

## A Supplemental Material

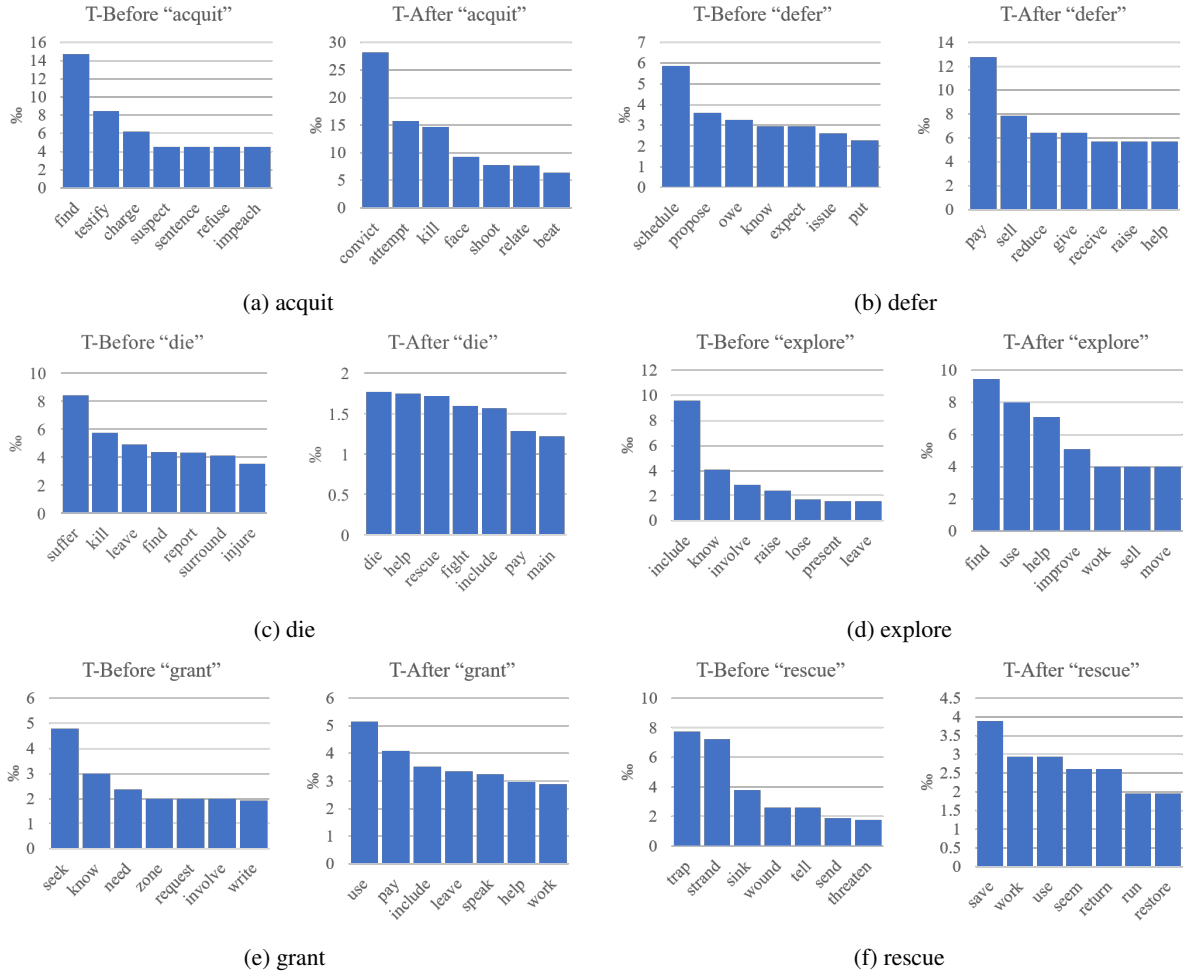


Figure 2: A few more examples in addition to Fig. 1.

