A Model Hyperparameters and Training Details

Component	Dimensions	Layers
Word Embeddings	100	
Character Embeddings	50	
Character LSTM	100	1
Sentence LSTM	250	2
Label Feedforward Network	250	1

Table 2: The sizes of the components used in our model.

Our model hyperparameters are summarized in Table 2. We train using the Adam optimizer (Kingma and Ba, 2014) with its default hyperparameters for 40 epochs. We evaluate on the development set 4 times per epoch, selecting the model with the highest overall development performance as our final model. When performing a word embedding lookup during training, we randomly replace words by the <UNK> token with probability $1/(1 + \text{freq}(w))$, where $\text{freq}(w)$ is the frequency of a word w in the training set. We apply dropout with probability 0.4 before and inside each layer of each LSTM. Our system is implemented in Python using DyNet (Neubig et al., 2017).

B Character LSTM Word Feature Classification

Binary Feature	Majority Class	Char-LSTM Classifier	Binary Feature	Majority Class	Char-LSTM Classifier
all-letters	77.22%	99.77%	suffix = “s”	82.65%	99.99%
has-letter	89.18%	99.97%	suffix = “ed”	92.52%	99.98%
all-lowercase	56.95%	99.95%	suffix = “ing”	93.26%	99.95%
has-lowercase	85.85%	99.90%	suffix = “ion”	97.75%	99.93%
all-uppercase	96.68%	99.90%	suffix = “er”	96.42%	99.97%
has-uppercase	67.77%	99.97%	suffix = “est”	99.63%	99.98%
all-digits	98.38%	99.99%	suffix = “ly”	97.56%	99.99%
has-digit	87.90%	99.91%	suffix = “ity”	99.30%	99.94%
all-punctuation	99.93%	99.98%	suffix = “y”	92.97%	99.93%
has-punctuation	79.04%	99.75%	suffix = “al”	98.48%	99.92%
has-dash	88.89%	99.95%	suffix = “ble”	99.30%	99.90%
has-period	92.55%	99.95%	suffix = “e”	89.57%	99.99%
has-comma	98.02%	99.97%			

Table 3: Classification accuracy for various binary word features using the character LSTM representations for words induced by a pre-trained parser. Performance substantially exceeds that of a majority class classifier in all cases, reaching 99.7% or higher for all features. The majority class is `True` for the first four features in the left column and `False` for the rest.