Exploring Integrated Co-occurrence and Semantic Mechanisms for Long Term Memory Retrieval

Lily Gebhart (gebhart@oxy.edu)
Justin Li (justinnhli@oxy.edu)
Occidental College, 1600 Campus Road
Los Angeles, CA 90041, USA

Abstract

Semantic and co-occurrence memory associations aid the retrieval of relevant memory elements from long term memory, but little is understood about how semantics and co-occurrence interact in this process. This paper explores the relationship between these associations via computational memory modeling in a Bayesian framework. We assessed the performance of eleven candidate mechanisms on two linguistic tasks - the Word Sense Disambiguation task and the Remote Associates Test. The most successful mechanisms use co-occurrence associations to modulate semantic associations by removing from or adding to the context or pool of candidates for retrieval, consistent with recent experimental work in memory retrieval. Although these results are a promising first step for understanding the relationship between semantic and co-occurrence associations in memory retrieval, more empirical human data is needed to validate the proposed interactions between these associations.

Keywords: Long-term Memory Retrieval; Semantics; Co-occurrence; Bayesian Memory

1 Introduction

Memory retrieval is a complex process, dependent on factors including past retrieval history, working memory context, and associations between elements in long-term memory. These associations could come from many sources; two common sources are that of co-occurrence and semantic associations, which have both been shown to affect memory retrieval. The effects of semantic and co-occurrence associations have often been studied separately, both experimentally and with computational modeling (Schatz, Jones, & Laird, 2022; Greenberg & Verfaellie, 2010). Recent experimental work in psychology and neuroscience, however, suggests that the relationship between co-occurrence and semantics is also important to their role in memory retrieval, though the nature of this relationship has yet to be fully explored (Manning, Sperling, Sharan, Rosenberg, & Kahana, 2012; Ferreira, Charest, & Wimber, 2019; Greenberg & Verfaellie, 2010). For cognitive modeling in particular, previous computational memory models have not consistently defined co-occurrence and semantic retrieval mechanisms, or considered their relationship to be important to their functionality, despite recent experimental progress (Chater et al., 2020; Roelke et al., 2018; Hofmann, Kleemann, Roelke-Wellmann, Vorstius, & Radach, 2022).

In this paper, we investigate the nature of the relationship between semantics and co-occurrence using Bayesian memory models. Using computational models, rather than experimental approaches, has the potential to clarify the relationship between semantics and co-occurrence for human long term memory retrieval and representations of this process in cognitive models. Computational models enable the exploration of possible memory retrieval mechanisms in a simplified framework that can be compared to experimental data later on if necessary. As such, with this work, we explore the space of potential relationships between semantics and co-occurrence for long term memory retrieval with the aim of illuminating psychological features of their relationship and inspiring its consideration in future computational modeling implementations and psychological experiments. [1]

2 Bayesian Memory Modeling

Co-occurrence and semantics refer to associations between memory elements that aid and influence long term memory retrieval. As an example, consider the words "animal", "rat", and "pack". The words "rat" and "animal" have a semantic association since rats are animals; on the other hand, "rat" and "pack" have a co-occurrence association, since they are more likely to occur together in the compound word, "packrat". In this paper, we are interested in how the presence of both "animal" and "pack" together might influence the retrieval of "rat".

We explore this question within a probabilistic Bayesian approach, commonly used in psychological models to determine which memory elements to retrieve into working memory given some immediate context (Tulving & Craik, 2005). Evidence suggests that human memory is only approximately Bayesian, since it is subject to error in encoding, storage, and recall, and is limited to only the biased group of experiences we've encountered (Chater et al., 2020). Nonetheless, in the idealized case, the memory retrieved m is one from the set of viable retrieval candidates M with the highest posterior probability given some working memory context C; that is,

$$\underset{m \in M}{\operatorname{argmax}}\ P(m|C) = \underset{m \in M}{\operatorname{argmax}}\ \frac{P(C|m)P(m)}{P(C)}$$

$P(m)$ and $P(C)$ do not depend on the relationship between co-occurrence and semantics, and thus we are primarily interested in the likelihood $P(C|m)$, which is defined differ-

[1] The code for this paper is available at https://github.com/Lily-Gebhart/Exploring-Integrated-Mechanisms

ently for retrieval mechanisms such as semantics and co-occurrence.