In Sec. 3.4.2, we have shown some interesting examples with events that temporally precede or follow a certain event. In addition to those examples shown in Fig. 1, it is actually very easy to find more interesting examples. We have picked some and shown them in Fig. 2. Again, there are potential errors in these figures, but the overall quality is intuitively appealing.

In Sec. 4.1, we have shown the quality analysis of TEMPROB in Tables 3-4 via the TimeBank-Dense dataset (Cassidy et al., 2014) and the EventCausality dataset (Do et al., 2011). Here we further provide confidence level analysis on some randomly selected event pairs. Specifically, we performed a 10-fold bootstrapping. In each fold, we randomly selected 50% of the graphs in TEMPROB (i.e., from  $\{G_i\}_{i=1}^N$ ) with replacement. Then we re-calculated the prior statistics  $\{f_r(v_i, v_j)\}$  (Eq. (5)) in each fold. By considering  $\{f_r(v_i, v_j)\}$  as a random variable, we now obtained 10 realizations of it, so we can evaluate the confidence level. In Fig. 3 (with color), we show the confidence levels of several randomly selected pairs of events. From Fig. 3, we can see that the prior statistics are actually rather stable, indicating that the prior statistics represented by TEMPROB is indeed a notion of knowledge underlying natural language text.

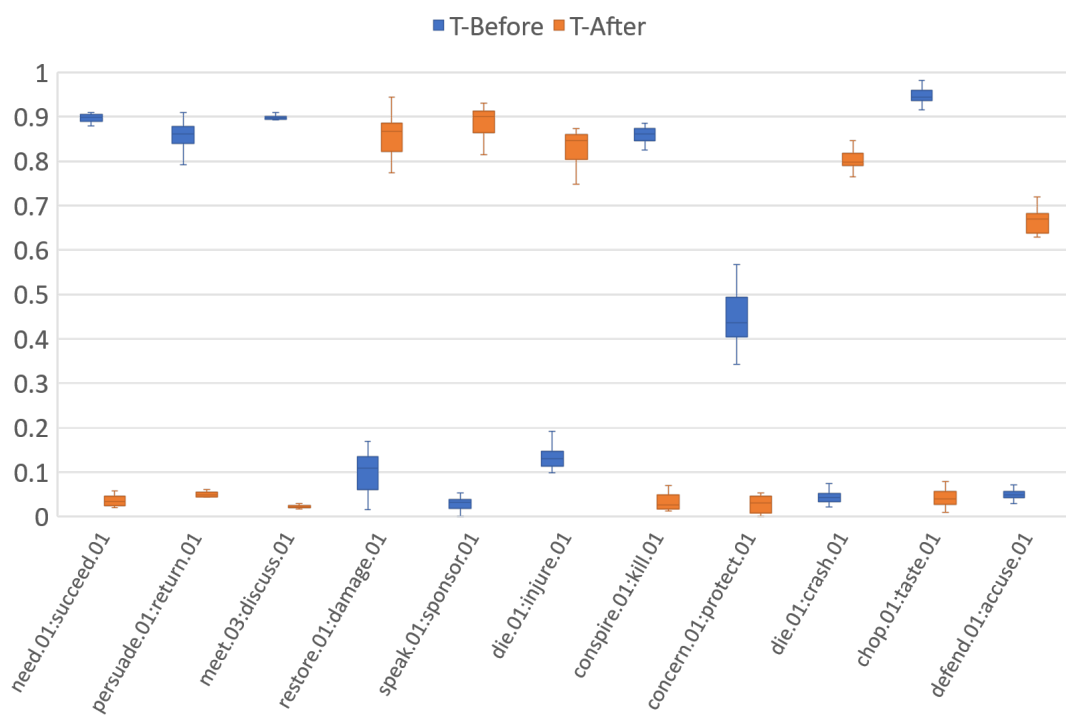


Figure 3: **Confidence intervals** from 10-fold bootstrapping.