



# How emotion is learned: Semantic learning of novel words in emotional contexts<sup>☆</sup>

Bryor Sneffjella<sup>\*,1</sup>, Nadia Lana, Victor Kuperman

McMaster University, Canada

## ARTICLE INFO

**Keywords:**  
Word learning  
Eye tracking  
Semantic prosody  
Vocabulary acquisition

## ABSTRACT

The present paper addresses two under-studied dimensions of novel word learning. We ask (a) whether originally meaningless novel words can acquire emotional connotations from their linguistic contexts, and (b) whether these acquired connotations can affect the quality of orthographic and semantic word learning and its retention over time. In five experiments using three stimulus sets, L1 speakers of English learned nine novel words embedded in contexts that were consistently positive, neutral or negative. Reading times were recorded during the learning phase, and vocabulary post-tests were administered immediately after that phase and after one week to assess learning. With two of three stimulus sets, the answer to (a) was positive: readers learned both the forms, definitional meanings and emotional connotations of novel words from their contexts. We confirmed (b) in two of three stimulus sets as well. Items were learned more accurately (by 10% to 20%) in positive rather than negative or neutral contexts. We propose the transfer of affect to a word from its collocations to be a virtually unstudied yet efficient mechanism of learning affective meanings. We further demonstrate that the transfer that occurs over a few exposures to a novel word in context is sufficient to elicit a long-lasting positivity advantage previously shown in existing words only. Null results in one stimulus set suggest that contextual transfer of affect is contingent on other contextual properties, such as text complexity. These findings are pitted against theories of vocabulary acquisition.

## 1. Introduction

Human capacity for learning new words is a key topic in both basic and applied language research. This is hardly surprising given the amount of word learning that people experience even after childhood, as fluent adult speakers and readers of one or more languages. For instance, an average 20-year-old native speaker of American English is estimated to learn three lemmas every week (Brysbaert, Stevens, Mandera, & Keuleers, 2016), and a similarly high rate of vocabulary acquisition is likely to characterize any native language. Given the

privileged role of language faculties such as speaking and reading in one's social, professional and academic well-being (Graham, 1987; Vinke & Jochems, 1993), understanding linguistic, environmental and cognitive factors that contribute to success of vocabulary acquisition has both a theoretical and practical merit (Frost, Siegelman, Narkiss, & Afek, 2013). The present study focuses on *affect* as one of the factors that psycholinguistic research has long associated with both the learning and representation of words in the mental lexicon.

The body of knowledge regarding interactions of affect and language learning is well-developed but, as we argue below, incomplete.

<sup>☆</sup> Kind acknowledgments are due to Sadaf Rahmanian, Constance Imbault, Gaisha Oralova, Alyssa Janes, Melda Coskun, and Joyce Bauman for assistance with data collection. Thanks also to Gabriella Vigliocco and Gail Mauner for consultation during the initial stages of the project, as well as the audience of the Society for the Scientific Study of Reading 2019 conference.

<sup>\*</sup> Corresponding author at: Department of Psychology, UCLA, Pritzker Hall 7581, Los Angeles, CA 90095-1563, United States.

*E-mail address:* [bsneffjella@psych.ucla.edu](mailto:bsneffjella@psych.ucla.edu) (B. Sneffjella).

<sup>1</sup> Bryor Sneffjella's contribution was supported by the Social Sciences and Humanities Research Council of Canada Joseph-Armand Bombardier Canada Graduate Scholarship 767-2016-2299 <http://www.sshrc-crsh.gc.ca/home-accueil-eng.aspx>, and an Ontario Graduate Scholarship. Nadia Lana's contribution was supported by Social Sciences and Humanities Research Council of Canada Partnership Training Grant 895-2016-1008 (Libben, PI). Victor Kuperman's contribution was partially supported by the Natural Sciences and Engineering Research Council of Canada [http://www.nserc-crng.gc.ca/index\\_eng.asp](http://www.nserc-crng.gc.ca/index_eng.asp) Discovery grant RGPIN/402395-2012 415 (Kuperman, PI), the Ontario Early Researcher Award <https://www.ontario.ca/page/early-researcher-awards> (Kuperman, PI), the Canada Research Chair <http://www.chairs-chaire.gc.ca/program-programme/index-eng.aspx> (Tier 2; Kuperman, PI), the Canada Foundation for Innovation Leaders Opportunity Fund <https://www.innovation.ca/about/news/leaders-opportunity-fund> (Kuperman, PI), and the Social Sciences and Humanities Research Council of Canada Partnership Training Grant 895-2016-1008 (Libben, PI). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

One such interaction is under discussion in several theories of semantic knowledge. While different in the explanations they offer, these theories claim that emotional knowledge is crucial for learning of abstract concepts, i.e., concepts that are primarily learned through language rather than through sensorimotor experience with the object that the concepts denote (e.g., Barsalou, 2009; Vigliocco, Meteyard, Andrews, & Kousta, 2009; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011, among others). Indeed, it is a robust finding that children learn abstract concepts earlier and recognize abstract words with less effort if these concepts have stronger emotional connotations (Lund, Sidhu, & Pexman, 2019; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Ponari, Norbury, & Vigliocco, 2017; Vigliocco, Ponari, & Norbury, 2018). For related work on abstractness and emotion in word learning, recognition, categorization in children and adults see e.g., Newcombe, Campbell, Siakaluk, and Pexman (2012), Siakaluk, Knol, and Pexman (2014), and Siakaluk et al. (2016).

Another well-studied interaction of affect and lexical processing is observed in stable semantic representations of known words. Affect has long been known to influence cognitive processes responsible for recognition, recall and retention of existing words (e.g., Ashby, Isen, & Turken, 1999; Bower, Monteiro, & Gilligan, 1978; Singer & Salovey, 1988). Negative stimuli tend to elicit slower responses than neutral or positive ones in tasks requiring word recognition, and come with lower recall rates in recognition memory or recall tasks (see Algom, Chajut, & Lev, 2004; Altarriba & Bauer, 2004; Kuperman, 2015; Kuperman, Estes, Brysbaert, & Warriner, 2014; Williams, Mathews, & MacLeod, 1996; Wentura, Rothermund, & Bak, 2000, but see Adelman & Estes, 2013; Kensinger & Corkin, 2003; Kousta, Vinson, & Vigliocco, 2009, for diverging findings). A proposed explanation for the advantage in processing and memorization of positive words and penalty to negative words is automatic vigilance towards negative words (Erdelyi, 1974; Pratto & John, 1991). Because they denote concepts, events and objects that may be threatening to one's survival (e.g., epidemics, rifle, tiger), negative stimuli engage attention longer than other stimuli (Fox, Russo, Bowles, & Dutton, 2001; Öhman & Mineka, 2001) causing longer response latencies.

The theoretical accounts above primarily address mechanisms that underlie acquisition of word meaning per se. Either via a definition given through language or via situational contexts in which a word occurs, a person learns to associate *sin* with negative affect, *vacation* with positive and *cosine* with neutral affect. These connotations – along with the experiences that accompany them – influence how easily the word is learned, recognized, recalled and retained in memory. Yet this family of approaches overlooks an important aspect of semantic learning. A word's meaning does not only encode affect of the word itself, but also affective characteristics of the linguistic contexts in which it occurs.

Affect may transfer to a word from its context. Such transfer is documented in known words under the label of *semantic prosody*, “a consistent aura of meaning with which a form is imbued by its collocates” (Louw, 1993: 157, see also Hunston, 2007; Sinclair, 1996, 2004). For instance, the words *cause*, *utterly* and their synonyms *produce*, *totally* are all relatively neutral in their valence (5.26, 5.22, and 5.63, 5.42 on a 1–9 negative-to-positive scale of Warriner, Kuperman, & Brysbaert, 2013), but both words in the former pair tend to co-occur with largely negative collocates (e.g., *cause harm*, *damage*, *war*; *utterly ridiculous*, *dependent*, *alone*), whereas collocates of *produce* or *totally* are neutral on average. Thus, semantic prosody of *cause* is more negative than that of *produce*. Sneffjella and Kuperman (2016) further demonstrated that semantic prosody is not only found in a select subset of lexical items. They operationalized context for a given word as a 10-word window surrounding that word and calculated average valence, arousal, and concreteness of each context in which the target word occurs in a large corpus of natural writing. Semantic prosody of the target word was then defined as the average valence, arousal or concreteness across all contexts of the target word. Further analyses showed that it is a sweeping

tendency of natural language for words to absorb emotional and sensorimotor information from their linguistic contexts, in addition to the emotional connotations that a word would acquire if it were learned or used in isolation.

While they show a weak positive correlation, valence (psychological positivity) of a word is demonstrably different from average valence of the contexts in which that word occurs. For instance, words *patriotic* and *rewarding* are judged as much more positive than average valence of their respective contexts; the situation is the opposite with words like *blinded* and *motherless*, see Table 4 in Sneffjella and Kuperman (2016). Such incongruencies are not accidental or inappropriate. They are statistical facts about word use in context and – as shown below – are learned by language speakers along with other information about lexical contexts (e.g., common collocates, genre and register of use, context diversity and such, see references below). In the framework of the Lexical Quality hypothesis (Perfetti & Hart, 2002; Perfetti, 2007), information about emotional and sensorimotor characteristics of contexts becomes part of a rich and nuanced semantic representation of the word critical for successful word learning (Perfetti, 2007).

Sneffjella and Kuperman (2016) have found independent contributions of word valence and context valence to explained variance in a large number of language tasks. Thus, words were recognized faster in a lexical decision task, acquired at an earlier age and predicted better serial recall in the lexical recognition memory task both if they were more positive and, independently, if they had a more positive context valence (for supporting empirical evidence see also Ellis, Frey, & Jalkanen, 2009; Winter, 2016). Emotional and sensorimotor properties of contexts explained unique variance in all tasks and in some tasks they were more influential as predictors of lexical memory and processing than the respective properties of words per se.

Thus, when processing known words, both sources of affective information – the word per se and its linguistic contexts – appear to be tapped into. Yet existing work has established this processing mechanism for words that have reached a “steady state” of this aspect of vocabulary knowledge. How does semantic prosody emerge in previously unseen words? Does it take place immediately, within a few exposures to a new word, or accumulates slowly, over years of experience? Is context valence as influential for concept learning as is valence of the concept itself? To our knowledge, semantic prosody has never been demonstrated in novel words, and its potential role in word learning is unexplored. In a series of experiments on novel word learning, the present study tracks the development of semantic prosody in novel word learning in L1 and estimates the influence of context valence on success of immediate and delayed vocabulary acquisition.

#### *The present study*

The motivation for this study is the phenomenon of *semantic prosody*, a robust finding that word meanings absorb some affective polarity of the contexts in which they tend to occur in natural language use. So far, evidence regarding semantic prosody mostly comes from the “steady state” of word learning, i.e., known words with relatively established semantic representations that include knowledge of the word's contexts. What is unknown is how semantic prosody comes about at the initial stages of novel word learning and whether context affect can confer independent benefits or penalties to learning outcomes. This paper investigates whether semantic prosody occurs in novel words, that is, words that do not initially have either a denotation or a connotation of their own.

Our first goal is to determine whether a transfer of affect from the linguistic context to a novel word can occur over a few exposures to that novel word when it consistently co-occurs with emotionally charged words. There are at least two mechanisms that may enable this transfer of affect from linguistic context to a novel word. The embodied account of meaning acquisition (Barsalou, Santos, Simmons, & Wilson, 2008) offers one mechanism. Specifically, it suggests that emotional and bodily states experienced by the reader are encoded along with the symbolic aspects of the word meaning at every word occurrence.

Furthermore, every encounter with a word is not confined to the activation of a literal word meaning but also simulates the encoded emotional and bodily state (Barsalou & Wiemer-Hastings (2005); Barsalou et al., 2008; Fischer & Zwaan, 2008; Pexman, 2019, see also Kiverstein & Miller, 2015). This notion finds support in studies on acquisition of emotional, concrete and abstract words (Borghi et al., 2017; Ferré, Ventura, Comesaña, & Fraga, 2015; Hald, van den Hurk, & Bekkering, 2015; Inkster, Wellsby, Lloyd, & Pexman, 2016; Ponari, Norbury, & Vigliocco, 2018). It also converges with recent work on “emotional experience”, a measure designed to tap into the ease with which words evoke emotional states in an individual. Emotional experience is found to be predictive of word recognition and categorization, especially for abstract words (e.g., Newcombe et al., 2012; Siakaluk et al., 2016). The embodied account predicts that if a substantial number of words in the linguistic context of a novel lexical item have a consistent polarity and a substantial magnitude of affect, that item will be learned in a specific emotional state and its meaning will encode a degree of that affective connotation, regardless of what it denotes.

Another possibility is of a cognitive rather than affective nature. As reported in Sneffella and Kuperman (2016), words tend to co-occur with words of a similar emotional polarity: e.g., an average valence of a 10-word window around a positive word will be more positive than that around a negative word. With this distributional knowledge about language use, speakers and readers may tend to transfer affective polarity of the novel word’s context to the connotation of the novel word itself.

Our first hypothesis is that semantic prosody transcends the domain of existing words, and can occur even during very few exposures to a novel word in an experimental paradigm that is relatively impoverished in terms of linguistic or semantic diversity. A demonstration that such a transfer took place under controlled experimental conditions would indicate a potential initial point of the learning trajectory, the end-point of which is an established effect of semantic prosody on lexical memory and processing of known words (see review in Winter, 2019). We also discuss implications of our findings for the two theoretical mechanisms described above.

Our second goal is to examine whether affective characteristics of linguistic contexts can influence the quality of novel word learning both immediately and one week after the learning took place. Processing advantages to words with more positive valence are well known (see above), and Sneffella and Kuperman (2016) reported similarly strong and independent processing advantages to words occurring in more positive contexts. We hypothesized that novel words learned in consistently positive contexts will demonstrate an advantage similar to that shown by positive words in a variety of lexical tasks, as compared to neutral or negative counterparts. Specifically, we expect that semantic knowledge for novel words in positive contexts will be stronger than for those learned in neutral or negative contexts, and that this advantage will hold over time, in immediate and deferred testing sessions with vocabulary post-tests.

Whether semantic prosody occurs rapidly and whether it influences the quality of novel word learning has importance because of the attention that the language instruction and language acquisition literature affords to this phenomenon in particular and to the role of affect in general (e.g., Guo et al., 2011; McGee, 2012; Zhang, 2009). If affect expressed in linguistic contexts influences word learning, this presents a potential intervention to boost word learning. This could be achieved by, for example, manipulating instructional materials to place hard to learn words in more or less affectively biased contexts, or to select words with “naturally occurring” affectively biased contexts for earlier or later instruction.