For semantic retrieval, $P(C|m)$ is calculated based on spreading activation and base-level activation, which has been shown to reflect a Bayesian estimate of need odds (Anderson & Milson, 1989; Anderson & Schooler, 1991). Our SEMANTIC retrieval agent implements this mechanism over task-specific semantic networks. When a memory element is retrieved, it is *activated*, with additional diminished activation spreading throughout the connections of a semantic network, beginning with neighbors of the retrieved memory element (Thomson & Lebiere, 2013; Taatgen, Lebiere, & Anderson, 2006). Spreading activation is defined as

$$P(C|m) \propto \text{act}_m = \sum_{k=1}^{n} \frac{1}{2^{\text{dist}(m,k)}} t_k^{-d}$$

where the activation for the memory element of interest m is the sum of the spreading activation it has received previously from each of the n memory elements k, with $\text{dist}(m,k)$ the distance between memory elements m and k in the semantic network, t_k the time since the spreading activation from k occurred, and d a time decay term.

For co-occurrence, $P(C|m)$ is extracted from the statistical co-occurrence of words or concepts in natural language, approximated through sources such as the Google Books ngrams database (Michel et al., 2011) and SemCor corpus (Miller, Leacock, Tengi, & Bunker, 1993). We make the simplifying assumption that the co-occurrence probabilities between each candidate memory element m_i and each context memory element c_j in the current working memory context C are independent, meaning that the likelihood is the product of all $P(m_i|c_j)$ probabilities. This mechanism is the basis of our CO-OCCURRENCE agent that uses task-specific statistics.

3 Integrated Mechanisms

While the likelihood term is well defined for semantics and co-occurrence associations separately, we are interested in the question of how it might be defined when both associations are involved. In Table 1 below, we define eleven candidate mechanisms that explore the relationships between the semantic and co-occurrence associations. These mechanisms can be organized into four categories based on the nature of the psychological relationships they capture, with the goal of exploring the space of potential psychological relationships between these associations.

Two of the mechanisms assume that semantics and co-occurrence are *Independent* in the memory retrieval process, and combine their respective probabilities accordingly. Since semantics and co-occurrence have been primarily studied in isolation from each other in the literature (Greenberg & Verfaellie, 2010), our proposed mechanisms make the simplifying assumption that they are derived from different sources with separate roles in memory retrieval. Two other mechanisms also consider semantic and co-occurrence information as probabilities, but determine retrieval candidates based on features of the semantic and co-occurrence probability distributions, namely their variance and maxima. Intuitively, higher confidence retrievals will result when the retrieved candidates have significantly higher retrieval probabilities than the other viable candidates; these *Distribution-Based* mechanisms will then select the result from the mechanism that has more distinct, or more confident, retrieved candidates.

The remaining categories of mechanisms consider how semantic associations might modify co-occurrence retrieval mechanisms or vice versa. Each of these modification affects different parts of the Bayesian likelihood probability $P(C|m)$ introduced above. The *Probability Modification* category mechanisms use co-occurrence or semantics to directly modify the retrieval probabilities of semantics or co-occurrence, respectively. These mechanisms suggest that co-occurrence and semantic relationships are highly intertwined, and that the calculation of likelihood depends on both associations at a deep level. The *Context Modification* mechanisms add additional memory elements to the context for semantics or co-occurrence, and the *Candidate Modification* mechanisms add or remove memory elements from the pool of viable retrieval candidates. This affects the context C and the retrieval candidates M respectively, since altering the number of memory elements in the context or the number of elements considered as viable retrieval candidates will affect the result of the overall memory retrieval. The intuition behind these mechanisms is to use one association to either narrow the focus of, or to fill in missing information from, the other association.

