Table 1: Performance of the proposed method with different semantic relation types.

Method	RG	MC	RW	SCWS	MEN	sem	syn	total	SemEval
corpus only	0.7523	0.6398	0.2708	0.460	0.6933	61.49	66.00	63.95	37.98
Synonyms	**0.7866**	**0.7019**	0.2731	**0.4705**	**0.7090**	61.46	**69.33**	**65.76**	**38.65**
Antonyms	0.7694	0.6417	0.2730	0.4644	0.6973	61.64	66.66	64.38	38.01
Hypernyms	0.7759	0.6713	0.2638	0.4554	0.6987	61.22	68.89	65.41	38.21
Hyponyms	0.7660	0.6324	0.2655	0.4570	0.6972	61.38	68.28	65.15	38.30
Member-holonyms	0.7681	0.6321	0.2743	0.4604	0.6952	**61.69**	66.36	64.24	37.95
Member-meronyms	0.7701	0.6223	0.2739	0.4611	0.6963	61.61	66.31	64.17	37.98
Part-holonyms	0.7852	0.6841	0.2732	0.4650	0.7007	61.44	67.34	64.66	38.07
Part-meronyms	0.7786	0.6691	**0.2761**	0.4679	0.7005	61.66	67.11	64.63	38.29

relation type. Specifically, we minimize (4) for different λ values, and use the learnt word representations to measure the cosine similarity for the word-pairs in the **WS** dataset. We then select the value of λ that gives the highest Spearman correlation with the human ratings on the **WS** dataset. This procedure is repeated separately with each semantic relation type R. We found that λ values greater than 10000 to perform consistently well on all relation types. The level of performance if we had used only the corpus for learning word representations (without using a semantic lexicon) is shown in Table 1 as the **corpus only** baseline. This baseline corresponds to setting $\lambda = 0$ in (4).

From Table 1, we see that by incorporating most of the semantic relations found in the WordNet we can improve over the corpus only baseline. In particular, the improvements reported by synonymy over the **corpus only** baseline is statistically significant on **RG**, **MC**, **SCWS**, **MEN**, **syn**, and **SemEval**. Among the individual semantic relations, synonymy consistently performs well on all benchmarks. Among the other relations, part-holonyms and member-holonyms perform best respectively for predicting semantic similarity between rare words (**RW**), and for predicting semantic analogies (**sem**) in the Google dataset. Meronyms and holonyms are particularly effective for predicting semantic similarity between rare words. This result is important because it shows that a semantic lexicon can assist the representation learning of rare words, among which the co-occurrences are small even in large corpora (**?**), The fact that the proposed method could significantly improve performance on this task empirically justifies our proposal for using a semantic lexicon in the word representation learning process. Table 1 shows that not all relation types are equally useful for learning word representations for a particular task. For example, hypernyms and hyponyms report lower scores compared to the corpus only baseline on predicting semantic similarity for rare (**RW**) and ambiguous (**SCWS**) word-pairs.

In Table 2, we compare the proposed method against previously proposed word representation learning methods that use a semantic lexicon: **RCM** is the relational constrained model proposed by **?** (**?**), **R-NET**, **C-NET**, and **RC-NET** are proposed by **?** (**?**), and respectively use relational information, categorical information, and their union from the WordNet for learning word representations, and **Retro** is the retrofitting method proposed by **?** (**?**). Details of those methods are described in Section 2. For **Retro**, we use the publicly available implementation[4] by the original authors, and use pre-trained word representations on the same ukWaC corpus as used by the proposed method. Specifically, we retrofit word vectors produced by CBOW (**Retro (CBOW)**), and skip-gram (**Retro (SG)**). Moreover, we retrofit the word vectors learnt by the corpus only baseline (**Retro (corpus only)**) to compare the proposed *joint* learning approach to the *post-processing* approach in retrofitting. Unfortunately, for **RCM**, **R-NET**, **C-NET**, and **RC-NET** their implementations, nor trained word vectors were publicly available. Consequently, we report the published results for those methods. In cases where the result on a particular benchmark dataset is not reported in the original publication, we have indicated this by a dash in Table 2.

Table 2: Comparison against prior work.

Method	RG	MEN	sem	syn
RCM	0.471	0.501	-	29.9
R-NET	-	-	32.64	43.46
C-NET	-	-	37.07	40.06
RC-NET	-	-	34.36	44.42
Retro (CBOW)	0.577	0.605	36.65	52.5
Retro (SG)	0.745	0.657	45.29	65.65
Retro (corpus only)	0.786	0.673	61.11	68.14
Proposed (synonyms)	**0.787**	**0.709**	**61.46**	**69.33**

Among the different semantic relation types compared in Table 1, we use the synonym relation which reports the best performances for the proposed method in the comparison in Table 2. All word embeddings compared in Table 2 are 300 dimensional and use the WordNet as the sentiment lexicon. From Table 2, we see that the proposed method reports the best scores on all benchmarks. Except for the smaller (only 65 word-pairs) **RG** dataset where the performance of retrofitting is similar to that of the proposed method, in all other benchmarks the proposed method statistically significantly outperforms prior work that use a semantic lexicon for word representation learning.

We evaluate the effect of the dimensionality d on the word representations learnt by the proposed method. For the limited availability of space, in Figure 1 we report results when we use the synonymy relation in the proposed method and on the semantic similarity benchmarks. Similar trends were observed for the other relation types and benchmarks.