We set out to address these two goals by using a common setup of novel word learning studies (e.g., Elgort, Brysbaert, Stevens, & Van Assche, 2018). In our study readers were presented with sets of short passages (i.e., linguistic contexts) that contain one novel word each and differ from one another only in a small number of critical existing words

strongly associated with positive, neutral or negative affect: full passages for all experiments are contained in Tables S2–S4 in the supplementary materials and see example below and details in the Methods section:

1. Some friends were planting flowers in the garden. They used a (NONWORD) to dig a hole.
2. Some people were working on the outer grounds. They used a (NONWORD) to dig a hole.
3. Some murderers needed to dispose of a body. They used a (NONWORD) to dig a hole.

We embedded a small number of novel words in short passages with consistent emotionality (positive, neutral or negative), and used behavioral measures to monitor the real-time dynamics of learning as well as vocabulary post-tests to estimate learning outcomes, see below. These experimental conditions were created to simulate a scenario when a new word is coined and its contextual use in natural language is biased towards positive, neutral, or negative contexts. As we discuss above, such biases are not accidental nor rare, but rather satisfy natural needs of communication. While natural contexts of different emotionality are likely to be intermixed and their tendencies towards one of the emotional extremes or neutrality would not be absolute, we keep contexts consistently close to either the extremes or the middle of the affective scale to generate a stronger behavioral signal. Since the denotation of the novel words in our experiments are emotionally neutral (boat, instrument, kitchen utensil), any difference that we would see between conditions in valence judgments to those words after exposure in contexts would be because of the contexts. Specifically, it would indicate a transfer of affect from linguistic contexts to the word during initial phases of learning, akin to semantic prosody in known words (see review in Winter, 2019).

Multiple options in the nature of stimuli and the task were available to us, yet any given experiment can engender only one set of choices. We chose the options that we believed would emphasize the effect of emotion. Specifically, we opted for an intentional learning paradigm, in which participants are made aware that the novel words they encounter will be used in post-tests. We also chose the syntactic role of each novel word as a noun in a direct or indirect object position in the respective sentence, and we selected intended meanings of the novel words to be relatively concrete (e.g., a tool, a kitchen utensil, or a musical instrument). We discuss the implications of our choices and speculate on how they generalize over other types of tasks or stimuli in the General Discussion.

We employed a combination of experimental tasks to investigate processes relevant to word learning both in real time (eye-tracking and other chronometric measurements of reading passages with novel words), immediately after learning (vocabulary post-tests) and one week later (another session with vocabulary post-tests). This combination offers advantages that single paradigms do not afford (Chaffin, Morris, & Seely, 2001; Elgort et al., 2018; Godfroid, Boers, & Housen, 2013; Godfroid, Winke, & Rebuschat, 2015; Rayner, 1998, among others). For instance, multiple studies in L1 and L2 word learning have pointed out that a higher quality of word knowledge (showing in higher scores in vocabulary post-tests) is achieved if either the learned items elicited longer reading times, or their contexts elicited longer reading times, or both (see e.g. reviews in Godfroid et al., 2013; Godfroid & Schmidtke, 2013; and Williams & Morris, 2004). The Noticing Hypothesis by Schmidt (1990, 2001) offers an explanation for this correlation. It argues that detection and especially conscious awareness of the new form and meaning – labeled together as noticing – are a necessary and sufficient condition of word learning, even in the incidental mode. Godfroid, Housen, and Boers (2010), Godfroid et al. (2013) and Godfroid and Schmidtke (2013) further argue that detection and awareness can be operationalized experimentally as the amount of time the eyes spend on a word, as well as reading times to entire texts or

specific text elements (e.g., Brusnighan & Folk, 2012; Godfroid & Schmidtke, 2013; Pellicer-Sánchez, 2016; Roberts & Siyanova-Chanturia, 2013; Schmidt, 1990, 1993; Williams & Morris, 2004). We registered eye movements to every word and passage in the laboratory-based experiments and reading times for passages in web-based experiments. This enabled us to verify earlier reports on the role of exposure (the number of learning events for a novel word) and inspection time (the metric of attention towards the novel word and the passage) in novel word learning.

Using the same experimental settings, we created three sets of stimuli (see Tables S2–S4 in the supplementary materials) to offer a strong test of validity and generalizability of our findings: these constitute experiments 1, 2, and 3. All experiments are presented jointly.

## 2. Methods

For experiments 1 and 2, we first collected both eye-tracking and post-test data from a relatively small sample of participants (around 36 participants each): these are labeled Experiments 1L (where L stands for lab) and 2L, respectively. Since these samples were underpowered (see power analysis below) we complemented this effort by web-based studies (Experiments 1O, where O stands for online, and 2O), in which only the learning phase and the post-tests were administered to larger samples and the eye-tracking was omitted. Experiment 3O was constructed to have tighter control of syntactic structure and semantic cues to the nonwords meanings. Experiment 3O was conducted solely online. Data from all experiments is available in an online at <https://osf.io/yghx3/>.

**Participants** Thirty-six monolingual English-speaking undergraduate students participated in the lab-based Experiment 1L for course credit (average age = 20.78, *sd* = 2.37; 34 females). Two participants completed only the eye-tracking portion of the study and did not complete the post-tests. Thirty-two monolingual English-speaking undergraduate students (average age = 21.8, *sd* = 5.8; 28 females) participated in the lab-based Experiment 2L for course credit (average age = 20.9, *sd* = 2.5, 30 females). All had normal or corrected-to-normal vision and did not report any learning or visual impairments. All studies were approved by the McMaster Research Ethics Board (protocol 2011–165).

Power analysis (reported below) revealed that the lab-based samples did not offer high power for the observed effect sizes. Additional Experiments 1O, 2O, and 3O recruited samples of participants from Amazon's Mechanical Turk online crowdsourcing platform (mturk.com). All participants had USA as their place of residence and were compensated with \$4 USD. For Experiments 1O and 2O, in each sample we targeted 120 participants as indicated by the power analysis. For Experiment 3O, we were forced to omit the second testing session due to a change in Amazon's Mechanical Turk platform software, and instead recruited 120 participants. After removing participants whose native language was not English and participants who took part in more than one experiment, the resulting sample sizes were 113 in Experiment 1O (average age = 37.27, *sd* = 12.07; 50 females), 104 in Experiment 2O (average age = 36.22, *sd* = 10.58; 46 females), and 118 in Experiment 3O (average age = 36.86, *sd* = 11.89; 44 females).

**Materials:** Nine novel words served as targets in all experiments (e.g., plurk, ceammy). We opted against a larger number of items because of a limited human capacity for learning new information, discussed as early as Miller (1956). Our literature review reveals that prior studies using the present experimental paradigm (embedding novel words in contexts without a prior familiarization phase) include between 5 and 10 novel words (6 words in Pellicer-Sánchez, 2016; 6 words in Joseph & Nation, 2018; 10 words in Webb, 2007). Larger numbers of experimental items are possible but require paradigms that either use reading of fiction with foreign words (Saragi, 1978; Godfroid et al., 2018), or a prolonged learning phase (16 novel words learned over 5 days in Joseph, Wonnacott, Forbes, & Nation, 2014). Since our

design requires the number of items to be a factor of 3 (to represent the positive, neutral, and negative contexts), we can only sensibly present up to 9 novel words without risking a very low and uninformative memory performance. Importantly, reliance on 9 items does not create a statistical problem, given a large number of repeated exposures to each item (9 items x 5 times) and a substantial number of participants.

All novel words were phonotactically legal in the English language and generated with the help of the Wuggy software (Keuleers & Brysbaert, 2010) which generates nonwords with maximally similar orthographic properties to one or more input strings. Each stimulus was selected to have three homophones (e.g. plurk, plirk, plerk) used in vocabulary post-tests. All novel words and their homophones were rated in a separate norming study for their positivity on the 1 (sad) to 9 (pleasant) scale. Sixty participants (mean age = 34.5, *sd* = 11.39; 31 females) were recruited via Amazon's Mechanical Turk crowdsourcing platform and paid \$0.50. None of them took part in any other study reported here. From a larger pool of candidate stimuli, we retained the nine that showed no difference in valence between all homophones of a given wordform, as indicated by a regression model fitted to valence ratings to candidate stimuli (not shown). The nine target words ranged in their average valence ratings from 2.44 (*rotch*) to 4.22 (*ceammy*). Table S1 in the supplementary materials contains the target nonwords and their homophones. Since all our mixed-effects regression models included novel word as a random effect, we had statistical control over potential differences that may have emerged from placing novel items with varying positivity in negative, neutral and positive contexts. Also, we argue, and demonstrate below, that valence ratings obtained out of context are a poor measure of a novel word's as-yet-unlearned affective connotation.

In designing our stimuli, we relied on the findings and instruments proposed in the prior literature. Perhaps the most robust finding of the word learning literature is the facilitatory effect of the number of exposures to the novel word, either on the processing speed or knowledge of that word or both (see e.g., Blythe et al., 2012; Ellis, 2002; Godfroid et al., 2018; Joseph et al., 2014; Mokhtar, Rawian, Yahaya, Abdullah, & Mohamed, 2017; Pellicer-Sánchez, 2016; Teng, 2016). A novel word that is seen more times or in more diverse contexts is learned better, both in terms of orthography and semantics (Pagán & Nation, 2019). We kept the number of exposures to each novel word constant. Randomized order of novel words for each participant was used to avoid an order-of-acquisition effect reported in Joseph et al. (2014).

Researchers have also explored multiple additional factors influencing word learning (see among others Brusnighan & Folk, 2012; Chaffin et al., 2001; Eskenazi, Swischuk, Folk, & Abraham, 2018; Godfroid et al., 2018; Joseph & Nation, 2018; Lowell & Morris, 2017; Schwanenflugel, Stahl, & McFalls, 1997; van den Broek, Takashima, Segers, & Verhoeven, 2018; Webb, 2008; Williams & Morris, 2004; see also Fang, Perfetti, & Stafura, 2017). We controlled for context length (e.g., Swanborn & De Gloppe, 1999; Wochna & Juhasz, 2013) keeping the amount of linguistic material around the novel word constant within each stimulus set. Since context informativity or constraint also influence learning (e.g., Bolger, Balass, Landen, & Perfetti, 2008; Daneman & Green, 1986) we took measures to keep most of the context identical across critical conditions and only changed 2–3 words in each context that had positive, neutral or negative connotations.

Each word in each Experiment was presented in five short passages: every participant saw a total of 45 (9 x 5) passages. The passages were created in triplets, see (1–3) above. The stimuli passages differed between Experiments 1L and 1O, Experiments 2L and 2O and Experiment 3O, but were constructed in a similar manner. Each passage in a triplet had an identical last sentence, which contained one occurrence of the novel word. Yet the passages in each triplet differed in whether their initial sentences contained a number of highly positive (Example 1), highly negative (Example 3) or emotionally neutral words (Example 2). Any given novel word for a given participant always appeared in contexts with the same level of emotionality. To avoid confounds, we



counterbalanced the association between a target novel word and context emotionality over participants: e.g., for participant A a novel word *plurk* would occur in 5 positive contexts, for participant B in 5 negative contexts, and for participant C in 5 neutral contexts. Each participant was exposed to the same overall number of neutral, negative, and positive passages (15 each), and the order of passage presentation was randomized.

Semantic diversity of contexts in which a novel word occurs is consequential for its learning, with a greater diversity boosting learning (Johns, Dye, & Jones, 2016). Also, as the Syntactic Bootstrapping hypothesis argues (e.g., Gleitman, 1990; see review by Fisher, Gertner, Scott, & Yuan, 2010), syntactic cues (e.g., part of speech or grammatical role in the sentence) can considerably constrain the semantic field that a novel word falls into. For example “The rotnch heard the arguments and issued a verdict” can identify a novel word *rotnch* as an animate noun that can be an agent. We made sure to keep semantic and syntactic diversity constant. Thus, each novel word was associated with one denotation only. All novel words had a syntactic role of a noun used as a direct or an indirect object in the sentence.

Passages provided information that made it possible to infer the intended denotation of novel words as a broad semantic category (e.g., a tool as in Examples 1–3, food, musical instrument or vehicle). A full counterbalancing of the 9 intended denotations, 3 conditions, and 9 nonwords was not feasible. We therefore created 9 stimuli lists, and opted to balance novel word occurrences across conditions, (i.e. “plurk” was seen in negative, positive, and neutral conditions) and intended denotations (i.e. a “plurk” was seen in each of the 9 intended meanings). This means that some combinations of denotations, nonwords, and emotional contexts (e.g., a “plurk” with the intended denotation of “tool” in negative context) did not occur in the experiment.

An additional factor of influence is conceptual difficulty of the context. Novel words are learned better in contexts with simpler rather than more demanding content, which can be gauged as higher and lower values of readability, respectively (e.g., Herman, Anderson, Pearson, & Nagy, 1987; Nagy, Anderson, & Herman, 1987). Passages used in Experiments 2L, 2O were an edited version of the Experiment 1L, 1O stimuli, changed for increased readability. Simplification of passages was meant to ensure that the novel word is likely to be the only unknown word to the reader in the passage, even for native speakers with a lower proficiency level for reading English: our web-based experiments recruited from a wide variety of skills and abilities, well beyond university participant pools. Specifically, we replaced several nouns and verbs in the passages with higher-frequency synonyms and implemented a minor syntactic change in two passages. We also edited some sentences to contain words that were semantically unrelated to the denotations of the nonwords, intended for use in a semantic priming post-task, described below. Both the increased length and the reduced linguistic complexity have been reported to benefit novel word learning outcomes (see the Introduction) and so the Experiment 2L, 2O stimulus set was expected to render it an easier material for novel word learning. Since in all stimulus sets a novel word only occurred once in each passage, and always in the same part of speech (noun) and grammatical role (direct or indirect object), we did not expect a difference in the quality of syntactic cues to emerge. The structure and the number of the passages in Experiments 2L, 2O and Experiment 3O replicated the one used in Experiments 1L, 1O, as did the selection of nine target novel words and their homophones.

As reported below, this simplification and modification of the 2L, 2O passages yielded a null effect of the manipulation of context on subsequent valence judgments to the novel words. Therefore, we constructed a new set of passages for the replication study Experiment 3O. For these passages, we used a strict template for each sentence containing emotionally loaded words:

Determiner (semantically unrelated negative/neutral/positive adjective) + (semantically related neutral noun) + (semantically

related verb) + determiner + (semantically related negative/neutral/positive adjective) + (semantically related negative/neutral/positive noun),

where semantically related refers to a high level of semantic similarity between a word and the intended meaning for nonwords within that passage, as determined by the ConceptNet Numberbatch co-occurrence model of semantics (Speer, Chin, & Havasi, 2017). This rigid structure provided tighter matching of conditions on passage length, syntactic structure and how strongly the passages cue the intended denotation of the nonword.