4 Task Descriptions

We evaluate these candidate co-occurrence and semantic relationships with two computational tasks: Word Sense Disambiguation (WSD) and the Remote Associates Test (RAT). Both tasks have been used for assessing long term memory retrieval, especially in computational modeling applications (Kwong, 2012; Dutta & Basu, 2012; Schatz et al., 2022; Marko, Michalko, & Riečanskỳ, 2019). We selected these tasks for exploring and evaluating the different integrated mechanisms because it has been demonstrated that co-occurrence retrieval mechanisms outperform semantic mechanisms on the WSD task (Montoyo, Suárez, Rigau, & Palomar, 2005; Krovetz & Croft, 1992) and semantic mechanisms outperform co-occurrence mechanisms on the RAT (Schatz et al., 2022). This tradeoff enables us to explore how different candidate integrations of semantics and co-occurrence perform on co-occurrence and semantics-based retrieval tasks. Here, we review the rationale for this tradeoff and describe how we implemented the standard CO-OCCURRENCE and SEMANTICS agents for each task.

4.1 Word Sense Disambiguation (WSD) Task

The WSD task asks an agent to identify the sense of a word given other context words in the sentence. For example, "pack" in the sentence "The pack of rats were swimming" could mean either a group of animals or a small container;

Category	Mechanism	Description
Independent	Joint Probability (JPR)	The probability distributions of SEMANTICS and CO-OCCURRENCE are assumed to be mathematically independent, and the product of probabilities corresponding probability elements is taken to produce a joint probability distribution. This suggests that co-occurrence and semantics are derived from independent sources and contribute to the retrieval process without additional interactions.
	Additive Probability (APR)	The probability distributions of the SEMANTICS and CO-OCCURRENCE agents for each trial are added together. This is proportional to the average of the two distributions. Similar to the previous mechanism, this suggests that the distributions have minimal interaction before they are involved in memory retrieval and give separate estimates of the probability that is averaged over.
Distribution Based	Maximum Probability (MPR)	One of the results from the standard SEMANTICS or CO-OCCURRENCE mechanisms is used, depending on which has a higher probability. Here, co-occurrence and semantics are not directly combined, but one mechanism or the other is selected on a per-retrieval basis, based on which has less uncertainty, as estimated by the maximal probability memory element(s).
	Variance-Based Selector (VBS)	One of the results from the standard SEMANTICS or CO-OCCURRENCE mechanisms is used, depending on which distribution has higher "certainty". We explored two definitions of certainty: as the standard deviation of probabilities of the candidates, and as the difference between the highest and second highest maximal probability memory element(s) in each distribution.
Probability Modification	Semantics Boosted Co-occurrence (SBC)	If two words are semantically related, their co-occurrence conditional probabilities will receive a small boost according to some function $f : [0,1] \rightarrow [0,1]$. We explored two such functions, the sigmoid and square root, for the semantic "boost". This implies that semantic associations in memory retrieval further boost existing co-occurrence associations.
	Co-occurrence Weighted Semantics (CWS)	The edges between memory elements in the semantic network are weighted by conditional co-occurrence probabilities. This implies that co-occurrence associations in memory retrieval is modulate the degree to which memory elements are semantically related in the network.
Context Modification	Semantic Supplemented Co-occurrence (SSC)	Memory elements that are directly semantically related to each of the original context memory elements are also included as context; with this expanded context, the standard co-occurrence retrieval mechanism is used. This suggests that semantic associations are used to supplement the current working memory context for co-occurrence memory retrieval.
	Co-occurrence Supplemented Semantics (CSS)	Memory elements that are co-occurrence associations to the original semantic context are also included as context; this expanded context is activated as part of the standard semantic retrieval mechanism. Similar to SSC, this suggests that co-occurrence associations are used to supplement the current working memory context for semantic memory retrieval.
Candidate Modification	Co-occurrence Expanded Semantics (CES)	All co-occurrence associations are incorporated into the semantic network as if they were semantic associations themselves. This allows spreading to occur over co-occurrence associations as well, with the implication that co-occurrence associations help define the semantic network structure.
	Co-occurrence Filtered Semantics (CFS)	Associations between memory elements in the semantic network that do not also have a co-occurrence association are removed, before the standard semantic mechanism is used. This suggests that stored co-occurrence associations filter the semantic network so that only the most relevant semantic associations remain.
	Semantic Filtered Co-occurrence (SFC)	Co-occurrence associations that are not also semantically associated are removed, before the standard co-occurrence mechanism is used. This implies that co-occurrence associations between two memory elements are only maintained if the semantic activation probability between the elements is above a predefined threshold.

