sensitive to meaning relatedness suggests that the processing cost lies in semantic rather than task-specific decision-making processes.

Second, the results show that the slowing effect of learning was smaller for new related than unrelated meanings, even though the two did not differ in how well they were remembered. It will be recalled that our analyses of relatedness decisions included only those words for which participants could recall their new meanings, and that in those instances participants recalled as many semantic features for related meanings as they did for the unrelated counterparts. This proves problematic for the idea that the training effect is due to retrieval of additional semantic features of the target's word referents gained after the learning and comparing them to features of the probe's word referents. If such an explicit search and analysis of features was involved, we would expect new related and unrelated meanings, with comparable numbers of additional semantic features, to slow post-training responses to the same extent, which was not the case.

Third, if the ambiguity disadvantage, on the whole, was purely a task artefact, as Hino et al. (2006) and Pexman et al. (2004) suggest, it is difficult to understand why it repeatedly appeared across a number of tasks of varying response-selection demands. Competitive processes involved in understanding semantically ambiguous words have been observed in tasks requiring semantic relatedness (e.g., Gottlob et al., 1999) and categorisation decisions (e.g., Jager & Cleland, 2015), semantically primed (e.g., Balota & Paul, 1996) and unprimed lexical decisions (e.g., Rodd et al., 2002), sensicality judgements (e.g., Klepousniotou et al., 2008), and even sentence-reading tasks that do not require any response or decision (e.g., Duffy et al., 1988). Consistent with this research, the present study provides novel evidence from a language learning paradigm that supports the postulate of semantic competition in connectionist models and further challenges decision-making accounts of ambiguity effects (see also Armstrong & Plaut, 2016). We do, however, acknowledge that decision making and other conscious strategic processes have a pervasive impact on language comprehension in experimental settings. We trust future studies of learning new meanings (and ambiguity for that matter) will employ tasks (such as masked priming or sentence reading) that appear less sensitive to these factors, and therefore be able to resolve these issues.

Finally, it is important to note that competition from newly-acquired meanings bears a striking resemblance to the lexical competition reported in studies of word learning (e.g., Bowers, Davis, & Hanley, 2005; Gaskell & Dumay, 2003). The general finding of these studies is that consolidation of new word forms (e.g., “cathedruke”) slows the recognition of known neighbours (e.g., “cathedral”), in either the spoken or the written modality. Although there are differences between learning new meanings for familiar words and learning new words, it appears that integration of both types of information comes at a cost because of the way lexical-semantic representations are formed and accessed.

The implication is that, just like lexical competition has served as an index of consolidation of new word forms, semantic competition, documented in this study, can serve as an index of consolidation of new word meanings. Thus, our work provides researchers with a novel paradigm to address important questions about meaning consolidation, such as the nature of training (e.g., learning from naturalistic, semantically diverse context vs. dictionary definitions) and differences in learning performance across the lifespan. Future studies should in particular investigate the role of sleep and the time-course of meaning consolidation to better understand the degree of offline consolidation that is necessary to produce competition between the new and well-known meanings of words. It is also important to examine the time-frame of this competition effect. Experiment 1 suggested that multiple recent exposures to words in the well-known meaning can negate the effect. However, it is unclear whether this is an indication of how short-lived and weak competitive processes are in artificial language learning studies, or whether it is due to a temporary boost in access to the well-known meaning, similar to that observed for existing ambiguous words (see Rodd et al., 2013, 2016). Studies on the time-frame of competition would also help to determine the extent to which early learning processes contribute to this effect. There is evidence to suggest that the initial stage of encoding new meanings for familiar words involves inhibition of their existing meanings – the so-called “perturbation” of old knowledge (Fang & Perfetti, 2017, 2019). Although the current study tested participants four days after the learning phase, it would be invaluable to extend the delay (without further opportunities for consolidation) and confirm that the slower processing of existing meanings is due to semantic competition, rather than due to transient effects of this perturbation.

In summary, our novel finding that having learnt new meanings for known words slows the comprehension of their existing meanings has important implications for models of language acquisition and ambiguity processing. In particular, it lends support to the postulate of semantic competition in current models of semantic ambiguity, particularly those that predict at least some degree of competition for polysemous words (Rodd et al., 2004). Such competition in polysemy processing could be further modulated by the degree of overlap of the multiple senses (i.e., competition could be minimal or non-existent for the highly overlapping senses of metonyms but stronger for the less overlapping senses of irregular polysemes). The present experiments also add a novel type of evidence to the literature on the differential representation and processing of homonymy and polysemy. Using the artificial language learning paradigm, we demonstrate that relatedness in meaning influences the learning of new meanings and their subsequent impact on semantic processing. Further research into children’s and adults’ ability to learn new meanings for familiar words is of particular value. Not only does such research provide a novel avenue for testing predictions from the ambiguity literature, but it can also help us delineate mechanisms underlying successful language learning. Although there has been much progress in understanding how children learn new words or new meanings for words they already know (e.g., Casenhiser, 2005; Doherty, 2004; Storkel & Maekawa, 2005), and despite the fact that language is rife with semantic ambiguity, current models of vocabulary acquisition have largely ignored learning words with multiple interpretations (see Dautriche, Chemla, & Christophe, 2016), and how we continually expand our vocabulary throughout the lifespan.

Acknowledgements

The data for Experiment 1 was collected when the first author was a Master's student at University College London. Experiment 2 is part of the first author's PhD at the University of Leeds. This work was funded by an Economic and Social Research Council (ESRC) doctoral studentship (ES/J500215/1) and a University College London postgraduate research grant awarded to the first author. We would like to thank two anonymous reviewers for their helpful comments on an earlier draft of the manuscript.

