processing is applied, but as soon as the context vectors are included, the spectrum of similarity scores widens up for SGNS. This investigation may explain why SVD is unable to manifest paradigmatic and syntagmatic relations at the same time.

SVD does not get a huge benefit from more training data or the post-processing step for inclusion of the context vectors. The underlying reason is that SVD always uses a sub-space of the entire similarity spectrum [-0.5, 1.0] so everything is squeezed – we refer to this phenomenon as *space compression*, which we hypothesize is due to the limitations of the dimensionality reduction mechanism. On the other hand, the distribution of words in the vector space obtained from SGNS changes drastically both by training on more data and considering context vectors.

As Figure 3 shows, SGNS has the capacity to use up the entire similarity spectrum [-1.0, 1.0], i.e., *space expansion*. We conjecture that this is due both to the design of the objective function and to the larger number of parameters in the neural model being updated independently, making it a more flexible method to encode fine-grained differences between word groups, while keeping them in meaningful clusters. More data helps the model fine-tune its parameters. Furthermore, averaging the word and context vectors provides an ensemble voting for syntagmatic (relatedness) and paradigmatic (similarity) at the same time.

3.5 Word Clusters in the Semantic Space

The space expansion of the SGNS model by inclusion of the context vectors can be visualized with a 2-dimensional projection of the vectors obtained from *w* vs. *w+c* post-processing conditions, depicted in Figures 4 and 5 respectively. A comparison between the two plots shows how the vicinity of paradigmatically similar words (interchangeable words such as *cat* and *mouse*) can be preserved while syntagmatic clusters are emphasized (*cat* and *chase*) by inclusion of context vectors.

It is important, however, to note that higher-level paradigmatic relations are negatively affected as the model tries to bring syntagmatically related words closer to one another. For example, verbs and nouns (clustered in gray ovals in Figures 4), which are paradigmatically different, get mixed up once the syntagmatic clusters start to shape (gray rectangles in Figure 5). On the other hand, nouns referring to animate categories (that have some level of paradigmatic similarity) fall apart in the *w+c* space (red dashed cluster in Figures 4, distorted in Figure 5). These observations emphasize the importance of the post-processing choices based on the final inferences we expect from the model. When generalized to a natural language setting, the models depending on the w+c parameterization would demonstrate synonymy, similarity and associative relatedness differently.

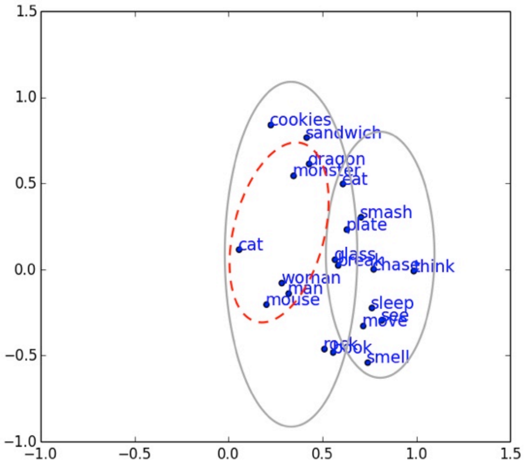

Figure 4. Paradigmatic clusters in SGNS *w* vector space; Syntagmatic clusters not easily identified (10K corpus, dim = 14, neg = 1)

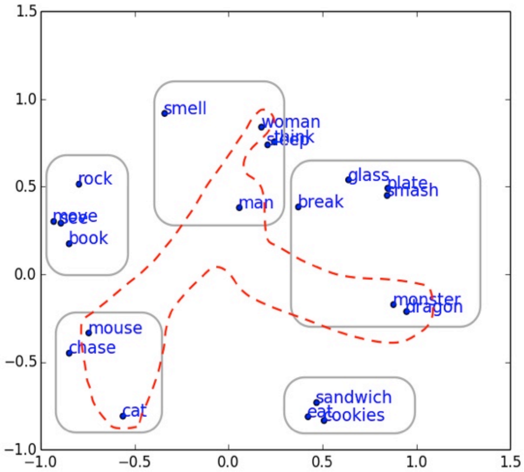

Figure 5. Clear syntagmatic clusters in SGNS *w+c* vector space; some paradigmatically related words are kept together and some have fallen apart (10K corpus, dim = 14, neg = 1)

One should consider that while dimensionality reduction to two dimensions is possible and helpful for visualization purposes, these images do not reflect the exact distances between words in the high-dimension space. Therefore, these observations should be understood in