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.
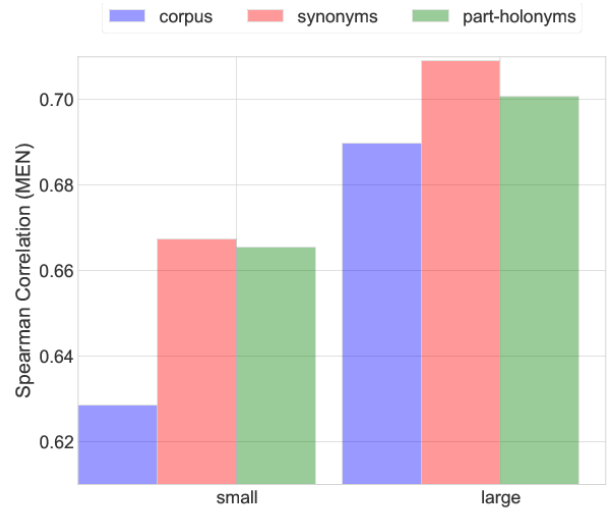
## 5 Conclusion

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,