Overall, five sets of passage triplets were created, to a total of 15 passages in each of the three stimulus sets. A norming study with 60 raters who did not participate in any other experiment reported here rated positivity of resulting passages on the 1 (sad) to 9 (pleasant) scale for each distinct stimulus sets used in experiments 1L and O, 2L and O, and 3O respectively. These participants were recruited via Amazon’s Mechanical Turk crowdsourcing platform and compensated \$1. Analyses revealed that our manipulation led to significant differences in perceived valence of the passages in stimuli used in Experiments 1L, 1O: negative  $M = 2.47$  ( $sd = 1.43$ ), neutral  $M = 4.56$  ( $sd = 1.32$ ), and positive  $M = 6.03$  ( $sd = 1.27$ ). Pairwise contrasts between all levels were highly reliable ( $p < 0.001$ ) in a linear mixed-effects model (Table 1) with rating as a dependent variable, condition as a predictor and sentence as a random effect. Supplementary materials S2, S3, and S4 report the stimuli passages and their emotionality ratings.

In passages in Experiment 2L, 2O the context manipulation still led to significant differences in perceived valence of the passages: negative  $M = 3.57$  ( $sd = 1.81$ ), neutral  $M = 5.41$  ( $sd = 1.89$ ), and positive  $M = 6.78$  ( $sd = 1.77$ ). Pairwise contrasts between all levels were highly reliable ( $p < 0.001$ ) in a linear mixed-effects model (Table 1) with rating as a dependent variable, condition as a predictor and sentence as a random effect. Overall, the perceived affect of new passages in Experiment 2L, 2O was higher ( $p < 0.001$ ) than in Experiments 1L, 1O by 1.13 points on a 1–9 point scale. The contrast between the negative versus neutral conditions was reduced by .26 points, and negative versus positive was reduced by .38 points in Experiments 2L, 2O. Supplementary materials S2 report the stimuli passages and their valence ratings.

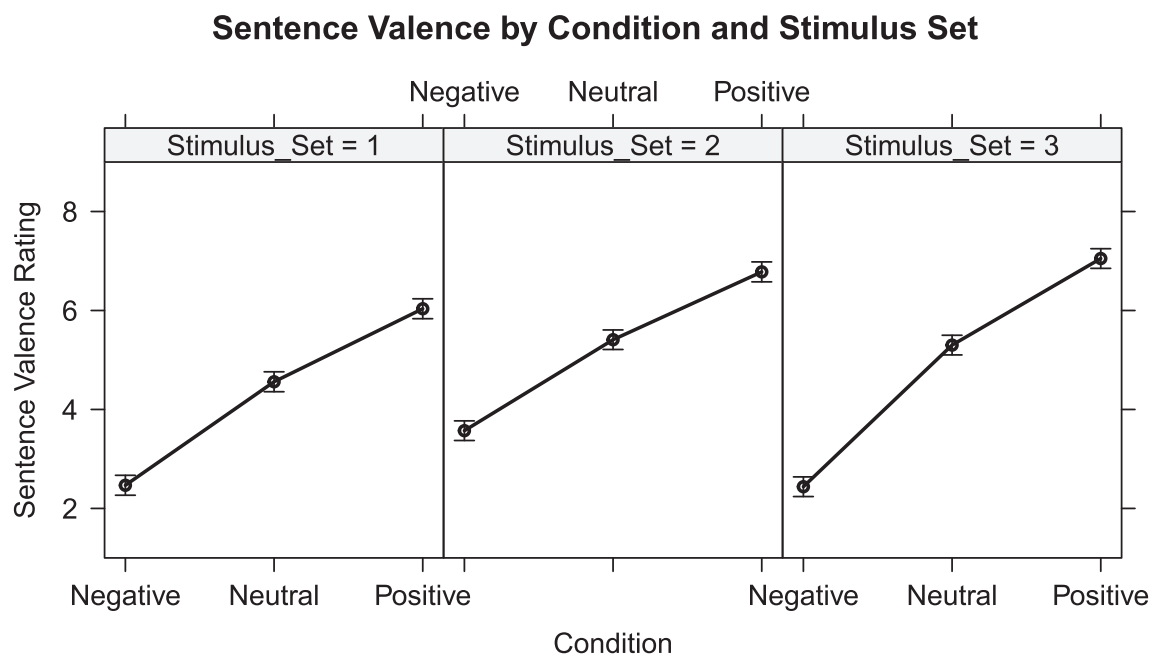
For the Experiment 3O stimuli the context manipulation led to significant differences in perceived valence of the passages: negative  $M = 2.44$  ( $sd = 1.68$ ), neutral  $M = 5.3$  ( $sd = 1.51$ ), and positive  $M = 7.05$  ( $sd = 1.68$ ). With these stimuli, the neutral condition was .74 points more positive than in Experiment 1, and the positive condition

**Table 1**

Results of sentence norming on valence, for stimulus sets used in Experiments 1L, 1O (first column), Experiments 2L, 2O (second column), and Experiment 3O (third column).

	Stimuli 1: (1L & 1O)	Stimuli 2: (2L & 2O)	Stimuli 3: (3O)
(Intercept)	2.47*** (.10)	3.57*** (.10)	2.44*** (.10)
Condition Neutral	2.09*** (.14)	1.84*** (.14)	2.86*** (.15)
Condition Positive	3.57*** (.14)	3.21*** (.14)	4.61*** (.15)
AIC	7690.22	10063.14	9579.62
BIC	7718.99	10092.27	9608.86
Log Likelihood	−3840.11	−5026.57	−4784.81
Num. obs.	2327	2507	2562
Num. groups: Sentence	135	135	135
Var: Sentence	.36	.29	.38
(Intercept)			
Var: Residual	1.44	3.05	2.27

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$



**Fig. 1.** Results of norming passage stimuli for valence for the three stimulus sets. Error bars are 95% confidence intervals. Column 1 depicts stimuli for Experiments 1L and 1O, column 2 Experiments 2L and 2O, and column 3 Experiment 3O.

was 1 point more positive than in Experiment 1. The Experiment 3O stimuli therefore have the greatest spread in rated valence of the three stimulus sets. The Experiment 2L, 2O stimuli have the least spread over the range of valence. Fig. 1 displays the results of norming all three stimulus sets for valence.

**Procedure and Apparatus:** Experiments 1L and 2L were conducted in an eye-tracking laboratory in two sessions separated by one week. The first session included all tasks described below. It began with participants signing a consent form and filling out a short demographic questionnaire, and then proceeded to the reading task during which participants' eye-movements were recorded. The second session only administered vocabulary post-tests, described below. No additional exposure to target words in contexts was made available in that second session. The same participants completed the first and the second session in each experiment.

**Passage reading:** Participants were instructed to read passages silently for comprehension while their eye-movements were recorded. Participants were informed that "After [they] have read all the passages, [they] will perform some tasks related to the spellings and meanings of the unfamiliar words." Thus, we opted for the intentional type of learning. As outlined above, each participant read 45 passages, which represented 9 novel words in 5 emotionally consistent contexts. Three novel words appeared in positive, neutral and negative contexts each. Emotionality of context was a critical manipulation of the study.

For the eye-tracking phase of the English segment of the study, participants were seated in a comfortable chair approximately 65 cm in front of an NEC MultiSync LCD 17-inch computer monitor with a resolution of 1600 x 1200 and screen refresh rate of 60 Hz. Tahoma 20 point fixed-size was used for text passages, resulting in about 3 characters subtending 1 degree of visual angle. Eye-movements were recorded with an EyeLink 1000 desktop eye-tracker (SR Research, Kanata, Ontario, Canada) at a sampling rate of 1000 Hz. We used DataViewer software (SR Research, Kanata, Ontario, Canada) for aggregation of eye-tracking samples into fixations and saccades and reporting of all eye-movement measures summarized by word. Default DataViewer settings were used for removal of blinks and very short fixations, as well as merging of fixations that were close (within one character) to one another.

Calibration was performed using a series of nine fixed targets

distributed around the display, followed by a 9-point accuracy test to validate eye position. Stimuli were viewed binocularly, but eye-movement data from only one eye was analyzed. Prior to the presentation of the trial stimuli, a dot appeared on the monitor screen, 20 pixels to the left of the first word in the passage. Once the participant had fixated on it, the trial would begin. Drift correction took place at the beginning of each trial and calibration was monitored and redone by the researcher if necessary. Each passage appeared on a separate screen and occupied 2–4 lines. Participants were instructed to press a key when their reading of a passage was completed. The order of passage presentation was randomized for each participant.

**Vocabulary post-tests:** These tests evaluated acquired knowledge of the novel word form and meaning. Word form knowledge was tested via an *orthographic choice* test, where participants were presented with the nine seen novel words (e.g., plurk) and their nine homophones (plirk), for a total of 18 items. Participants were instructed to press one of two keys if that wordform was seen or unseen during the passage readings. The proportion of correctly discriminated word forms was taken as an index of the strength of the learned orthographic representations. Due to a programming error, orthographic choice data are not available for Experiment 2L.

To differentiate depth of vocabulary knowledge, word-learning researchers distinguish between meaning recognition and meaning recall (see e.g., Laufer, Elder, Hill, & Congdon, 2004, Laufer and Goldstein, 2004). Meaning recall requires generating a word's meaning, whereas recognition merely requires choosing a meaning from a set of options. Recall is therefore considered the more difficult task, and indicates a more advanced knowledge of a word's meaning. We used two tasks intended to gauge the depth of knowledge of the novel words obtained by the participants, and whether our manipulation of emotionality of contexts made it easier to choose between meanings (recognition) or to generate a meaning (recall).

The *definition prompting* test tapped into meaning recall. Participants were presented with the nine seen novel words and nine unseen fillers (that were not homophones and were generated in the Wuggy application). Participants were asked to skip an item if they believed it was a filler. If the participant thought they had read the word during the passage reading, they were asked to provide a definition. If the experimenter judged the definition to be accurate (it corresponded to the

meaning encoded in the passages, e.g. a tool or a food), 3 points were awarded. If the participant could not generate a response or their response was wrong, the experimenter verbally provided a hint. The hint was the initial fragment of one passage in which that novel word occurred during passage reading. This hint therefore contained the emotionally loaded words seen earlier. A correct definition after this hint granted 2 points, while an incorrect answer prompted a second hint from a different passage. If the response was correct after two hints, 1 point was awarded. 0 points were awarded if the responses was incorrect, including if one of the novel words was thought to be a filler.

A *definition matching* task was used to tap into meaning recognition. In this test, participants were provided with a list of nine seen novel words and nine fillers (none of which were used in other tasks), as well as a list of the nine intended denotations for the novel words (a tool, food, musical instrument, etc.) and nine semantic categories that were not targeted in experimental passages (e.g., a plant or piece of clothing). The task was to connect seen lexical items with their definitions and ignore unseen fillers. Accuracy in this task indicated how strongly readers associated newly learned forms with context inferred meanings. Accuracy was calculated as providing the correct item for one of the non-filler definitions. By this definition of accuracy, randomly assigning items to definitions yields a chance accuracy of 5.6%.

*Valence ratings:* To assess whether semantic prosody (i.e., transfer of affect from emotion-laden context to an intrinsically neutral novel word) takes place in a word learning paradigm, we collected ratings of valence for novel words. Participants were asked to estimate how a novel word makes them feel on a scale ranging from 1 (sad/unpleasant) to 9 (happy/pleasant). We also included 9 filler words generated with Wuggy and not used in other tasks. Participants were instructed to press 0 on the keyboard if they thought an item was not used during the passage reading portion of the experiment.

The order of vocabulary post-tests in all experiments was as follows: orthographic choice, valence ratings, definition prompting, and definition matching. Experiment 2L additionally contained a semantic priming task. It was intended to gauge whether the novel words differentially prime semantically related/unrelated, and seen/unseen words, and thereby comment on depth of integration into the lexicon (for motivation for including this task, see Bordag, Opitz, Rogahn, & Tschirner, 2018; Dagenbach, Horst, & Carr, 1990). However, a programming error made the data from this task unusable, and we do not report results from it.

Web-based Experiments 1O, 2O were also administered in two sessions separated by one week. In each experiment, the same participants completed both sessions. The first session consisted of reading passages with embedded novel words and a battery of vocabulary post-tests (administered in this order: orthographic choice, valence ratings, definition prompting, and definition matching); the second session contained the same battery of vocabulary post-tests in the same order.

Unlike Experiments 1L and 2L, the readers' eye-movements were not registered during passage reading in Experiments 1O, 2O, and 3O. The semantic priming task from Experiment 2L was not implemented in Experiments 1O, 2O, or 3O either. Another difference in procedure occurred in administration of the definition prompting task. In lab-based experiments (1L, 2L) accuracy of the response was evaluated by the experimenter in real-time, and additional hints were or were not offered accordingly. Conversely in the web-based experiments (1O, 2O, and 3O) the participant always received a novel word without a hint and was asked to respond "I don't know this word" or provide a definition. Regardless of the response, the first hint was presented and a response was requested again, and then the second hint was presented and a response was requested again. In this case, we calculated the definition prompting score based on when the first correct response was given (3 points without hints, 2 points after one hint, 1 point after 2 hints, and 0 points otherwise).

#### Variables

*Reading:* Eye-movement measures during reading, available in

Experiments 1L and 2L serve as real-time indices of word learning and the use of contextual cues. We selected the following measures as dependent variables: total fixation time on the word (a summed duration of all fixations on the word) as a cumulative measure of cognitive effort of word recognition, skipping rate as an index of how many opportunities readers took to direct their overt attention to the word, and regression rate from the word as a measure of how often readers consulted already seen context to infer the novel word meaning. While eye-tracking can point to more fine-grained indices of learning (e.g., what parts of the texts were mostly relied on for recovering meanings of novel words), in this paper we only considered information that was immediately relevant to our central question: does emotion influence novel word learning?

As a global measure of the reading effort, we considered total passage reading time. Longer reading times suggest more attentive reading of the entire passage. In the web-based Experiments 1O, 2O and 3O, we also used a measure of total passage reading time. This measurement was defined as a time interval between the presentation of the text and the participant-initiated transition to the next trial.

The measures of processing times played a dual role in the analyses. First, we investigated the role of context emotionality on the eye-movement measures and the global passage reading times. This was done to evaluate the role of emotion in the real-time process of novel word learning, from the first to the last exposure to that novel word. Second, we used passage reading times (registered either through eye-tracking or via timestamps in the web-based experiments) as a control predictor of the scores in vocabulary post-tests. As proposed in the Noticing hypothesis (Schmidt, 1990) and demonstrated in a number of empirical studies (see the Introduction), a longer processing time of a passage points to a greater allocation of attention to that passage and may translate into better semantic knowledge and better performance in the post-tests.