Disclosure of interest

The authors report no conflict of interest.

References

Annett, M. (1967). The binomial distribution of right, mixed and left handedness. *The Quarterly Journal of Experimental Psychology*, *19*(4), 327-333.

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 273-278). Austin, TX: Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, *31*(7), 940-966.

Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple targets that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 827-845.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bartoń, K. (2014). MuMIn: Multi-model inference. Retrieved from http://CRAN.*R*-Project. org/package=MuMIn.

Bates, D., Mächler, M., & Bolker, B., (2011). Lme4: Linear mixed-effects modeling using S4 classes. Retrieved from http://CRAN.R-Project.org/package=lme4.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition*, *97*(3), B45-B54.

Briggs, G. G., & Nebes, R. D. (1975). Patterns of hand preference in a student population. *Cortex*, *11*(3), 230-238.

British National Corpus Consortium (2007). *British National Corpus Version 3*. Oxford, UK: Oxford University Computing Services.

Brocher, A., Foraker, S., & Koenig, J. P. (2016). Processing of irregular polysemes in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1798-1813.

Casenhiser, D. M. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language*, *32*(2), 319-343.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language, 55*(4), 767-811.

Clark, H. H., & Gerrig, R. J. (1983). Understanding old words with new meanings. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 591-608.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology, 33*(4), 497-505.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3773-3800.

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word learning: Homophony and the distribution of learning exemplars. *Language Learning and Development*, *12*(3), 231-251.

De Rosario-Martínez, H. (2015). Phia: Post-hoc interaction analysis. Retrieved from http://cran.r-project.org/web/packages/phia.

Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, *31*(1), 203-214.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429-446.

Eddington, C. M., & Tokowicz, N. (2015). How meaning similarity influences ambiguous word processing: The current state of the literature. *Psychonomic Bulletin & Review, 22*(1), 13-37.

Fang, X., & Perfetti, C. (2017). Perturbation of old knowledge precedes integration of new knowledge. *Neuropsychologia, 99*, 270-278.

Fang, X., & Perfetti, C. (2019). Learning new meanings for known words: Perturbation of original meanings and retention of new meanings. *Memory & Cognition*, 1-15.

Fang, X., Perfetti, C., & Stafura, J. (2017). Learning new meanings for known words: Biphasic effects of prior knowledge. *Language, Cognition and Neuroscience*, *32*(5), 637-649.

Fellbaum, C. (1998). WordNet. Retrieved from http://wordnetweb.princeton.edu.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*(2), 181-200.

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1366-1383.

Frisson, S., & Pickering, M. J. (2007). The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes*, *22*(4), 595-613.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105-132.

Gottlob, L. R., Goldinger, S. D., Stone, G. O., & Van Orden, G. C. (1999). Reading homographs: Orthographic, phonologic, and semantic dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 561-574.

Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? *Journal of Memory and Language*, *55*(2), 247-273.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718-730.

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 385-402.

Hulme, R. C., Barsky, D., & Rodd, J. M. (2018). Incidental learning and long-term retention of new word meanings from stories: The effect of number of exposures. *Language Learning*, 1-26.

Jager, B., & Cleland, A. A. (2015). Connecting the research fields of lexical ambiguity and figures of speech: Polysemy effects for conventional metaphors and metonyms. *The Mental Lexicon*, *10*(1), 133-151.

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944-946.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language*, *32*(4), 474-516.

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, *81*(1), 205-223.

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, *20*(1), 1-24.

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1534-1543.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978-990.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmerTest’. Retrieved from http://cran.r-project.org/web/packages/lmerTest.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306-353.

Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics*, 1(2), 207-246.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476-490.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315.

McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: Why starting Dickens is taxing but reading Dickens isn't. *Cognitive Science*, *30*(1), 181-192.

McRae, K. (2004). Semantic memory: Some insights from feature-based connectionist attractor networks. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 45, pp. 41-86). San Diego: Elsevier.

Medler, D.A., & Binder, J.R. (2005). MCWord: An on-line orthographic database of the English language. Retrieved from http://www.neuro.mcw.edu/mcword.

Murphy, G. L. (2006). Comprehending new words beyond their original contexts. *Skase Journal of Theoretical Linguistics*, 3(2), 2-8.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3*(2), 143-184.

Parks, R., Ray, J., & Bland, S. (1998). *Wordsmyth English dictionary-thesaurus*. Chicago, IL: University of Chicago.

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1252-1270.

Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory & Cognition*, *28*(4), 657-666.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 785-794.

R Development Core Team (2004). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14*(3), 191-201.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition, 40*(7), 1095-1108.

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language*, *87*, 16-37.

Rodd, J., Gaskell, M. G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46*(2), 245-266.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, *28*(1), 89-104.

Rodd J. M., Lopez Cutrin B., Kirsch H., Millar A., & Davis M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language*, *68*(2), 180–198.

Schneider, W., Eschman, A., & Zuccolotto, A. (2010). E-prime [computer software]. Pittsburgh, PA: Psychology Software Tools.

Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, *30*(6), 627-643.

Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua, 157*, 124-152.

Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology*, *62*(4), 245-272.

Srinivasan, M., & Snedeker, J. (2014). Polysemy and the taxonomic constraint: Children's representation of words that label multiple kinds. *Language Learning and Development*, *10*(2), 97-128.

Storkel, H. L., & Maekawa, J. (2005). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language*, *32*(4), 827-853.

Vicente, A. (2018). Polysemy and word meaning: An account of lexical meaning for different kinds of content words. *Philosophical Studies*, *175*(4), 947-968.

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, *18*(4), 742-750.