Table 1: Integrated Mechanisms and Agents

deciphering its meaning is crucial to understanding what the speaker or writer means. Past literature has shown that the WSD task relies heavily on statistical co-occurrence information (Montoyo et al., 2005), making it a meaningful metric for how candidate integrated mechanisms are able to utilize co-occurrence information in memory retrieval.

Our implementation of the WSD task uses text from the SemCor corpus. A subset of WordNet that contain the semantic relations of words in the SemCor corpus were used to generate the semantic network (Bird, Klein, & Loper, 2009). We further divided the corpus into six partitions of 5,000 sentences to limit the spreading of activation and make our models more computationally tractable. We measure task performance as the percentage of correct trials averaged over all partitions. If the agent suggests that n senses are equally likely and one of them is correct, the agent's performance is discounted proportionally as $\frac{1}{n}$.

In each trial, the SEMANTICS agent activates all other words in the sentence at the same time. Spreading activation then occurs, and the most activated element is used as the answer. The decay parameter and the spreading activation depth were kept constant. The SEMANTICS WSD agent is additionally parameterized by whether how far activation spreads, and by whether the network activations are cleared after every trial, cleared after all trials in a sentence, or never cleared; otherwise, the repeated activation of the words in each sentence accumulates.

Meanwhile, the CO-OCCURRENCE agent answers using word with the maximum posterior probability, with a naive Bayes assumption for all context words. The probabilities are learned from the SemCor corpus itself, but we parameterize the agent by either considering the correct senses of the context words, or only the context words themselves. For example, in the sentence, "The pack of rats were swimming", the conditional probability for the retrieved sense of "pack" could be based on co-occurrence with the correct sense of "rats", or merely the co-occurrence with the word "rats" regardless of its sense.

4.2 Remote Associates Test (RAT)

In RAT, an agent is provided with three context words and asked to find a word that relates to all three. For example, if given the words "animal", "back", and "rat", the correct response would be "pack" to form the compound words "pack animal", "backpack", and "pack rat" (Bowden & Jung-Beeman, 2003). RAT is often used as a test of creativity and mental acuity in experimental psychology, with relatively poor human performance due to the limited context given in the task (Wu, Huang, Chen, & Chen, 2020). As such, it is also a good test of whether candidate integrated mechanisms are able to utilize relational, semantic associations, although many of the solutions to RAT problems also rely on commonly co-occurring words.

Our implementation of the RAT uses the 142 question bank from Bowden and Jung-Beeman (2003), which is frequently used in both experimental and computational studies. Given the three context words for each trial, each agent only guesses once, selecting the candidate target with the maximum probability or activation as its guess (Schatz et al., 2022). RAT is a difficult task in general, with performance here further constrained by the fact that the co-occurrence and semantic data sources do not contain solutions to every RAT problem. Therefore, it is reasonable to expect low accuracy on the RAT in general, regardless of the mechanism.

The SEMANTICS agent uses a semantic network generated from the South Florida Free Association Norms (SFFAN) corpus. Since the SFFAN corpus is a large database of human free association norms, it may contain some co-occurrence relations in addition to semantic relations (Nelson, McEvoy, & Schreiber, 2004). While prior work have used other semantic networks (Schatz et al., 2022), SFFAN contains many of the RAT associations in a small network, and therefore is relatively computationally tractable for use in experiments. In each trial, the agent activates all context words, then uses the word with the highest resulting activation (that is not also a context word) as the answer. All activations in the semantic network were cleared after each trial.

The CO-OCCURRENCE agent uses co-occurrence conditional probabilities from the Google Books English One Million dataset, which provides a relatively unbiased, representative source of co-occurrence probabilities required for the RAT (Michel et al., 2011). We only use bigrams and not all n-grams for computational tractability and consistency with the smaller network sizes used in our implementation of the WSD task. In each trial, for each candidate answer, its bigram probabilities with all three context words is multiplied together; the word with the highest joint probability is selected as the trial guess. The task accuracy is computed using the same methods employed to compute the accuracy on the WSD.