In eye-tracking studies 1L and 2L we also investigated the role of exposure on reading behavior. We examined how the first to the fifth repetitions of the word impact the reader's eye-movements. We note that all vocabulary post-tests were administered after the learning phase (i.e., after five exposures to each word), thus it is impossible to analyze the effect of individual exposures on the word's orthographic or semantic knowledge. All post-test analyses necessarily tap into a cumulative effect of all exposures.

*Word knowledge:* Additional dependent variables stem from the vocabulary post-tests. These include accuracy scores from the orthographic choice test as well as the scores received in the semantic definition prompting and definition matching tests. Furthermore, valence ratings to novel words further evaluate the existence and strength of the semantic prosody effect (see above).

A critical manipulation of this study is emotionality of contexts in which a novel word occurs. We expected the valence of context to propagate to the semantic connotation of the novel word and have an effect on both the eye-movements to the novel word and context during the learning phase (in Experiments 1L and 2L) and on the scores in vocabulary post-tests. There are several ways of quantifying this effect. One possibility is to treat context valence as a categorical variable with three levels (negative, neutral, and positive): this variable would reflect our intended manipulation of the context. Another possibility is to tap directly into the valence ratings of a novel word: this variable is a more immediate index of how our manipulation of context actually influenced representation of the target word. This metric allows for estimating how the emotional connotation learned by a specific participant in a specific novel word affects their individual reading effort and semantic knowledge of that word. We found the latter operationalization of valence to be statistically more acceptable than the former. In all regression models (below), individual valence ratings performed better as predictors of dependent variables than a categorical three-level variable. Individual valence ratings are also conceptually more attractive because they estimate an effect of positivity when a novel word was

actually learned as positive rather than when we as experimenters expected it to be learned as positive. Thus, the critical predictor of the reading and learning of a novel word was defined as an individual valence rating to that novel word. Importantly, while the categorical treatment of contextual affect showed less reliable results, those results were consistent with the continuous treatment of contextual affect.

Additional independent variables included the number of exposures to the target word (an ordinal number from 1 to a maximum of 5 exposures) and individual scores in the Author Recognition Task in Experiments 1L and 2L.

**Statistical Considerations** We applied generalized linear mixed effect regression models (using the lme4 package version 1.1–21, Bates, Mächler, Bolker, & Walker, 2015) in the statistical software environment R version 3.6.1 (R Core Team, 2019). Although dependent variables changed by task, all models had a similar structure and series of comparisons across tasks. The Gaussian distribution family was used in models fitted to continuous variables, the Poisson family to count variables (e.g., test scores ranging from 0 to 3), and the binomial family to binary variables. Optimizer bobyqa was used in the latter two types of models to reduce processing time. We conducted post hoc comparisons between specific conditions and examined interactions using cell-means coding and the glht function in the multcomp package (Hothorn et al., 2015).

*Random Effects*

In all models, we included by-subject, by-word, and by-denotation random intercepts. The random intercept for the denotation is one more random intercept than standard crossed effects of subjects and items in single word processing studies (Baayen, Davidson, & Bates, 2008). The justification for this random intercept of denotation is however the same as for items. In this study, the stimuli are particularly complex. We are sampling from two populations of stimuli, novel words and denotations, and intend for our results to generalize to both other denotations and other nonwords not included in this study. Additionally, we tested for a by-subject random slope of the continuous treatment of valence during model comparisons, as recommended for achieving an optimal Type I error rate (Barr, Levy, Scheepers, & Tily, 2013).

*Fixed Effects and Model Comparisons*

We fitted six models to each pairing of an experiment and a dependent variable from the vocabulary post-test (e.g., definition prompting score in Experiment 1L). Model 1 had novel word valence represented as a categorical variable mirroring the experimental manipulation of the contexts. Model 2 added a continuous rating of valence obtained for a specific novel word from a specific participant in the experiment. Model 3 removed the categorical operationalization of valence and only retained the continuous one, modeled as a linear predictor. Model 4 added the continuous variable as a by-subject slope in the random effects structure of the model. Model 5 tested potential non-linearity of the valence rating by modeling it as a quadratic parabola. Finally, model 6 tested non-linearity as a natural cubic spline with 3 knots. All models contained the same set of control predictors (mean passage reading time per novel word), session (1 or 2). All factors used treatment (dummy) coding with a negative condition as the baseline category. Model comparisons were conducted using the log-likelihood ratio test implemented as anova() function in R. A comparison of models 1–3 indicated which operationalization of novel word valence improved performance of the model, while a comparison of models 4–6 indicated if there was an effect of valence as a continuous predictor and whether it was linear or nonlinear.

**Power analysis.**

To decide on the number of participants for web-based Experiments 1O, 2O and 3O, we conducted two power analyses using the simr R package (Green & MacLeod, 2016) for Monte Carlo simulation mixed-model power analysis. We tested effects of condition (i.e. the difference between a baseline negative condition and the positive condition) varying between the original observed effect sizes in Experiment 1L, and effect sizes 3/4, 1/2 and 1/4 as large ( $d = .3, .225, .150, .075$  in

definition prompting,  $d = .4, .3, .2, .1$  in definition matching), for the semantic knowledge tasks. We assumed an equal attrition between the lab- and web-based experiments, when determining how increasing the number of participants increases power. We also assumed residual variances of the random intercepts of novel word, participant, and denotation to be equal in the web-based experiment.

For definition prompting, effect sizes from 1/4 to the originally estimated effect size yielded powers of 20.00% (17.56, 22.62), 66.00% (63.58, 69.52), 97.00% (95.74, 97.97), and 100% (99.63, 100) at 117 participants. For definition matching, effect sizes yielded powers of 65.60% (65.60), 87.27% (87.70, 89.67), 96.60% (96.60, 97.63), and 99.60% (98.98, 99.89) at 117 participants. Thus, under the assumptions given above, we estimate good power (of over 80% and higher) to detect effect sizes of condition half as large in the definition matching task, and good power to detect effect sizes three quarters in size in definition prompting, at 117 participants. We therefore set our sample sizes at 120 participants. Because of the removal and attrition of participants in Experiments 1O and 2O, the sample sizes were somewhat smaller ( $N = 113$  and  $104$ , respectively), but our estimates indicate that they still provide a higher than nominal power (>80%) for an effect size that is half as small as the one observed in Experiments 1L and 2L. In our replication Experiment 3O, we set a larger sample of 120 but did not conduct a second session, resulting in 118 participants with usable data.

**3. Results and discussion**

We begin with summarizing the results of the post-tests, and further zoom in on the eye-movement measures of real-time text processing. Table 2 reports descriptive statistics of the outcome variables of all vocabulary post-tests and select numeric predictors.

*Vocabulary post-tests*

*Valence ratings:* After the learning phase, participants gave valence

**Table 2**

Descriptive statistics of vocabulary post-tests and predictors, presented by experiment.

Experiment	Variable	mean	sd	median	min	max
Exp 1L	Valence rating	5.42	2.01	5.00	1.00	9.00
Exp 1O	Valence rating	5.44	2.03	6.00	1.00	9.00
Exp 2L	Valence rating	5.25	2.00	5.00	1.00	9.00
Exp 2O	Valence rating	5.59	1.96	6.00	1.00	9.00
Exp 3O	Valence rating	5.90	1.87	6.00	1.00	9.00
Exp 1L	Orthographic choice accuracy	0.75	0.43	1.00	0.00	1.00
Exp 1O	Orthographic choice accuracy	0.81	0.39	1.00	0.00	1.00
Exp 2L	Orthographic choice accuracy					
Exp 2O	Orthographic choice accuracy	0.86	0.35	1.00	0.00	1.00
Exp 3O	Orthographic choice accuracy	0.94	0.23	1.00	0.00	1.00
Exp 1L	Matching score	0.41	0.49	0.00	0.00	1.00
Exp 1O	Matching score	0.51	0.50	1.00	0.00	1.00
Exp 2L	Matching score	0.45	0.50	0.00	0.00	1.00
Exp 2O	Matching score	0.47	0.50	0.00	0.00	1.00
Exp 3O	Matching score	0.58	0.49	1.00	0.00	1.00
Exp 1L	Prompting score	1.04	1.25	0.00	0.00	3.00
Exp 1O	Prompting score	1.89	1.20	2.00	0.00	3.00
Exp 2L	Prompting score	1.22	1.26	1.00	0.00	3.00
Exp 2O	Prompting score	1.80	1.26	2.00	0.00	3.00
Exp 3O	Prompting score	1.97	1.24	3.00	0.00	3.00
Exp 1L	Passage reading time, log10	3.59	0.10	3.60	3.35	3.85
Exp 1O	Passage reading time, log10	4.93	1.07	4.87	0.34	11.04
Exp 2L	Passage reading time, log10	3.69	0.18	3.69	3.03	4.40
Exp 2O	Passage reading time, log10	5.11	1.14	5.15	0.64	11.88
Exp 3O	Passage reading time, log10	4.90	1.07	4.88	0.69	8.94



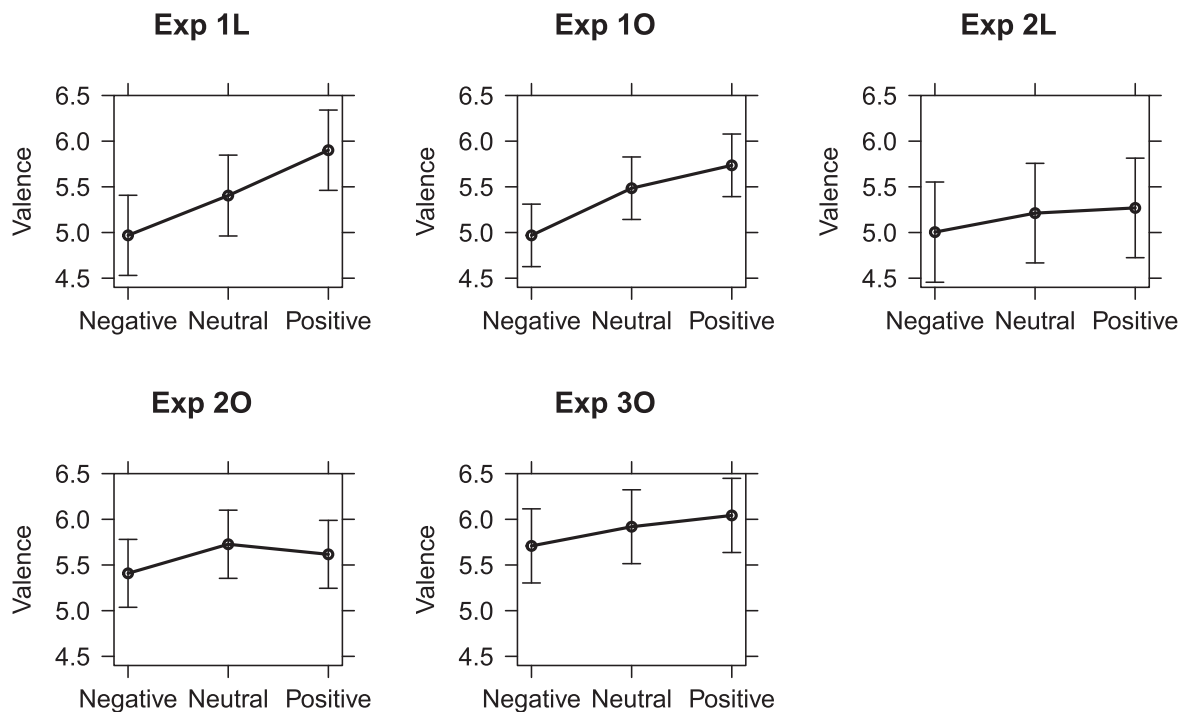


Fig. 2. Partial effects of experimental condition on valence ratings to target novel words per experiment. Error bars represent the 95% confidence interval.

ratings to the novel words they learned, and the test was repeated a week after, in Session 2. Participants indicated whether they had seen a stimulus before and, if so, provided a valence rating to the stimulus. We only present analyses of responses to seen novel words. A total of 558, 1647, 612, 1071, and 1062 responses were recorded in Experiments 1L, 2L, 1O, 2O, and 3O, with accuracy rates of 88, 91, 90, 93, and 90%, respectively.

Fig. 2 summarizes valence ratings to novel words presented in negative, neutral, and positive contexts. The set of passages used in both Experiments 1L and 1O elicited clear differences in valence ratings (1L: negative  $M = 4.95$ , neutral  $M = 5.45$ , and positive  $M = 5.87$ ; 1O: 5.04, 5.53, 5.77), see Fig. 2 top panels. Differences between valence ratings associated with experimental conditions were relatively subtle, amounting to about 0.8 points (or 10%) on a 1–9 point scale between the negative and positive conditions. Regression models confirmed differences between all conditions as reliable in Experiments 1L, 1O (see Tables S5–S9 in the supplementary materials). The small effects are perhaps not surprising given that readers learn emotional connotations of novel words solely from the context and over few exposures; this is also in line with the reported subtle nature of semantic prosody in existing words (e.g., *cause* vs *produce*, see the Introduction). We remind the reader that valence ratings to target words presented in isolation were relatively low (range 2.44–4.22). Yet our manipulation increased perceived positivity of these words in all conditions, and especially in the positive condition. This strongly suggests that the out-of-context valence ratings of novel words are not indicative of the affect that a linguistic context may impose on that novel word. It is also worth noting for the interpretation of results in Experiments 2L, 2O (below) that differences between conditions in valence rating of the nonwords are attenuated compared to the difference in valence between the sentences.

