pairs that are both paraphrases and textual entailment are pairs #1 and #2 in Table 7.8. Furthermore, both paraphrases and entailment have a negative correlation with contradiction, which is expected and confirms the quality of our data.

Specificity does not have any strong correlation with any of the other relations, showing that it is independent of those in our corpus.

**Table 7.6** Correlation between all relations

|      | TE  | UTE | BiTE | Cont | Spec | ∅ Sim |
|------|-----|-----|------|------|------|-------|
| PP   | .75 | .20 | .70  | -.25 | -.01 | 3.77  |
| TE   |     | .57 | .66  | -.30 | -.01 | 3.59  |
| UTE  |     |     | -.23 | -.17 | -.04 | 3.21  |
| BiTE |     |     |      | -.20 | -.01 | 3.89  |
| Cont |     |     |      |      | -.09 | 1.45  |
| Spec |     |     |      |      |      | 2.27  |

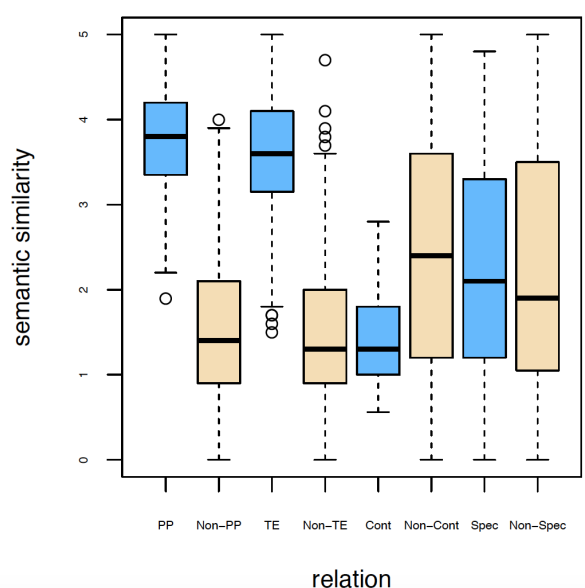## 7.4.1.2 Binary Relations and Semantic Similarity



Figure 7.1: Similarity scores of sentences annotated with different relations

We look at the average similarity for each relation (see Table 7.6) and show boxplots between relation labels and similarity ratings (see Figure 7.1). Table 7.6

shows that bi-directional entailment has the highest average similarity, followed by paraphrasing, while contradiction has the lowest.

Figure 7.1 shows plots of the semantic similarity for all pairs where each relation is present and all pairs where it is absent. The paraphrase pairs have much higher similarity scores than the non-paraphrase pairs. The same observation can be made for entailment. The contradiction pairs have a low similarity score, whereas the non-contradiction pairs do not have a clear tendency with respect to similarity score. In contrast to the other relations, pairs with and without specificity do not have any consistent similarity score.

## 7.4.2   Overlap of Relation Labels

Table 7.7 shows the overlap between the different binary labels. Unlike Pearson correlation, the overlap is asymmetric - the % of paraphrases that are also entailment (UTE in PP) is different from the % of entailment pairs that are also paraphrases (PP in UTE). Using the overlap measure, we can identify interesting interactions between phenomena and take a closer look at some examples.

**Table 7.7** Distribution of overlap within relations

|          | PP   | UTE  | BiTE | Contra | Spec |
|----------|------|------|------|--------|------|
| In PP    |      | 28 % | 64 % | 0      | 73 % |
| In UTE   | 52 % |      | -    | 0      | 73 % |
| In BiTE  | 94 % | -    |      | 0      | 72 % |
| In Contra| 0    | 0    | 0    |        | 63 % |
| In Spec  | 30 % | 17 % | 21 % | 11 %   |      |

### 7.4.2.1   Entailment and Paraphrasing Overlap

In a more theoretical setting, bi-directional entailment is often defined as being paraphrases [Madnani and Dorr, 2010, Androutsopoulos and Malakasiotis, 2010, Sukhareva et al., 2016]. This implies that paraphrases equal bi-directional entailment. In our corpus, we can see that only 64% of the paraphrases are also annotated as bi-directional entailment. An example of a pair that is annotated both as paraphrase and as bi-directional entailment is pair #10 in Table 7.8. However, in the corpus we also found that 28 % of the paraphrases are only uni-directional entailment, while in 8% annotators did not find any entailment. An example of a pair where our annotators found paraphrasing, but not entailment is sentence pair #5 in Table 7.8. The agreement on the paraphrasing for this pair was 80%, the agreement on (lack of) forward and backward entailment was 80% and 70% respectively. Although the information in both sentences is nearly identical, there is

no entailment, as "having a higher chance of getting smth" does not entail "getting smth" and vice versa.

**Table 7.8** Annotations of sentence pairs on all meaning relations taken from our corpus

| # | Sentence 1 | Sentence 2 | PP | FTE | BTE | Cont | FSpec | BSpec | Sim |
|---|---|---|---|---|---|---|---|---|---|
| 1 | The importance of technology in modern education is overrated. | Technology is not mandatory to improve education | ✓ | ✓ | ✓ | | | | 2.8 |
| 2 | Machines cannot interact with humans. | No machine can communicate with a person. | ✓ | ✓ | ✓ | | | | 4.9 |
| 3 | The modern assistants make finding data slower. | Today's information flow is greatly facilitated by digital assistants. | | | | ✓ | | ✓ | 1.9 |
| 4 | The bible is in Hebrew. | Bible is not in Latin. | | ✓ | | | ✓ | | 2.7 |
| 5 | All around the world, girls have higher chance of getting a good school education. | Girls get a good school education everywhere. | ✓ | | | | | ✓ | 4.7 |
| 6 | Reading the Bible requires studying Latin. | The Bible is written in Latin. | | ✓ | ✓ | | | ✓ | 3.6 |
| 7 | Speaking more than one language can be useful. | Languages are beneficial in life. | ✓ | ✓ | ✓ | | | ✓ | 4.4 |
| 8 | You can find a good job if you only speak one language. | People who speak more than one language could only land pretty bad jobs. | | | | ✓ | | | 2.3 |
| 9 | All Christian priests need to study Persian, as the Bible is written in Ancient Greek. | Christian clergymen don't read the bible. | | | | | | ✓ | 0.9 |
| 10 | School makes students antisocial. | School usually prevents children from socializing properly. | ✓ | ✓ | ✓ | | | ✓ | 3.9 |

If we look at the opposite direction of the overlap, we can see that 52% of the uni-directional and 94% of the bi-directional entailment pairs are also paraphrases. This finding confirms the statement that bi-directional entailment is paraphrasing (but not vice versa).

There is also a small portion (6%) of bi-directional entailments that were not annotated as paraphrases. An example of this is pair #6 in Table 7.8. Although both sentences make each other true, they do not have the same content.

Neither paraphrasing nor entailment had any overlap with contradiction, which further verifies our annotation scheme and quality.

These findings are partly due to the more "relaxed" definition of paraphrasing adopted here. Our definition is consistent with other authors that work on paraphrasing and the task of paraphrase identification, so we argue that our findings are valid with respect to the practical applications of paraphrasing and entailment and their interactions.