4.3 Task Baselines

To contextualize the standard SEMANTICS and CO-OCCURRENCE agents, as well as the integrated agents' performance on the WSD task and RAT, we compare their performance to two additional baselines: the ORACLE and uniform RANDOM agents. The ORACLE is correct if either of CO-OCCURRENCE or SEMANTICS is correct, thereby indicating the expected maximum performance considering only the abilities of the SEMANTICS and CO-OCCURRENCE agents. It is possible for an integrated agent to perform better than ORACLE if it is able to find synergy between co-occurrence and semantic information. The ORACLE is therefore not necessarily an upper bound on the performance of integrated agents.

The uniform RANDOM baseline r assesses how well integrated candidates perform relative to chance. For each of n trials, $r = \sum_{i=1}^{n} \frac{1}{s_i}$. For the WSD task, s_i is the number of valid word senses of each trial and for the RAT task, s_i is the number of words that form bigrams with all three trial words; This s_i value for the RAT task provides a better comparison of performance than statically using the size of the SFFAN network because many of the candidates in the network have zero probabilities each trial and the candidates with non-zero probabilities fluctuate between trials.

5 Experiment Results

Figure 1 provides a comparison of the relative performance of each agent on the WSD and RAT tasks. For simplicity, for each agent, we only show the results of the most successful combination(s) of agent and task parameters. As expected, there is a tradeoff in the performance of the standard CO-OCCURRENCE and SEMANTICS agents, with CO-OCCURRENCE outperforming the SEMANTICS on WSD and vice versa on RAT, confirming our initial assumptions. These results and the results of the integrated agents to follow are further contextualized by the performance of the ORACLE (WSD: 100%, RAT: 37.3%) and RANDOM baselines (WSD: 37.9%, RAT: 3.1%), representing the maximum anticipated and chance-level performance on each task.

In general, the results of the integrated agents on either task are not consistent with the original psychologically-based framework for organizing the integrated mechanisms. For each mechanism the retrieval probabilities of all viable retrieval candidates for each trial results in a probability distribution; the average variance and ranking of the solution in each trial distribution for a given agent are used to interpret the results of each agent on either task. It is relevant to note that there is a much larger number of candidates with nonzero probabilities for the RAT compared to the WSD, which contributes to the differences in distribution shape on each task as discussed below. Additionally, mechanisms perform better

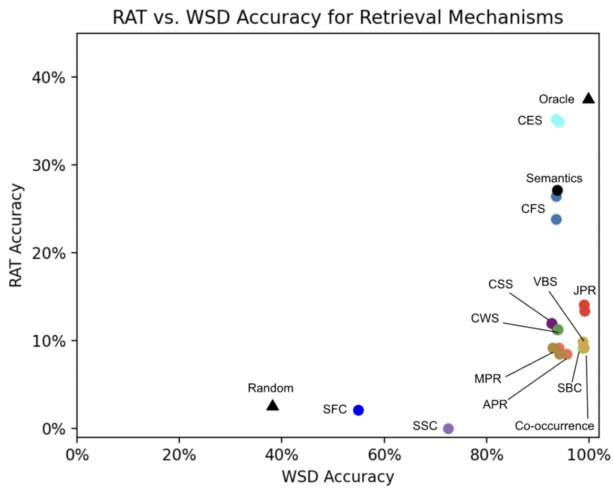

Figure 1: A performance comparison of the accuracies of the Co-occurrence, Semantic Spreading, and integrated mechanism candidates on WSD and RAT.