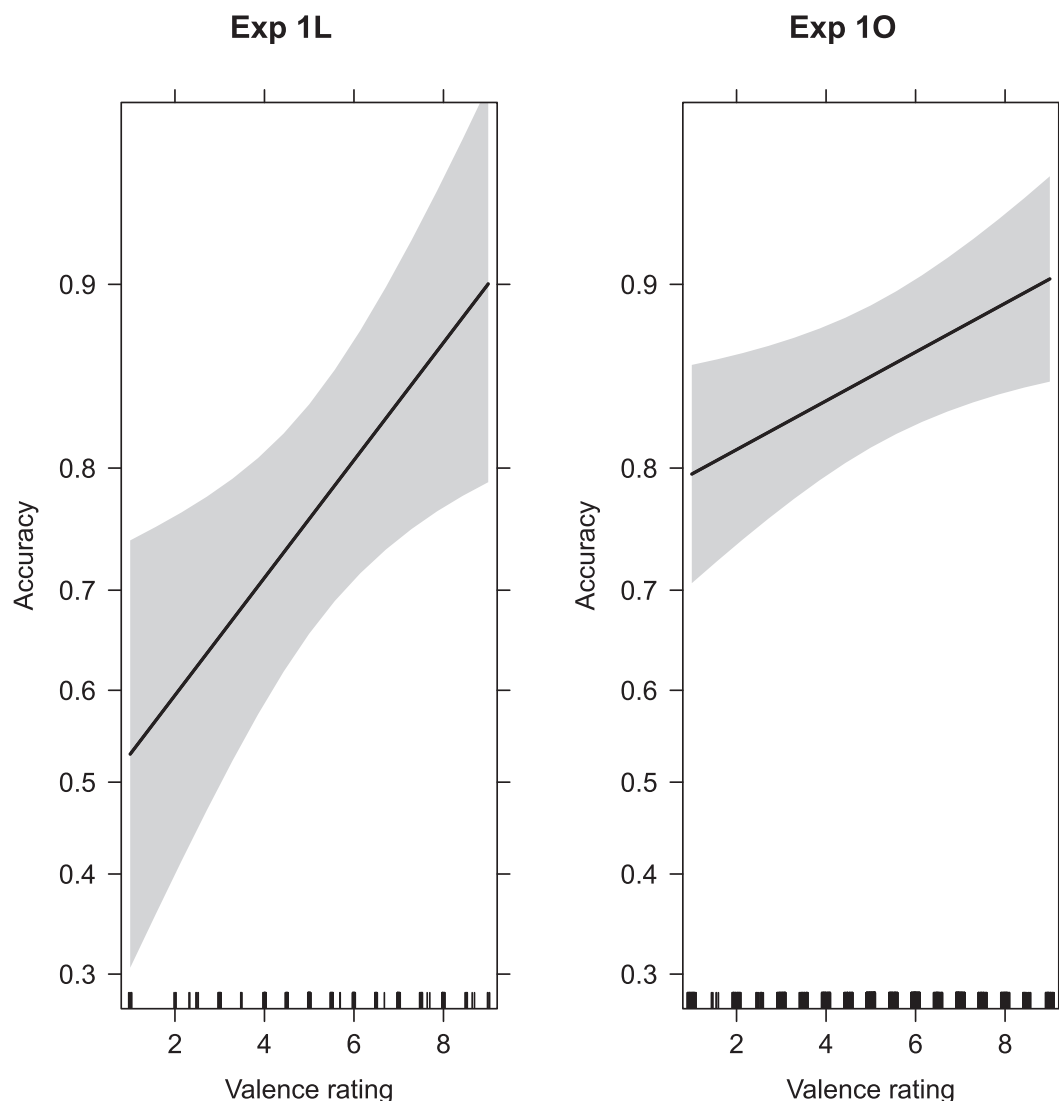
The set of passages used in both Experiment 2L and 2O failed to elicit differences in valence ratings to novel words as a function of emotionality of context in which they occurred (Fig. 2 bottom panels). The only difference that reached statistical significance was the one between the negative ( $M = 5.42$ ) and neutral ( $M = 5.75$ ) conditions in Experiment 2O. Clearly, the nature of the linguistic contexts in which novel words are embedded is crucial for whether emotionality of

contexts propagates to those words through semantic prosody. As noted above, the stimulus set used in these experiments has the least variability in valence of the three experiments. As the effects of context on learned nonword connotation are subtle, it could be that an initial strong emotional contrast is required for semantic prosody to emerge. We elaborate on this and other possible explanations for this null effect of experimental condition in the General Discussion.

Our online study Experiment 3O, using the more tightly controlled stimuli, elicited differences in valence ratings as a function of experimental context. (1L: negative  $M = 5.69$ , neutral  $M = 5.92$ , and positive  $M = 6.1$ ). This amounts to .42 points or about 5% of the 1 to 9 scale. Regression models, taking into account random effects of nonword, participant, and the intended meaning for the nonwords, confirmed the difference between negative and positive, and negative and neutral, conditions as significant (see Table S9 in the supplementary materials). Thus, we once again observed semantic prosody when emotional contrasts between the stimuli are larger. Control predictors showed reliable effects in some of the models (see Tables S5–S9 in the supplementary materials). The only consistent pattern was that participants in the lab-based Experiments 1L and 2L showed higher valence ratings in the second session, a week after the learning phase; no such effect was found in web-based Experiments 1O and 2O.

**Orthographic choice:** A total of 268, 1648, 1036, and 959 valid responses were recorded in Experiments 1L, 1O, 2O, and 3O: orthographic choice data for Experiment 2L are unavailable due to a programming error. The overall accuracy of discriminating seen novel words from their unseen homophones was relatively high: 81%, 81% and 86% and 94%. This accuracy was somewhat lower than in the valence rating task. This is likely because the foils in the orthographic choice task were homophones of seen items, while the foils in the valence rating task were not.

Valence ratings to novel words influenced orthographic choice accuracy in Experiments 1L and 1O, see linear effects in Fig. 3. Orthographic knowledge was better for words that elicited higher positivity ratings. No effect of valence was found in Experiments 2O or 3O. Regression models reported in Tables S10–S13 in the supplementary materials additionally reveal a consistent role of total passage reading time: longer inspection times led to a strong advantage in orthographic



**Fig. 3.** Partial effects of valence ratings to target novel words on orthographic choice accuracy in lab-based Exp 1L, left and web-based Exp 1O, right. The 95% confidence interval is shown.

choice accuracy in all Experiments. This supports the Noticing hypothesis (Schmidt, 1990): increased attention to the text that includes the novel word and cues towards its meaning leads to better learning outcomes. Furthermore, orthographic knowledge deteriorated between the two testing sessions separated by a week. Accuracy of discriminating seen words from homophonic foils was reliably lower in Session 2. No interactions were observed between valence and session.

Since some novel words did not receive valid valence ratings in the respective task, we additionally examined whether these words constitute a systematic pattern of missing data i.e., if fewer words in a particular condition receive valid valence ratings. We observed no significant relationships between condition and whether a word received a valence rating or not, as determined by a binomial generalized linear model (not shown) predicting whether a nonword received a valid valence rating or not as a binary variable.

**Definition matching:** This task tested meaning recognition for the novel words, by asking to match seen words with their intended meanings and to ignore unseen words and definitions. Below we report responses to seen novel words. A total of 511, 1588, 477, 1018 and 959 responses were recorded for Experiments 1L, 1O, 2L, 2O and 3O, respectively.

Fig. 4 reveals strong non-linear or linear effects of valence ratings on definition matching score in all experiments: regression models confirm

these effects as reliable (see Tables S19–S23 in the supplementary materials). In all cases, semantic knowledge was stronger for novel words that were judged as more positive. For these positively judged words, the likelihood was higher for a participant to match a form and a definition correctly. In three experiments of out of five, the non-linear functional form of the effect suggests that most advantage in semantic knowledge occurs in novel words with very high positivity. In Experiments 2L and 2O the effect is linear and characterizes the entire valence range. All effects were of a similar magnitude. It is worth a reminder that the baseline performance in this test is 5.6% if forms and definitions were matched randomly. Thus, even the lowest estimated performance in this task at around 20–30% correct suggests that semantic knowledge is successfully acquired and participants were not guessing randomly. Remarkably, their likelihood of a correct match was estimated to increase by another 30% (to the 60–70% success rate) if they encoded a given novel word as highly positive.

Additional effects of control variables were as follows (see regression models in Tables S19–S23 in the supplementary Materials): in all experiments except 2L, a longer inspection time of the passages with novel words led to a significant increase in the definition matching score. As argued above, the amount of noticing, which is reflected in a greater allocation of attention to the context and novel word, improves semantic knowledge of that word. Moreover, web-based 1O and 2O

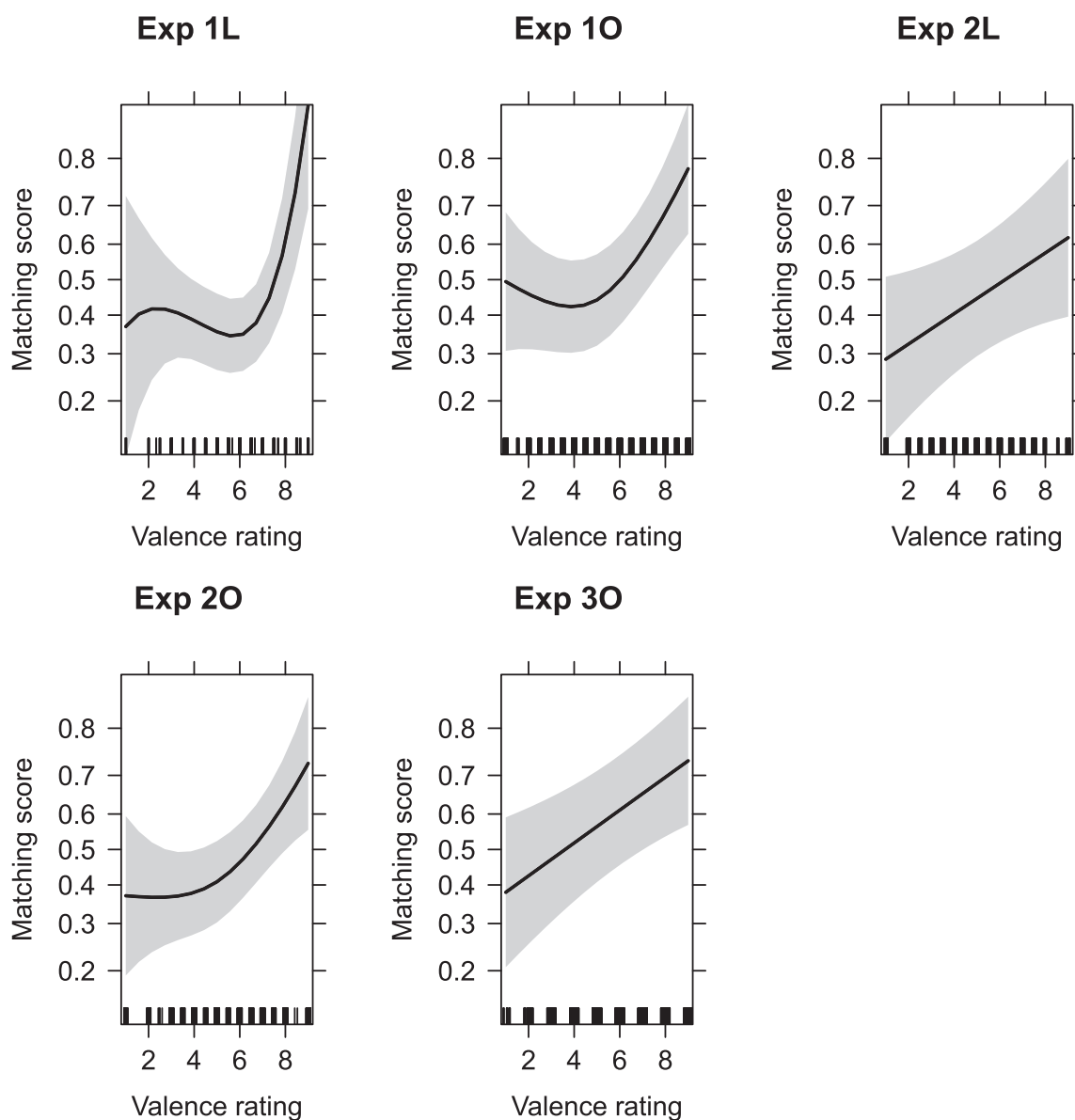


Fig. 4. Partial effects of valence ratings to novel words on definition matching scores per experiment. The 95% confidence interval is shown.

showed a strong negative effect of session on the outcomes. Their memory for the novel words' semantics was weaker in Session 2, a week after learning. Lab-based experiments did not show this difference. Our argument rests on semantic prosody yielding stable gains for word learning. A significant session by valence interaction allows us to test if gains from learning words in positive contexts are undone by a week long gap. For our argument to hold, we should see a general degradation of performance, regardless of contextual affect. This would indicate that while overall learning suffers over time, positive context still boosts learning. To test this, we added additional models interacting session with the best-performing functional shape of the valence effect in each study of 1L, 10, 2L, and 20. None of these interactions reached significance or significantly improved model fit in an additional model comparison. Experiment 1L additionally showed a positive effect of ART scores: individuals with a greater exposure to print demonstrated improved semantic knowledge, reflected in higher definition matching scores.

**Definition prompting:** This task tested meaning recall and the depth of semantic knowledge in an expressive way, without any explicit meanings to discriminate between. We report responses to seen novel words below. A total of 520, 1597, 479, 1016, and 959 responses were

recorded for Experiments 1L, 10, 2L, 20, and 30 respectively.