[4]https://github.com/mfaruqui/retrofitting

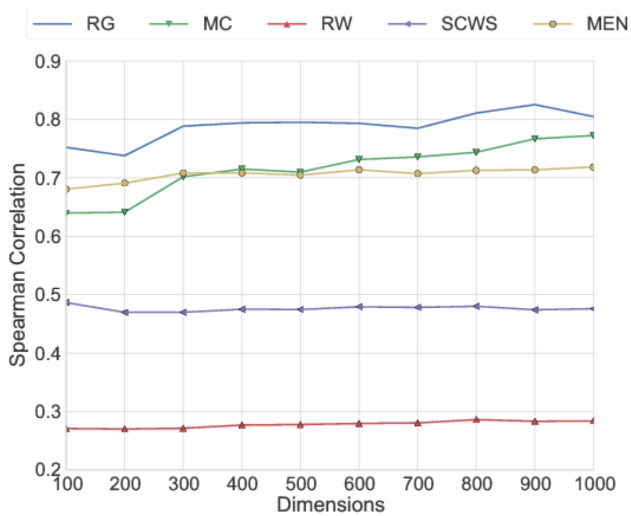

Figure 1: The effect of the dimensionality of the word representations learnt by the proposed method using the synonymy relation, evaluated on semantic similarity prediction task.

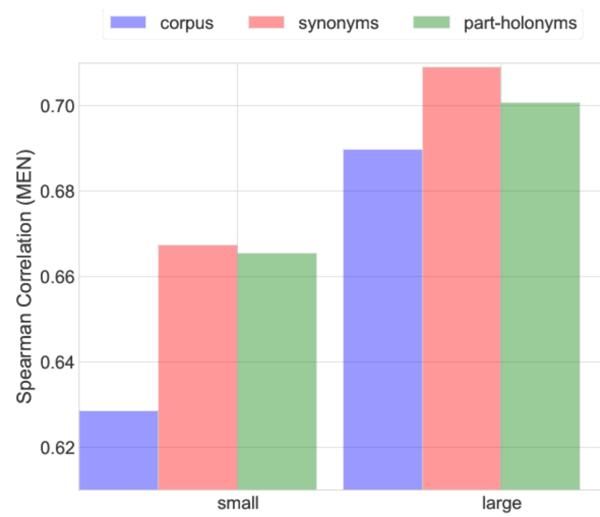

Figure 2: The effect of using a semantic lexicon under different corpus sizes. The performance gain is higher when the corpus size is small.

From Figure 1 we see that the performance of the proposed method is relatively stable across a wide range of dimensionalities. In particular, with as less as 100 dimensions we can obtain a level of performance that outperforms the corpus only baseline. On **RG**, **MC**, and **MEN** datasets we initially see a gradual increase in performance with the dimensionality of the word representations. However, this improvement saturates after 300 dimensions, which indicates that it is sufficient to consider 300 dimensional word representations in most cases. More importantly, adding new dimensions does not result in any decrease in performance.

To evaluate the effect of the corpus size on the performance of the proposed method, we select a random subset containing 10% of the sentences in the ukWaC corpus, which we call the *small* corpus, as opposed to the original *large* corpus. In Figure 2, we compare three settings: **corpus** (corresponds to the baseline method for learning using only the corpus, without the semantic lexicon), **synonyms** (proposed method with synonym relation), and **part-holonyms** (proposed method with part-holonym relation). Figure 2 shows the Spearman correlation coefficient on the **MEN** dataset for the semantic similarity prediction task. We see that in both small and large corpora settings we can improve upon the corpus only baseline by incorporating semantic relations from the WordNet. In particular, the improvement over the corpus only baseline is more prominent for the smaller corpus than the larger one. Similar trends were observed for the other relation types as well. This shows that when the size of the corpus is small, word representation learning methods can indeed benefit from a semantic lexicon.

5 Conclusion

We proposed a method for using the information available in a semantic lexicon to improve the word representations learnt from a corpus. For this purpose, we proposed a global word co-occurrence prediction method using the semantic relations in the lexicon as a regularizer. Experiments using ukWaC as the corpus and WordNet as the semantic lexicon show that we can significantly improve word representations learnt using only the corpus by incorporating the information from the semantic lexicon. Moreover, the proposed method significantly outperforms previously proposed methods for learning word representations using both a corpus and a semantic lexicon in both a semantic similarity prediction task, and a word analogy detection task. The effectiveness of the semantic lexicon is prominent when the corpus size is small. Moreover, the performance of the proposed method is stable over a wide-range of dimensionalities of word representations. In future, we plan to apply the word representations learnt by the proposed method in downstream NLP applications to conduct extrinsic evaluations.

Ex eligendi voluptatum consectetur architecto ut ratione quod quisquam, nisi sequi consectetur.Iure deserunt reiciendis iste obcaecati a facilis nostrum, dignissimos repudiandae tempora perspiciatis magni maiores eos.Ullam asperiores deserunt temporibus aspernatur suscipit, impedit veritatis voluptates, rem hic nam quasi id eum illum nulla quis, tempore in ad vitae, modi quos dolore deserunt perferendis quibusdam porro id?Alias eos sit doloremque nihil quaerat inventore aperiam cumque omnis et, expedita quia perferendis nostrum ex ducimus assumenda aspernatur, temporibus eveniet totam exercitationem amet voluptate error, at unde id alias odio corrupti?Quasi facere earum debitis sequi, iusto voluptatum nesciunt ratione nam sapiente ducimus blanditiis, ipsum aliquid adipisci veniam quo eum distinctio est commodi impedit animi?Id accusantium molestiae non distinctio deleniti provident laborum,