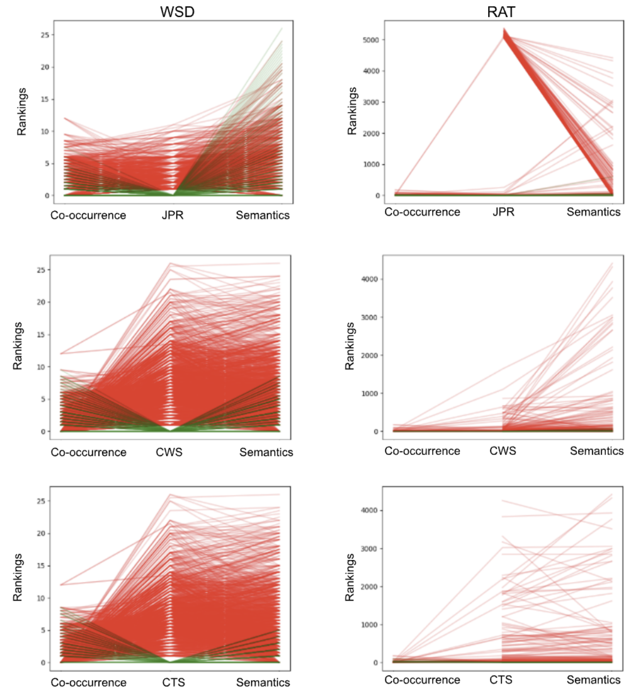

Figure 2: The rankings of the solution to each task trial in CO-OCCURRENCE, SEMANTICS, and integrated agent distributions is compared for three integrated agents: CWS, JPR, and CTS. Lines connect the ranking of the solution in each of the CO-OCCURRENCE, integrated, and SEMANTICS distributions for a given trial with green lines indicating that the integrated agent correctly retrieved the solution element for the task and red lines that it did not.

on WSD when the variance of the distribution of their candidate retrieval probabilities is higher. Higher variance implies that distributions are more "peaked", or that the highest probability retrieval candidates are significantly more likely to be the retrieved than the other candidate elements. The same trend does not hold for RAT.

We will now discuss how the results of each integrated agent contributes to our understanding of the relationship between semantics and co-occurrence in facilitating long term memory retrieval.

The SFC and SSC agents perform poorly on both the WSD and RAT, indicating that integrating semantic information into the context or candidate pool of a co-occurrence based retrieval is ineffective for retrievals that largely rely on co-occurrence (WSD) or semantic (RAT) information. The poor performance of SFC can be attributed to the fact that influential co-occurrence associations that are not also semantic associations fail to be considered as retrieval candidates. Meanwhile, the SSC agent performs poorly on the WSD and RAT likely as a result of reduced variance between candidates, or equivalently, increased ambiguity of which candidate is correct; increasing the number of memory elements in the context results in a more diffuse retrieval probability distribution, lessening the likelihood that the ideal candidates will be retrieved.

The CSS, CWS, MPR, and APR agents comprise a cluster of mechanisms which perform about the same as CO-OCCURRENCE on RAT and about the same as SEMANTICS on WSD. The success of these agents can be attributed to factors specific to each mechanism including high retrieval ambiguity for APR, the ineffectiveness of selecting the maximum probability element as a proxy for "certainty", and increased selectivity by weighting and removing semantic network edges by CWS and CSS. Interestingly, though CWS performed more poorly than SEMANTICS on RAT, it does increase the ranking of the correct retrieval candidate compared to the CO-OCCURRENCE and SEMANTIC agents (shown in Figure 2. It is possible that more complex integrations of co-occurrence retrieval probabilities into determining edge weights are necessary for more effective retrieval using this strategy, but such weights were not explored here.

Another cluster of candidates including JPR, SBC, and VBS performs nearly identically to CO-OCCURRENCE on WSD and CO-OCCURRENCE on RAT, with JPR performing slightly better than the other candidates on RAT. As opposed to the other independent mechanism, APR, which suffers from high retrieval ambiguity, JPR seems promising as a candidate retrieval mechanism. As demonstrated prominently in the RAT ranking comparison plot in Figure 2, JPR generally ranks the correct retrieval candidate higher in the distribution, as long as the solution is both co-occurrently and semantically related to the context, otherwise it is ranked low in the distribution. Testing the JPR on different sources of co-occurrence and semantic associations may be useful in gathering more conclusive results on the efficacy of JPR, especially for semantically-based retrievals.

Finally, the CFS and CES agents perform best out of all integrated candidates, with performance similar to SEMANTICS on WSD and similar or better than SEMANTICS on RAT. The performance of CFS suggests that the removal of non co-occurrence associations from the network maintains edges that link context words to the ideal retrieval candidates while removing less relevant connections from the network. However, the random effects the removal of these candidates has on the ranking of the solution in each trial suggests that critical edges are also being removed from the network, as shown in Figure 2. Removing edges from the graph based on the strength rather than the lack of existence of co-occurrence associations as in the current CFS may prove more effective in both co-occurrence and semantic based retrievals. Meanwhile, the exceptional performance of CES, especially on RAT, may be partially attributed to the inclusion of co-occurrence relations in the free-association corpora used to generate the semantic network. Furthermore, by nature of the corpora used, each word has more co-occurrence associations than semantic associations, so the effect may have resulted from heightened spreading activation in the larger more interconnected network that resulted. Regardless, these results suggest that co-occurrence associations are helpful in modifying the structure of the semantic network for both co-occurrence and semantic based retrieval tasks. Further investigation is necessary to determine if adding and deleting edges from the semantic network based on co-occurrence associations would be more effective than CFS and CES.