Fig. 5 visualizes results from all experiments (1L, 10, 2L, and 30) in which an effect of valence ratings on definition prompting scores was found. In these four experiments, words that were rated as more positive came with higher scores in the definition prompting task, and thus revealed a better meaning recall ability. In the lab-based Experiment 1L, the effect was non-linear and mainly confined to very positive values of valence, while in Experiments 1L, 2L, and 30 the effect was linear and characterized the entire valence range; no reliable effect of valence was found in Experiment 20 (see Tables S14–S15 in the supplementary materials for regression models). The baseline level of performance, expected if the task is done at random, is a score of 0. All experiments demonstrated an above-random performance, suggesting that semantic learning took place: the lowest model-estimated score was around 0.7 points for lab-based Experiments 1L and 2L and 1.5 for Experiment 10 and 30, on a scale from 0 to 3 points. The much higher overall performance in the web-based task compared to the lab-based one is likely due to differences in their administration (with and without human feedback to each response, see the Methods). Importantly, in both web- and lab-based studies, the maximum difference in valence translated into a substantial difference in the definition

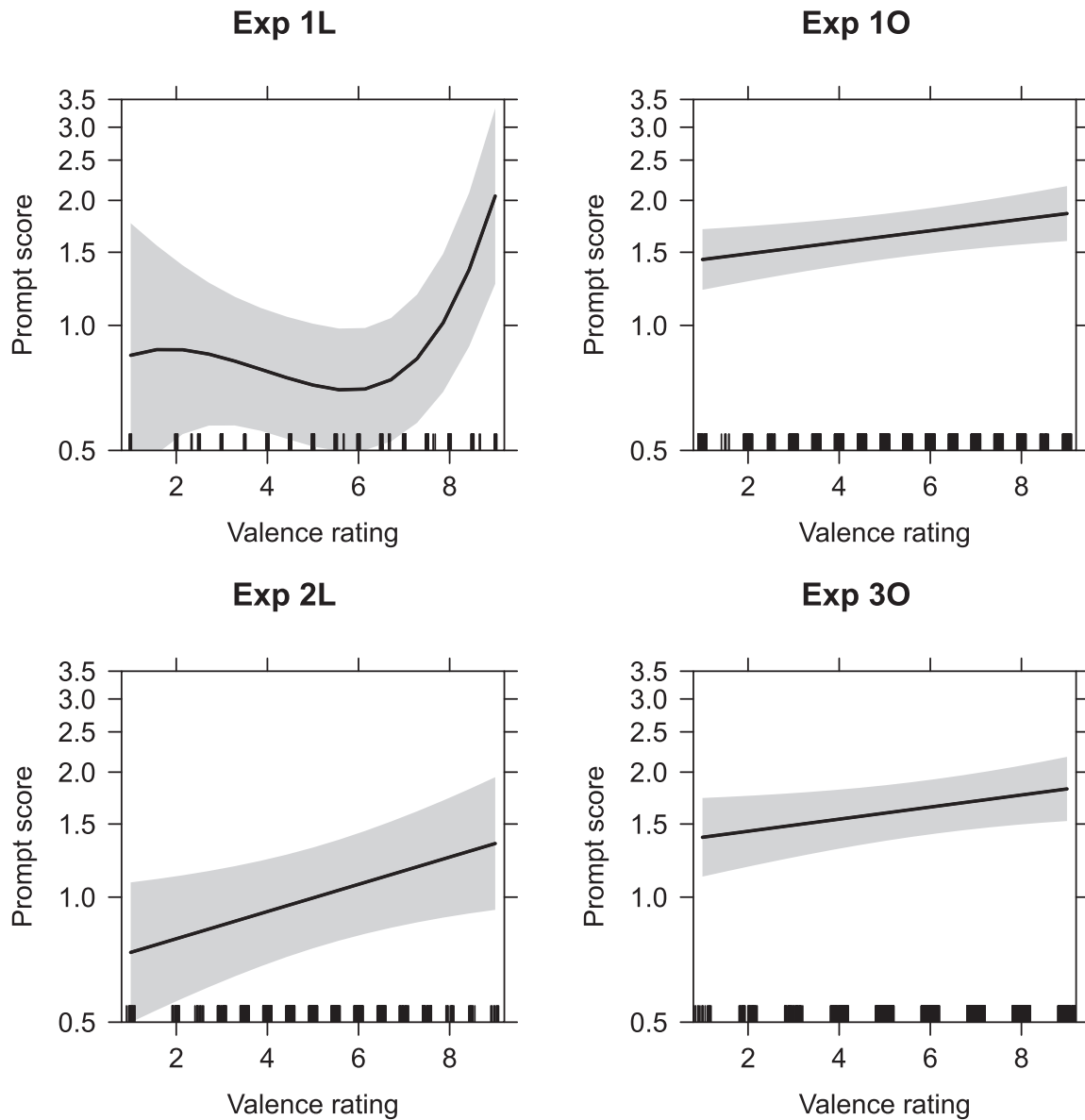


Fig. 5. Partial effects of valence ratings to novel words on definition prompting scores per experiment. The 95% confidence intervals are shown.

prompting scores: on the order of 0.7 points (23% of the scale) in 1L and 1O, 0.3 points (10% of the scale) in 2L, and .25 points in Experiment 3O (8% of the scale). As with the meaning recognition, the meaning recall was superior in those novel words that were encoded with a more positive connotation. The relatively weak effect in Experiment 2L was not observed when the definition prompting task was administered to a much larger sample of participants in 2O: it is possible that the effect in 2L is an artifact of an under-powered experiment.

Additional control variables showed some influence in the web-based Experiments 1O and 2O: specifically, definition prompting scores were reliably lower in Session 2, separated by one week from the learning phase. Session did not have an effect in Experiments 1L and 2L, nor did it interact with valence. Thus the positivity advantage was maintained. Inspection times for the passage did not consistently affect definition prompting scores either. Similarly to the definition matching task, we observed a strong main effect of the ART score on the definition prompting score in Experiment 1L: more experienced readers were able to better recall definitions for novel words.

*Eye movements:* The use of eye-tracking enabled us to monitor how the effects of emotion on word learning developed in real time, during the readers' exposure to the novel words in context. Table 3 reports

Table 3

Descriptive statistics of eye-movements: Experiments 1L and 2L.

	mean	sd	median	min	max
Exp 1L: Total fixation time	450.54	266.62	369.50	86.00	1941.00
Exp 1L: Skip rate	0.06	0.23	0.00	0.00	1.00
Exp 1L: Regression rate	0.31	0.46	0.00	0.00	1.00
Exp 2L: Total fixation time	730.05	593.98	538.00	85.00	3964.00
Exp 2L: Skip rate	0.08	0.27	0.00	0.00	1.00
Exp 2L: Regression rate	0.42	0.49	0.00	0.00	1.00

descriptive statistics of eye-movement dependent variables.<sup>2</sup>

In Experiment 1L, we removed one out of 36 participants because of excessive blinking. The initial data pool contained 25950 observations. For the analysis of target words we restricted the data pool to eye-movements to the target words, yielding 1596 data points. We further

<sup>2</sup> We note that the eye-tracking portion does not suffer in terms of statistical power compared to post-tests, as the trial level eye-tracking includes 5 observations per nonword per participant.



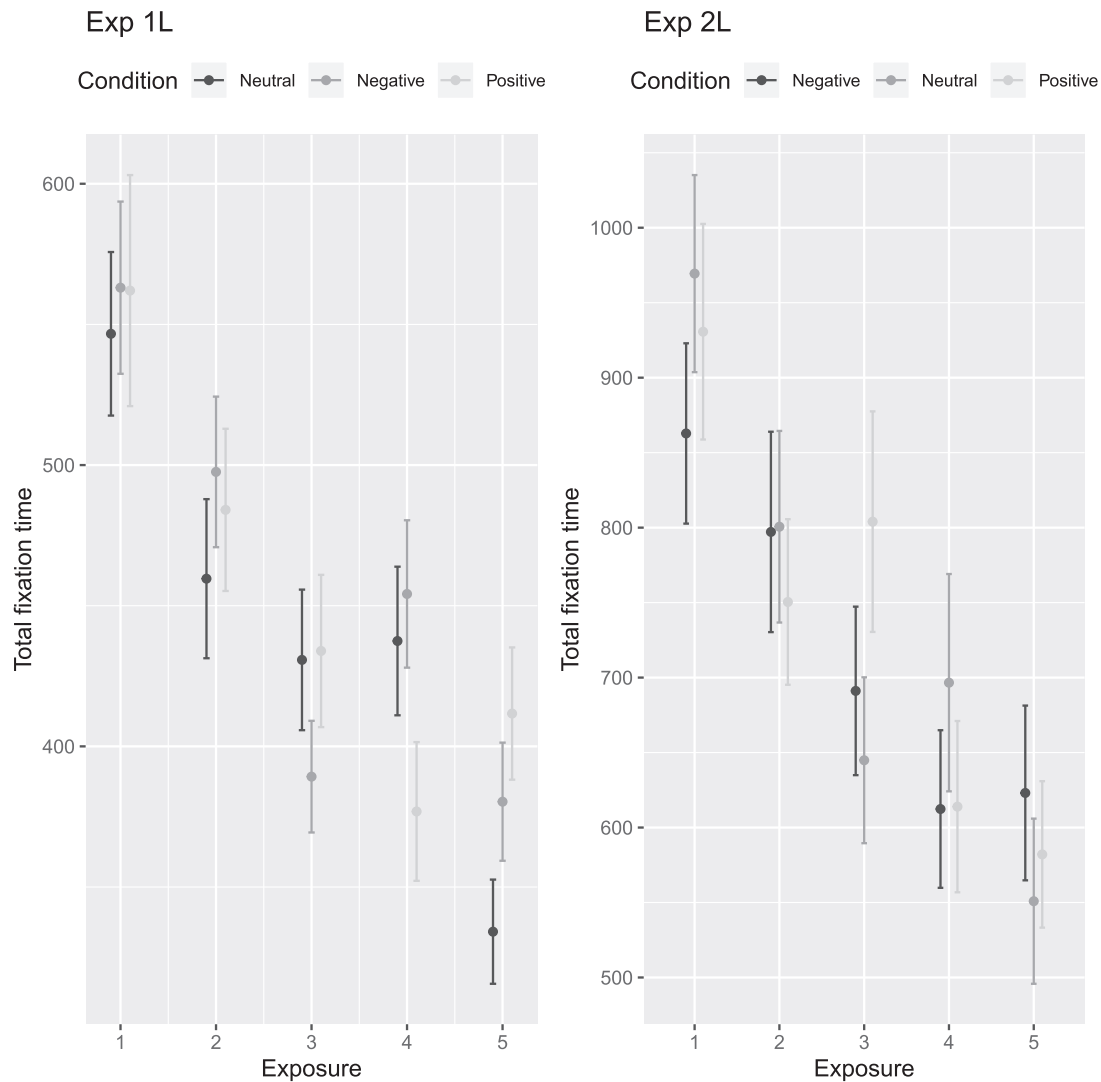


Fig. 6. Total fixation time per condition and exposure in Exp 1L (left) and Exp 2L (right). Error bars stand for the 95% confidence interval.

removed trials where target words elicited fixations shorter than 80 ms or longer than 1000 ms, or more than five fixations. The resulting pool contained 1390 data points. After restricting the data pool of Experiment 2L to target novel words and applying trimming procedures as described in Experiment 1, there remained 1355 data points in the cohort of 33 readers. Table 3 reports descriptive statistics of the eye-movement measures.

For the ease of interpretation, below we presented the effect of context emotionality as that of categorical variables with three levels (negative, neutral, and positive); the results are very similar if we use the continuous valence ratings instead. Fig. 6 summarizes total fixation durations on novel words across conditions of experimental condition and five exposures to the word (labeled here as factor Exposure) in Experiments 1L (left panel) and 2L (right panel). A few patterns emerged from the data. First, there was a substantial speed-up in total fixation times on novel words as a function of the number of exposures on the order of 200 ms. Second, no condition showed a consistent advantage over others across all exposures. To take as an example the positive vs the negative condition in Experiment 1L (Fig. 6 left panel), the first two exposures did not elicit a noticeable contrast between them. The remaining three exposures saw a significant advantage in processing speed for the negative condition over positive (exposure 3), that of a positive condition over negative (exposure 4) and, finally, that of the negative condition over positive again (exposure 5). A similar

fluctuating pattern characterized contrasts between all other conditions and in both eye-tracking studies.

Regression models (reported in supplementary materials S24 and S25) uncovered a strong main effect of Exposure, an unreliable main effect of experimental condition, and a reliable Emotion  $\times$  Exposure interaction. Additional models fitted to total fixation times at individual exposures (3, 4, and 5) supported the notion that, even when statistically significant, the contrasts flip their direction and do not constitute a consistent pattern that would point to a processing advantage of one condition over others. Similarly inconsistent were eye-movement patterns at different levels of Emotion observed in models fitted to regression rates and skipping rates (not shown). We concluded that context emotionality did not influence the real-time aspects of novel word learning in any systematic way.

In sum, intriguingly, eye-movements as online indices of word learning did not point to any impact of context emotionality, even though this impact was evident in semantic knowledge post-tests in Experiments 1L and 2L. We return to this apparent paradox below.

#### 4. General discussion

The present paper addresses two under-explored aspects of language learning: how do people learn to associate novel words with affective characteristics of their context, and how do these acquired affective

connotations influence the quality of learning the words' forms and meanings? We conducted five experiments with three stimuli sets in which proficient readers performed intentional learning of English novel words in context. The critical manipulation of each experiment was emotionality of context: for a given participant, each novel word consistently occurred in a positive, neutral or negative context (see Examples 1–3 in the Introduction). All novel words (e.g., *plurk*, *ceammy*) had an emotionally neutral and relatively concrete intended meaning or denotation (e.g., a tool, a kitchen utensil). We investigated whether the affect of context can transfer to a novel word with an emotionally neutral intended denotation: i.e., whether a few exposures to a novel word are sufficient to induce semantic prosody. Using a range of experimental tasks, we also tested the impact of emotionality on the real-time behavioral indices of learning during an initial exposure to the novel words (gauged via the readers' eye-movements) and the time spent on reading each text (chronometric measures), as well as on their acquisition and retention of formal and semantic knowledge immediately after the exposure and after one week (using vocabulary post-tests).

In three stimuli sets, emotionality of linguistic contexts of novel words was manipulated such that a participant saw a given novel word in consistently positive, neutral or negative contexts. Experiments 1L and 1O both used the first stimulus set, Experiments 2L and 2O the second stimulus set, and Experiment 3O used a set with tight matching and control of passage structure and semantic similarity to intended meanings. Experiments 1L and 2L incorporated eye-tracking during the learning phase of reading passages with novel words and were administered to relatively small samples ( $N = 36$ ); Experiments 1O, 2O, and 3O were administered online (without an eye-tracking component) to much larger samples ( $N > 100$ ) and served as a validation with high statistical power.

Our findings revealed that semantic prosody – or the transfer of affect to a word from its context – does take place even after few exposures to originally meaningless novel words. As well, by performing a second post-testing session one week after, we found that the positivity advantage remains. While overall performance was sometimes (in the online experiments) degraded in the second session, this decreased performance was unrelated to the learned affect. Positive words still elicited better performance in session two, even if overall performance was lower. We also uncovered evidence that positive semantic prosody gives a greater boost to intentional learning of novel words, compared to neutral or negative prosody, though this effect may be moderated by stimulus complexity and stimulus valence. The remainder of this paper presents each finding and pits them against existing accounts of cognitive and affective dimensions of novel word learning.

*Semantic prosody:* Experiments 1L and 1O (using the same set of passages) as well as Experiment 3O showed a robust effect of manipulated context emotionality on valence ratings, see Fig. 2. Words appearing in positive contexts were judged to be more positive in the valence rating task than words learned in neutral contexts, and both elicited higher valence ratings than words learned in negative contexts. For a meaningless string of characters to display properties tightly linked to its affect, it is a necessary condition that it acquires an affective connotation, i.e., its representation in one's semantic memory develops an association with some affective value. There may be at least two possibilities for this association to form. First, the learned literal meaning of the novel word may be a concept or object associated with a strongly valenced connotation (e.g., a type of disease or a type of food). This mechanism is unlikely to account for affect-related results in our study, because all intended denotations were fairly neutral (a tool, a kitchen utensil etc.). We do, however, find evidence that the affective content of linguistic contexts transfers to novel words and determines the polarity of their connotations. Thus, we are some of the first research showing that novel words can absorb some of the affect that their contexts convey: that is, they demonstrate semantic prosody well established in corpus-linguistic and psycholinguistic studies of existing

words (Hunston, 2007; Sinclair, 1996, 2004; Sneffjella & Kuperman, 2016; Winter, 2016). Remarkably, the transfer of affect from context to a novel word can occur over as few as five exposures made in the written modality which is devoid of additional affective cues that other modalities afford (e.g., gestures, facial expression, speech prosody, or pitch). This transfer persists one week after the learning phase, showing that it is not short-lived.

