

## **A Full Results**

The full results from our experiments are given in Table 4. In the main text, we present the average scores for the 4 – 6 STS tasks in each year. Here, we provide the results for all tasks. For a particular year, the tasks preceding it in the same section (i.e., not separated by a horizontal line) are in the same year. For example, the tasks for STS’12 are MSRpar, MSRvid, SMT-eur, OnWN, and SMT-news. Tasks in different years with the same name are not related (e.g., ‘headlines’ in STS’13 and STS’14).

| Supervision     | Wieting et al.(2017b) |             |             | Wieting et al. (2016b) |             |              | Arora et al. (2017) |        | Our Approach |             |             |
|-----------------|-----------------------|-------------|-------------|------------------------|-------------|--------------|---------------------|--------|--------------|-------------|-------------|
|                 | Weakly Supervised     |             |             | Unsupervised           |             |              | Weakly Supervised   |        | Unsupervised |             |             |
| Tasks           | LSTM AVG              | AVG         | GRAN        | PP-XXL                 | tfidf-GloVe | skip-thought | GloVe+WR            | PSL+WR | GloVe+UP     | PSL+UP      | Czeng+UP    |
| MSRpar          | 49.0                  | 45.9        | 47.7        | 44.8                   | 50.3        | 16.8         | 35.6                | 43.3   | 51.4         | 55.2        | <b>58.3</b> |
| MSRvid          | 84.3                  | 85.1        | 85.2        | 79.6                   | 77.9        | 41.7         | 83.8                | 84.1   | 86.4         | 85.4        | <b>90.1</b> |
| SMT-eur         | 51.2                  | 47.5        | 49.3        | 49.5                   | 54.7        | 35.2         | 49.9                | 44.8   | <b>54.9</b>  | 51.3        | 54.7        |
| OnWN            | 71.5                  | 71.2        | 71.5        | 70.4                   | 64.7        | 29.7         | 66.2                | 71.8   | 72.7         | <b>74.3</b> | 74.2        |
| SMT-news        | <b>68.0</b>           | 58.2        | 58.7        | 63.3                   | 45.7        | 30.8         | 45.6                | 53.6   | 59.3         | 62.8        | 64.1        |
| STS'12          | 64.8                  | 61.6        | 62.5        | 61.5                   | 58.7        | 30.8         | 56.2                | 59.5   | 64.9         | 65.8        | <b>68.3</b> |
| headlines       | 77.3                  | 76.9        | 76.1        | 73.9                   | 69.2        | 34.6         | 69.2                | 74.1   | 74.8         | 77.7        | <b>81.2</b> |
| OnWN            | 81.2                  | 72.8        | 81.4        | 73.8                   | 72.9        | 10.0         | 82.8                | 82.0   | 85.5         | 85.4        | <b>87.5</b> |
| FNWN            | 53.2                  | 50.2        | 55.6        | 47.7                   | 36.6        | 30.4         | 39.4                | 52.4   | 55.0         | <b>56.9</b> | 53.6        |
| SMT             | 40.7                  | 38.0        | 40.3        | 40.4                   | 29.6        | 24.3         | 37.9                | 38.5   | 39.1         | 41.0        | <b>42.0</b> |
| STS'13          | 63.1                  | 59.4        | 63.4        | 58.9                   | 52.1        | 24.8         | 56.6                | 61.8   | 63.6         | 65.2        | <b>66.1</b> |
| deft forum      | 56.6                  | 55.6        | 55.7        | 53.4                   | 37.5        | 12.9         | 41.2                | 51.4   | 52.3         | 55.0        | <b>58.5</b> |
| deft news       | 78.0                  | <b>78.5</b> | 77.1        | 74.4                   | 68.7        | 23.5         | 69.4                | 72.6   | 74.6         | 74.5        | 78.4        |
| headlines       | 74.5                  | 75.1        | 72.8        | 71.5                   | 63.7        | 37.8         | 64.7                | 70.1   | 69.7         | 73.3        | <b>78.7</b> |
| images          | 84.7                  | 85.6        | 85.8        | 80.4                   | 72.5        | 51.2         | 82.6                | 84.8   | 84.3         | 84.9        | <b>87.1</b> |
| OnWN            | 84.9                  | 81.4        | 85.1        | 81.5                   | 75.2        | 23.3         | 82.8                | 84.5   | 86.9         | 87.6        | <b>88.9</b> |
| tweet news      | 76.3                  | 78.7        | 78.7        | 77.4                   | 65.1        | 39.9         | 70.1                | 77.5   | 78.5         | <b>80.0</b> | 78.8        |
| STS'14          | 75.8                  | 75.8        | 75.9        | 73.1                   | 63.8        | 31.4         | 68.5                | 73.5   | 74.4         | 75.9        | <b>78.4</b> |
| answers-forum   | 71.8                  | 70.6        | <b>73.1</b> | 69.1                   | 45.6        | 36.1         | 63.9                | 70.1   | 72.4         | 72.2        | 72.0        |
| answers-student | 71.1                  | 75.8        | 72.9        | <b>78.0</b>            | 63.9        | 33.0         | 70.4                | 75.9   | 72.7         | 75.1        | 73.5        |
| belief          | 75.3                  | 76.8        | 78.0        | 78.2                   | 75.3        | 24.6         | 71.8                | 75.3   | 76.5         | 77.3        | <b>79.1</b> |
| headline        | 79.5                  | 80.3        | 78.6        | 76.4                   | 70.9        | 43.6         | 70.7                | 75.9   | 75.9         | 78.8        | <b>83.1</b> |
| images          | 85.8                  | 86.0        | 85.8        | 83.4                   | 72.9        | 17.7         | 81.5                | 84.1   | 83.0         | 84.6        | <b>87.5</b> |
| STS'15          | 76.7                  | 77.9        | 77.7        | 77.0                   | 60.6        | 31.0         | 71.7                | 76.3   | 76.1         | 77.6        | <b>79.0</b> |
| SICK14          | 71.3                  | 72.4        | 72.9        | 72.7                   | 69.4        | 49.8         | 72.2                | 72.9   | 73.0         | 72.3        | <b>73.5</b> |

Table 4: Results (Pearson’s  $r \times 100$ ) on textual similarity tasks. The highest score in each row is in bold. “GloVe+UP” is the application of uSIF-weighting (U) and piecewise common component removal (P) to GloVe word vectors; “PSL+UP” to PSL word vectors; “ParaNMT+UP” to ParaNMT word vectors.