6 Discussion

The computational agents implemented in this paper, and the underlying retrieval mechanisms they represent, explored several possible relationships between semantic and co-occurrence mechanisms in long term memory retrieval. Based on our results, the most successful agent across the two tasks uses co-occurrence associations to modulate the semantic retrieval mechanism by adding relevant co-occurrence associations to the retrieval context, which broadens the search over the semantic network. Other successful agents on both tasks modify the semantic network by adding or removing edges based on co-occurrence associations. In both cases, the core underlying retrieval mechanism relies on semantic networks, but co-occurrence associations provide additional information that modify the spreading activation process. While each of these co-occurrence influences on semantic retrieval were explored in independent mechanisms, it is likely that some or all of these mechanisms are at play, with co-occurrence associations modulating semantics at multiple stages in the retrieval process.

In a way, these results fit in with findings from psychology literature. Children learn co-occurrence associations at a young age by first associating concepts that directly co-occur in their natural environment, and only as their linguistic and conceptual abilities mature do children learn semantic associations as memory elements with similar sets of co-occurrence associations (Unger & Fisher, 2021; Savic, Unger, & Sloutsky, 2023). Recent developments in word embedding models such as word2vec further demonstrate how semantic relationships could be derived from natural usage statistics (Rohde, Gonnerman, & Plaut, 2006). In both children and such AI models, semantic associations serve as symbolic gists of the underlying co-occurrence statistics, but may not capture the nuances of the full joint probabilities. While many tasks may be solvable using semantic associations, it is possible that additionally incorporating lower-level co-occurrence information - using it to introduce additional context or to (de)emphasize particular semantic associations - could lead to better retrievals. If we consider episodic memory to be a source of co-occurrence information, this may also corroborate accounts that episodic memory facilitates retrieval from semantic memory (Greenberg & Verfaellie, 2010).

Even though our computational results are suggestive of trends in experimental work, it is important to note that this remains only an early step in understanding the relationship between co-occurrence and semantic associations in memory retrieval. While the mechanisms suggested several ways that semantics and co-occurrence may interact, there are many other possible mechanisms for their interaction. We did not explore mechanisms that integrate the associations in more complex ways, nor how multiple mechanisms could be incorporated into a larger framework. Furthermore, while we considered several variable parameters for the tasks and agents considered, there are a large number of parameters that we did not consider or left fixed in our models. It is also likely the case that co-occurrence associations are only independent in certain cases, not in general as we assumed here.

One difficulty is that co-occurrence and semantic effects have been surprisingly difficult to tease apart experimentally; we were only able to find one paper that attempts to measure how these associations affect retrieval separately and together (Roelke et al., 2018). This dearth of human data make modeling difficult, hence our use of simple linguistic tasks, which bring their own problems regarding the datasets and database from which we draw co-occurrence and semantic information. Beyond how these associations may not match those in people, the distinction between semantic and co-occurrence is also blurred in the data, as the semantic networks used in our experiments also contain co-occurrence associations. As mentioned above, such gray areas between different types of associations also exist in people, which further complicates efforts understand the distinct and combined contributions.

Nonetheless, although this work is preliminary, we believe it is a needed first step to combine the literature on co-occurrence and semantic associations. Computational models have not cleanly distinguished between the effects of these mechanisms, and have even used the same representations (such as semantic networks) to model both leading to difficulties in understanding their separate and joint effects on memory retrieval. The psychology literature has been much more careful in considering each association, and to follow suit, we carefully considered how they might be different and how their effects might be combined in a computational model. Much additional work is needed, both experimentally and via modeling, to fully understand the interplay of co-occurrence and semantic associations in memory retrieval.