These findings indicate an initial point of a semantic learning trajectory that leads to the well-described phenomenon of semantic prosody in relatively stable lexical representations of known words (see review in Winter, 2016). Our study is not designed to pin down the exact nature of the mechanism that underlies semantic prosody, but we outline two candidate theories, subject to future validation. As mentioned in the Introduction, one mechanism might be embodiment of a (novel) word's meaning in emotional and sensorimotor experiences during the learning of the word. In our studies, every encounter with a novel word took place in a linguistic context containing several words with extreme or neutral valence and the majority of emotionally neutral words. In line with the embodied cognition account, observed semantic prosody can be explained if the affective content of the context was encoded in the semantic memory together with the denotation and the form of the novel word. When retrieved from memory in the definition matching or prompting vocabulary post-tests, this affective connotation of the novel word would account for the advantage in semantic knowledge, similar to the positivity advantage in existing words (see below). An alternative mechanism is of a cognitive nature. As Sneffjella and Kuperman (2016) demonstrate in the analyses of written corpora, words tend to co-occur with words of a similar affective polarity: contexts of positive words are relatively positive, and those of negative words are relatively negative. If this statistical bias of language use is part of the rich knowledge that speakers have about their language, they may ascribe an emotional connotation to a novel word that is congruent with the dominant connotation of its context. In our stimuli, that would lead to a lesser or greater degree of positivity ascribed to novel words occurring in positive contexts or negative contexts. A more definitive answer regarding the mechanism of semantic prosody would likely require an experimental technique that is sensitive to behavioral expressions that are primarily relevant for affective rather than cognitive responses to stimuli such as galvanic skin response or heart rate measurements. Such a study is outside of the present scope of this paper.

Our demonstration of semantic prosody in Experiments 1L, 1O, and 3O has broad implications for the theory of language acquisition, especially under the embodiment account. The emotional aspect of lexical connotations has been proposed as a major vehicle for learning of abstract words (e.g., *sophisticated* or *gullible*, see among others Borghi et al., 2017; Kousta et al., 2011; Ponari et al., 2018; Sheikh & Titone, 2013). Since such words do not denote tangible objects or events in the material world, one view is that emotional cues arguably provide grounding to symbolic abstract meanings that – unlike their concrete counterparts – they cannot obtain from sensorimotor cues. This grounding enables crisp lexical representations to form even when meanings have to be learned from language rather than from physical experience. Another view is that there may be still be a degree of sensorimotor involvement in the grounding and learning of abstract concepts. Emotional experience with an abstract concept (e.g., “catastrophe”) may include knowledge about how catastrophes “feel”, including such sensorimotor experiences as increased heart rate or alertness (e.g., Kiverstein & Miller, 2015; Newcombe et al., 2012; Siakaluk et al., 2016)<sup>3</sup>. Under either view, however, emotional information is key to learning abstract concepts.

This account begs a question: how do people learn to associate new abstract words with affective values in the first place? To our

<sup>3</sup> We thank an anonymous reviewer for raising this point.

knowledge, most studies on acquisition of abstract words either consider existing words which have either already acquired their affective connotations through one's experience with language (see e.g., experimental data and analyses of behavioral mega-studies in Ferré et al. (2015), Kousta et al., 2011; Ponari et al., 2018; and a review in Borghi et al., 2017), or are explicitly associated with some degree of affect. Our data in Experiments 1L, 1O, and 3O reveal that affective connotations can be learned without prior semantic knowledge associated with a new lexical item, as long as it consistently appears in an emotionally loaded linguistic context.

The present findings also hint at a link between emerging effects of semantic prosody in novel words, shown here, and the tendency of established words to occur in a company of words with similar affective and sensorimotor connotations (Sneffjella & Kuperman, 2016). When learned, words with emotionally neutral connotations will tend to occur in a broad variety of contexts, some positive, some negative and some neutral. Temporarily, neutral words may acquire a degree of valence through the chance of co-occurring with valenced contexts. Yet through repeated exposure, the aggregated emotionality of contexts for such words will likely be neutral. Even a small bias towards emotionally congruent contexts for valenced (positive or negative) words will accumulate and – over sufficient exposure – lead to formation of similarly valenced stretches of texts, in line with corpus-based observations of Sneffjella and Kuperman. Thus a “steady state” system of connotations may arise from the very fast and malleable absorption of context emotionality reported here.<sup>4</sup>

Another important finding is that while semantic prosody is demonstrably possible it is not always present. The second stimulus set of passages (Experiments 2L and 2O) did not lead to a reliable effect of our context manipulation on valence ratings to novel words in the post-vocabulary tests, see Fig. 2. In our view, the null effect suggests that the affective influence on novel word learning interacts with other linguistic dimensions of context that are known to have an independent impact on word learning. As described above, this set of passages was derived from the one used in Experiments 1L and 1O and edited to be both longer and more readable (e.g., had higher frequency words and simpler syntactic structures). Some additional words not semantically related to the intended denotations were included in some passages, for the priming task that we were forced to discard due to programming error. Both context length and complexity have been shown to influence the quality of novel word learning (e.g., Bolger et al., 2008; Daneman & Green, 1986; Swanborn & De Groot, 1999; Wochna & Juhasz, 2013): longer and simpler contexts lead to better learning, see the Introduction. It is thus possible that affective cues (and the ensuing semantic prosody) become more relevant when other cues to the meaning of the novel word are less accessible, as was the case in shorter and more complex passages of Experiments 1L, 1O, and 3O. Another discrepancy between the two stimulus sets was revealed in our norming of the valence of passages: passages used in Experiments 2L, 2O were overall more positive (by 1 point on a 1–9 scale) than the respective passages used in Experiments 1L and 1O (negative: 2.47 vs 3.57, neutral: 4.56 vs 5.41, and positive: 6.03 vs 6.78 in Experiment 1). Similarly, compared to passages used in Experiment 3O, passages in 2L, 2O were more positive in the negative condition (2L, 2O: 3.57 vs. 3O: 2.44). Possibly, for semantic prosody to have a noticeable magnitude of behavioral expression, very negative stimuli have to be included in the stimulus set. This was true of Experiments 1L, 1O, and 3O but not Experiments 2L, and 2O.

The exact parameters of linguistic context that enable a greater magnitude of the semantic prosody effect remain to be explored. Yet the null result still provides some testable hypotheses: stimulus complexity and the strength of valence are now hypothesized to moderate the strength of learned semantic prosody. Extreme emotional valence in

contexts impoverished of other cues to meaning should maximize the transfer of semantic prosody. At present, most studies of novel word learning, including this one, only address one or two of the known dimensions: a systematic effort of coordinated multi-factorial manipulation is necessary to establish the relative importance of the causal factors of word learning and their interactions. In this paper, we confine ourselves to reporting a new contextual effect as relevant to vocabulary acquisition and pointing to some dimensions that may cancel it out.

*Positivity and word knowledge:* Vocabulary post-tests demonstrated a relatively high overall performance in tasks tapping into the orthographic knowledge of novel words, as well as meaning recognition (definition matching) and meaning recall (definition prompting). The orthographic choice task showed accuracy between 75 and 86%; the average definition matching scores ranged between 40 and 50% with a chance baseline of 5.6%; and the definition prompting scores ranged from 1 to 1.9 with a chance baseline of 0. Thus, our contexts were informative and numerous enough to allow for successful word learning.

Critically, novel words with positive connotations came with a consistent advantage to novel words in orthographic and semantic tasks relative to words with neutral or negative connotations in Experiments 1L, 1O and 3O. Thus, words that were judged as more positive came with a higher score in the orthographic choice task (Fig. 2), definition matching task (Fig. 4) and definition prompting task (Fig. 5). When the effects were nonlinear, they indicated that most of the advantage came from very positive words, while semantic knowledge of more negative and neutral words did not differ as much. All effects were somewhat weaker in Experiment 1O than in 1L, which is consistent with the notion that properly powered experiments are more accurate in estimating the true effect size whereas effects that come out reliable in under-powered experiments are often overestimated (Hedges & Olkin, 2014). Still, in both experiments the effects were sizable. The more modest estimates of Experiment 1O are that the model-predicted benefit of a novel word with the most positive rating, as compared to a word with the most negative rating, was on the order of 10% in orthographic choice, 20% of accuracy in definition matching task, and 20% in definition prompting task, using the scale of the respective task. This positive advantage held true both in the immediate and the deferred testing sessions. This suggests that the observed impact of affect on semantic memory for novel words is lasting (over one week).

These findings are important for theoretical accounts of word representation and memory. Existing literature emphasizes the difference between valenced (positive and negative) and neutral words in word recognition and recall tasks, yet it offers mixed results as to whether only positive, only negative or all valenced words are easier to memorize and then recall (e.g., Adelman & Estes, 2013; Kensinger & Corkin, 2003; Maratos, Allan, & Rugg, 2000). Our analyses of mega-studies of recognition memory (Sneffjella & Kuperman, 2016) showed a clear advantage in response accuracy in both the words that were relatively positive and the words occurring in more positive contexts. This advantage was also found in the speed and accuracy of lexical decision and naming (Kuperman et al., 2014) and linked to automatic vigilance towards and delayed disengagement of attention to negative words, which communicate threat and danger to one's survival (e.g., Algom et al., 2004; Erdelyi, 1974; Estes & Adelman, 2008a, 2008b; Larsen, Mercer, Balota, & Strube, 2008; Pratto & John, 1991; Wentura et al., 2000; Williams et al., 1996). Our present study is the first to demonstrate that the automatic vigilance account may not only apply to existing words with well-established semantic representations and contextual preferences, but it may also generalize over initially meaningless and affect-less novel words.

As discussed above, our experimental manipulation failed to elicit a transfer of the linguistic context's affect to the connotations of novel words in Experiments 2L and 2O. For this reason, context emotionality would not be expected to engender a difference in the formal or semantic knowledge of novel words in these experiments. This

<sup>4</sup> We thank an anonymous reviewer for raising this point.

expectation was confirmed to a large extent. Valence ratings to novel words did not elicit a reliable effect on the orthographic choice performance in Experiments 2O or 3O. While a reliable effect of valence rating emerged in Experiment 2L in the definition prompting task (words that were judged as more positive came with higher scores), this effect was not replicated in Experiment 2O and might be spurious, due to the low statistical power in 2L.

Somewhat surprisingly, words with more positive ratings showed significantly higher scores in the definition matching task, found both in Experiment 2L and 2O (Fig. 4), similar in shape and magnitude to the effects observed in Experiment 1L, 1O. If experimental condition was used in the regression model as a categorical predictor of emotionality (model not shown), no effect on definition matching scores was observed. This suggests that novel words that were encoded as having positive connotations elicit a boost in at least one aspect of semantic knowledge, i.e., meaning recognition, even in the absence of the effects of the experimental condition on valence ratings to novel words. One possibility is that valence ratings are not sensitive enough to reflect subtle differences in affect transferred from contexts to novel words.

An alternative explanation for the effect would be that participants gave higher valence ratings to the novel words that they felt they learned better, regardless of the context in which those words occurred. While we cannot rule out this explanation, we find that it is not parsimonious. Specifically, it is not required to explain the results of Experiment 1L, 2O, and 3O where differences in valence ratings can be traced back to experimental manipulation. Furthermore, this explanation would require that participants only feel more positive towards novel words if they can recognize their meanings better, but not if they are better at remembering their orthographic form (measured via the orthographic choice test) or at retrieving their meanings (measured via definition prompting).

As with semantic prosody above, we confine ourselves to reporting robust findings – more positive affect learned from the context strengthens formal and semantic knowledge of the novel word – and pointing at other factors that might interact with the proposed mechanisms – word knowledge may also be influenced by affective connotations that do not originate in the linguistic context.

The outcomes of our experiments are noteworthy for the applied issue of how to boost word learning in one's L1. If instructional materials place novel terminology in the contexts that are lexically and semantically diverse but are consistently emotionally positive, learners will respond with an increased level of semantic knowledge and better retention of word meanings over time. Even if the positivity advantage is relatively small, it may lead to substantial gains over time given how much and how frequently people need to learn words (see the Introduction).

*Encoding vs retrieval of novel words:* An intriguing finding in all our studies was that the eye-movement record did not reveal any consistent influence of emotionality conditions during active learning of the novel words. This stands in a stark contrast with the positive advantage observed in vocabulary post-tests of meaning recognition and recall in Experiment 1L 1O, and 3O, see above. It is likely that eye-movements tap into the effort of encoding novel words in memory, while vocabulary post-tests tap into the effort of retrieving them from memory. The present findings suggest that the affective dimension of learning did not influence memory encoding but did influence memory retrieval in the form of a positivity advantage.

We do find ample evidence, however, that the overall reading times for the passages predicted higher scores in orthographic and semantic tasks across all experiments. These findings support the Noticing Hypothesis (e.g., Schmidt, 1990, 2001) which suggests that a greater effort of inspecting the text at hand correlates positively with the amount of selective attention allocated to that text and improves encoding of the novel word's form and meaning. It also corroborates reports in the literature that longer reading times to the words and/or contexts come with higher indices of orthographic or semantic

knowledge (see the Introduction).

#### 4.1. Future directions

We consider the reported experiments to provide initial evidence of a role for semantic prosody in word learning. The evidence presented also raises many other questions. Perhaps the most important issue of these is to determine the conditions under which semantic prosody takes place with novel words embedded in context, and how it interacts with other known linguistic dimensions of word learning (see our discussion of the difference between our experiments 1 and 3 versus 2 above). An experiment manipulating both extremity of valence in the stimuli of the learning context and the complexity of the context could clarify the null result of Experiments 2L, 2O.

As discussed in the Introduction, we set up our experiments in the way that – in our intuition – would emphasize the effect of emotion. Specifically, we made participants aware that they will be tested on word knowledge (an intentional learning paradigm), and we restricted novel words to concrete nouns in salient syntactic positions in the sentence. Further directions may include a change in the learning paradigm. Incidental learning (with participants being unaware of post-tests) is considered to be more difficult for readers (e.g., Jenkins & Dixon, 1983; Konopak et al., 1987; Nagy et al., 1987). We expect that, in the absence of an explicit motivation for learning, the effects of emotionality on performance in formal and semantic knowledge tasks will be weaker. It is, however, possible that a more difficult task may highlight the role of affective cues in the linguistic contexts of novel words, similarly to a stronger role of affective manipulations in more difficult contexts that we observed in Experiment 1 vs 2. Thus, in incidental learning we may see stronger differences in valence ratings to novel words stemming from manipulations of context affect.