References

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*(4), 703.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological science*, *2*(6), 396–408.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. O'Reilly.

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior research methods, instruments, & computers*, *35*, 634–639.

Chater, N., Zhu, J.-Q., Spicer, J., Sundh, J., León-Villagrá, P., & Sanborn, A. (2020). Probabilistic biases meet the Bayesian brain. *Current Directions in Psychological Science*, *29*(5), 506–512. doi: 10.1177/0963721420954801

Dutta, S., & Basu, A. (2012). A cognitive approach to word sense disambiguation. In *Proceedings of the 13th international conference on computational linguistics and intelligent text processing* (pp. 211–224).

Ferreira, C. S., Charest, I., & Wimber, M. (2019). Retrieval aids the creation of a generalised memory trace and strengthens episode-unique information. *NeuroImage*, *201*, 115996.

Greenberg, D. L., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: Evidence from neuropsychology. *Journal of the International Neuropsychological society*, *16*(5), 748–753.

Hofmann, M. J., Kleemann, M. A., Roelke-Wellmann, A., Vorstius, C., & Radach, R. (2022). Semantic feature activation takes time: longer soa elicits earlier priming effects during reading. *Cognitive Processing*, *23*(2), 309–318.

Krovetz, R., & Croft, W. B. (1992). Lexical ambiguity and information retrieval. *ACM Transactions on Information Systems (TOIS)*, *10*(2), 115–141.

Kwong, O. Y. (2012). *New perspectives on computational and cognitive strategies for word sense disambiguation*. Springer Science & Business Media.

Manning, J. R., Sperling, M. R., Sharan, A., Rosenberg, E. A., & Kahana, M. J. (2012). Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *Journal of Neuroscience*, *32*(26), 8871–8878.

Marko, M., Michalko, D., & Riečanskỳ, I. (2019). Remote associates test: An empirical proof of concept. *Behavior research methods*, *51*(6), 2700–2711.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, G. B., … others (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Miller, G. A., Leacock, C., Tengi, R., & Bunker, R. T. (1993). A semantic concordance. In *Human language technology: Proceedings of a workshop held at plainsboro, new jersey, march 21-24, 1993.*

Montoyo, A., Suárez, A., Rigau, G., & Palomar, M. (2005). Combining knowledge-and corpus-based word-sense-disambiguation methods. *Journal of Artificial Intelligence Research*, *23*, 299–330.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Roelke, A., Franke, N., Biemann, C., Radach, R., Jacobs, A. M., & Hofmann, M. J. (2018). A novel co-occurrence-based approach to predict pure associative and semantic priming. *Psychonomic Bulletin and Review*, *25*(4), 1488–1493. doi: 10.3758/s13423-018-1453-6

Rohde, D. L., Gonnerman, L. M., & Plaut, D. C. (2006). An improved model of semantic similarity based on lexical co-occurrence. *Communications of the ACM*, *8*(627-633), 116.

Savic, O., Unger, L., & Sloutsky, V. M. (2023). Experience and maturation: The contribution of co-occurrence regularities in language to the development of semantic organization. *Child development*, *94*(1), 142–158.

Schatz, J., Jones, S. J., & Laird, J. E. (2022). Modeling the remote associates test as retrievals from semantic memory. *Cognitive Science*, *46*(6), e13145.

Taatgen, N. A., Lebiere, C., & Anderson, J. R. (2006). Modeling paradigms in ACT-R. *Cognition and multi-agent interaction: From cognitive modeling to social simulation*, 29–52.

Thomson, R., & Lebiere, C. (2013). Constraining Bayesian inference with cognitive architectures: An updated associative learning mechanism in ACT-R. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).

Tulving, E., & Craik, F. I. (2005). *The oxford handbook of memory*. Oxford University Press.

Unger, L., & Fisher, A. V. (2021). The emergence of richly organized semantic knowledge from simple statistics: A synthetic review. *Developmental Review*, *60*, 100949.

Wu, C.-L., Huang, S.-Y., Chen, P.-Z., & Chen, H.-C. (2020). A systematic review of creativity-related studies applying the remote associates test from 2000 to 2019. *Frontiers in psychology*, *11*, 573432.