We also predict weaker effects of emotionality on semantic knowledge, if novel words are used in less prominent thematic (instrument, location) or syntactic (e.g., attribute, modifier) roles. Yet according to studies on children by Schwanenflugel et al. (1997) and Wagovich and Newhoff (2004) we might expect the learning effects to be stronger for non-nominal parts of speech, such as verbs and adjectives.

Another important extension would be to compare learning of novel concrete and abstract words. As mentioned above, emotional information is argued to play a bigger role in learning of abstract words that are devoid of tangible denotations, as compared to concrete ones (e.g., Borghi et al., 2017; Kousta et al., 2011; Ponari et al., 2018). This notion leads to a falsifiable prediction: positivity advantage to novel words with abstract meanings will be stronger (e.g., forms and meanings will be learned easier and retained longer) compared to words with concrete meanings.

Yet another extension of the present results, suggested by an anonymous reviewer, is to associate novel words with intended meanings that are not affectively neutral as in the present studies (e.g., clothing, furniture) but are typically associated with negative or positive (e.g., weapon vs flower). We predict that congruent cases – when a novel word with an intended positive/negative connotation is learned in positive/negative contexts – will lead to better learning outcomes than incongruent cases – a positive/negative novel word learned in context of an opposite polarity.

Although we focus on L1 word learning in this paper, we acknowledge the importance of word learning in L2. Most people in the world speak multiple languages. For instance, over 50% of Europeans are proficient in at least one foreign language to the degree of being able to conduct a conversation (Eurobarometer, 2006), and over 1 billion people speak English as a foreign language (Crystal, 2012). A similar word learning study in an L2 context is an important continuation of the present work.



## 5. Conclusion

This series of experimental studies one of the first to demonstrate psychological reality of semantic prosody in novel word learning, i.e., the transfer of affect carried by linguistic context onto affective coloring (or connotation) of a word initially devoid of any meaning. We argue that this finding in the first exposures to the word marks an initial point of a learning trajectory that leads to semantic prosody observed in well-established lexical representations of known words. Thus, we show that semantic prosody is a pervasive mechanism by which readers come to associate emotional coloring with word forms and denotations. We further show that positive semantic prosody can lead to a considerable advantage in orthographic and semantic learning, similar to the advantage seen in existing words (Sneffjella & Kuperman, 2016). Meanings of novel words learned in positive contexts are more accessible during retrieval, and lead to a better performance in semantic knowledge tasks both immediately after the learning phase and after one week. To our knowledge, this advantage has not been reported earlier and can be utilized by applied linguists and language educators for improving learning outcomes for vocabulary acquisition.

## CRedit authorship contribution statement

**Bryor Sneffjella:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - review & editing. **Nadia Lana:** Investigation, Resources, Writing - review & editing. **Victor Kuperman:** Supervision, Formal analysis, Writing - original draft.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jml.2020.104171>.

## References

- Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, 129(3), 530–535.
- Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional stroop phenomenon: A generic slowdown, not a stroop effect. *Journal of Experimental Psychology: General*, 133(3), 323.
- Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American Journal of Psychology*, 389–410.
- Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review*, 106(3), 529–550.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- 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.
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1281–1289.
- Barsalou, L. W. & Wiemer-Hastings, K. (2005). *Situating abstract concepts. Grounding cognition: The role of perception and action in memory, language, and thought*, 129–163.
- Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. *Symbols, Embodiment, and Meaning*, 245–283.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Blythe, H. I., Liang, F., Zang, C., Wang, J., Yan, G., Bai, X., & Liversedge, S. P. (2012). Inserting spaces into Chinese text helps readers to learn new words: An eye movement study. *Journal of Memory and Language*, 67(2), 241–254.
- Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, 45(2), 122–159.
- Bordag, D., Opitz, A., Rogahn, M., & Tschirner, E. (2018). The strength of meaning. *The Mental Lexicon*, 13(1), 74–104.
- Borghini, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, 143(3), 263.
- Bower, G. H., Monteiro, K. P., & Gilligan, S. G. (1978). Emotional mood as a context for learning and recall. *Journal of Verbal Learning and Verbal Behavior*, 17(5), 573–585.
- Brunsnighan, S. M., & Folk, J. R. (2012). Combining contextual and morphemic cues is beneficial during incidental vocabulary acquisition: Semantic transparency in novel compound word processing. *Reading Research Quarterly*, 47(2), 172–190.
- Brysbart, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, 7, 1116.
- Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 225.
- Crystal, D. (2012). *English as a global language* (2nd ed.). Cambridge University Press <https://doi.org/10.1017/CBO9781139196970>.
- Dagenbach, D., Horst, S., & Carr, T. H. (1990). Adding new information to semantic memory: How much learning is enough to produce automatic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 581–591.
- Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25(1), 1–18.
- Elgort, I., Brysbart, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, 40(2), 341–366.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188.
- Ellis, N. C., Frey, E., & Jalkanen, I. (2009). The psycholinguistic reality of collocation and semantic prosody. *Exploring the Lexis-Grammar Interface*, 89–114.
- Erdelyi, M. H. (1974). A new look at the new look: Perceptual defense and vigilance. *Psychological Review*, 81(1), 1.
- Eskenazi, M. A., Swischuk, N. K., Folk, J. R., & Abraham, A. N. (2018). Uninformative contexts support word learning for high-skill spellers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(12), 2019–2025.
- Estes, Z., & Adelman, J. S. (2008a). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion*, 8(4), 441–444.
- Estes, Z., & Adelman, J. S. (2008b). Automatic vigilance for negative words is categorical and general. *Emotion*, 8(4), 453–457.
- Eurobarometer, S. (2006). *Europeans and their languages*. European Commission.
- Fang, X., Perfetti, C., & Stafura, J. (2017). Learning new meanings for known words: Biphase effects of prior knowledge. *Language, Cognition and Neuroscience*, 32(5), 637–649.
- Ferré, P., Ventura, D., Comesaña, M., & Fraga, I. (2015). The role of emotionality in the acquisition of new concrete and abstract words. *Frontiers in Psychology*, 6, 976.
- Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, 61(6), 825–850.
- Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(2), 143–149.
- Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General*, 130(4), 681.
- Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, 24(7), 1243–1252.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1), 3–55.
- Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, 21(3), 563–584.
- Godfroid, A. & Schmidtke, J. (2013). What do eye movements tell us about awareness? a triangulation of eye-movement data, verbal reports and vocabulary learning scores. *Noticing and Second Language Acquisition: Studies in honor of Richard Schmidt*, 183–205.
- Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, 35(3), 483–517.
- Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the noticing hypothesis in the context of vocabulary acquisition. *Cognitive Processing in Second Language Acquisition*, 169–197.
- Godfroid, A., Winke, P., & Rebuschat, P. (2015). Investigating implicit and explicit processing using L2 learners eye-movement data. *Implicit and Explicit Learning of Languages*, 325–348.
- Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL Quarterly*, 21(3), 505–521.
- Green, P., & MacLeod, C. J. (2016). simr: an R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504> Retrieved from <https://CRAN.R-project.org/package=simr>.
- Guo, X., Zheng, L., Zhu, L., Yang, Z., Chen, C., Zhang, L., & Dienes, Z. (2011). Acquisition of conscious and unconscious knowledge of semantic prosody. *Consciousness and Cognition*, 20(2), 417–425.
- Hald, L. A., van den Hurk, M., & Bekkering, H. (2015). Learning verbs more effectively through meaning congruent action animations. *Learning and Instruction*, 39, 107–122.
- Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic Press.
- Herman, P. A., Anderson, R. C., Pearson, P. D., & Nagy, W. E. (1987). Incidental acquisition of word meaning from expositions with varied text features. *Reading Research Quarterly*, 263–284.
- Hothorn, T., Bretz, F., Westfall, P., Heiberger, R., Schuetzenmeister, A., Scheibe, S., & Hothorn, M. (2015). *Package 'multcomp' version 1.4-0*.
- Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, 12(2), 249–268.
- Inkster, M., Wellsby, M., Lloyd, E., & Pexman, P. M. (2016). Development of embodied word meanings: Sensorimotor effects in children's lexical processing. *Frontiers in*

- Psychology*, 7, 317.
- Jenkins, J. R., & Dixon, R. (1983). Vocabulary learning. *Contemporary Educational Psychology*, 8(3), 237–260.
- Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, 23(4), 1214–1220.
- Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, 166, 190–211.
- Joseph, H. S., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, 133(1), 238–248.
- Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, 31(8), 1169–1180.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627–633.
- Kiverstein, J., & Miller, M. (2015). The embodied brain: Towards a radical embodied cognitive neuroscience. *Frontiers in Human Neuroscience*, 9, 237.
- Konopak, B., Sheard, C., Longman, D., Lyman, B., Slaton, E., Atkinson, R., & Thames, D. (1987). Incidental versus intentional word learning from context. *Reading Psychology: An International Quarterly*, 8(1), 7–21.
- Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General*, 140(1), 14.
- Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112(3), 473–481.
- Kuperman, V. (2015). Virtual experiments in megastudies: A case study of language and emotion. *The Quarterly Journal of Experimental Psychology*, 68(8), 1693–1710.
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, 143(3), 1065.
- Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (2008). Not all negative words slow down lexical decision and naming speed: Importance of word arousal. *Emotion*, 8(4), 445–452.
- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202–226.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399–436.
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. *Text and Technology: Honour of John Sinclair*, 157, 176.
- Lowell, R., & Morris, R. K. (2017). Impact of contextual constraint on vocabulary acquisition in reading. *Journal of Cognitive Psychology*, 29(5), 551–569.
- Lund, T. C., Sidhu, D. M., & Pexman, P. M. (2019). Sensitivity to emotion information in children's lexical processing. *Cognition*, 190, 61–71.
- Maratos, E. J., Allan, K., & Rugg, M. D. (2000). Recognition memory for emotionally negative and neutral words: An ERP study. *Neuropsychologia*, 38(11), 1452–1465.
- McGe, I. (2012). Should we teach semantic prosody awareness? *RELC Journal*, 43(2), 169–186.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81.
- Mokhtar, A. A., Rawian, R. M., Yahaya, M. F., Abdullah, A., & Mohamed, A. R. (2017). Vocabulary learning strategies of adult ESL learners. *The English Teacher*, 38, 133–145.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237–270.
- Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. *Frontiers in Human Neuroscience*, 6, 275.
- Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, 108(3), 483.
- Pagán, A., & Nation, K. (2019). Learning words via reading: Contextual diversity, spacing, and retrieval effects in adults. *Cognitive Science*, 43(1).
- Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, 38(1), 97–130.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy*, 11, 67–86.
- Pexman, P. M. (2019). The role of embodiment in conceptual development. *Language, Cognition and Neuroscience*, 34(10), 1274–1283.
- Ponari, M., Norbury, C., & Vigliocco, G. (2017). How do children process abstract concepts? Evidence from a lexical decision task. *Developmental Science*, 10, 10–11.
- Ponari, M., Norbury, C. F., & Vigliocco, G. (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, 21(2), e12549.
- Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, 61(3), 380.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, 35(2), 213–235.
- Saragi, T. (1978). Vocabulary learning and reading. *System*, 6(2), 72–78.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.
- Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. *Interlanguage Pragmatics*, 21, 42.
- Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.
- Schwanenflugel, P. J., Stahl, S. A., & McFalls, E. L. (1997). Partial word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research*, 29(4), 531–553.
- Sheikh, N. A., & Titone, D. A. (2013). Sensorimotor and linguistic information attenuate emotional word processing benefits: An eye-movement study. *Emotion*, 13(6), 1107.
- Siakaluk, P. D., Knol, N., & Pexman, P. M. (2014). Effects of emotional experience for abstract words in the stroop task. *Cognitive Science*, 38(8), 1698–1717.
- Siakaluk, P. D., Newcombe, P. I., Duffels, B., Li, E., Sidhu, D. M., Yap, M. J., & Pexman, P. M. (2016). Effects of emotional experience in lexical decision. *Frontiers in Psychology*, 7, 1157.
- Sinclair, J. M. (1996). The empty lexicon. *International Journal of Corpus Linguistics*, 1(1), 99–119.
- Sinclair, J. (2004). The search for units of meaning. *Trust the text* (pp. 34–58). Routledge.
- Singer, J. A., & Salovey, P. (1988). Mood and memory: Evaluating the network theory of affect. *Clinical Psychology Review*, 8(2), 211–251.
- Sneffella, B., & Kuperman, V. (2016). It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition*, 156, 135–146.
- Speer, R., Chin, J., & Havasi, C. (2017). *ConceptNet 5.5: An open multilingual graph of general knowledge*. (pp. 4444–4451). Retrieved from <http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14972>.
- Swanborn, M. S., & De Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69(3), 261–285.
- Teng, F. (2016). The effects of word exposure frequency on incidental learning of the depth of vocabulary knowledge. *GEMA Online Journal of Language Studies*, 16(3).
- van den Broek, G. S., Takashima, A., Segers, E., & Verhoeven, L. (2018). Contextual richness and word learning: Context enhances comprehension but retrieval enhances retention. *Language Learning*, 68(2), 546–585.
- Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, 1(2), 219–247.
- Vigliocco, G., Ponari, M., & Norbury, C. (2018). Learning and processing abstract words and concepts: Insights from typical and atypical development. *Topics in Cognitive Science*, 10(3), 533–549.
- Vinke, A. A., & Jochems, W. (1993). English proficiency and academic success in international postgraduate education. *Higher Education*, 26(3), 275–285.
- Wagovich, S. A., & Newhoff, M. (2004). The single exposure: Partial word knowledge growth through reading. *American Journal of Speech-Language Pathology*, 13(4).
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207.
- Webb, S. (2007). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, 11(1), 63–81.
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20(2), 232–245.
- Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach-and avoidance-related social information. *Journal of Personality and Social Psychology*, 78(6), 1024.
- Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional stroop task and psychopathology. *Psychological Bulletin*, 120(1), 3.
- Williams, R., & Morris, R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, 16(1–2), 312–339.
- Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia*, 49(5), 1105–1127.
- Winter, B. (2016). Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience*, 31(8), 975–988.
- Winter, B. (2019). *Sensory linguistics: Language, perception and metaphor*, vol. 20. John Benjamins Publishing Company.
- Wochna, K. L., & Juhasz, B. J. (2013). Context length and reading novel words: An eye-movement investigation. *British Journal of Psychology*, 104(3), 347–363.
- Zhang, W. (2009). Semantic prosody and ESL/EFL vocabulary pedagogy. *TESL Canada Journal*, 26(2), 1–12.