Empirical Analysis of the 19 Linguistic Patterns

I. Contribution of Individual Linguistic Patterns

We analyzed the contribution of each of the 19 linguistic patterns to the summarization accuracy of ASSORT$_S$. In particular, we created variants of ASSORT$_S$ by removing each linguistic pattern at a time, as well as a variant without all linguistic patterns. We use the same hyperparameters and dataset as in the original paper to train these model variants. Then, we measured the accuracy degradation of ASSORT$_S$ in terms of the F-1 score.

Table I shows the result. Column F1 shows the F1 score of the model variants in comparison to the original ASSORT$_S$ model (the first row in Table I). Column Δ shows the accuracy difference after removing the corresponding linguistic patterns. The last row in Table I shows the model accuracy after removing all linguistic patterns.

While removing all patterns led to a non-trivial 4% decrease in the F1 score, some patterns were more influential than others. For example, the removal of *"First, ..."* caused the most decrease, 2%, among all patterns. A possible reason for the strong influence of this pattern is that many Stack Overflow users tend to use "First" to start a sentence with important information in the answer post, followed by some elaborations.

TABLE I: Contribution of each linguistic pattern in ASSORT$_S$

	F1	Δ
ASSORT$_S$	**0.71**	**-**
—w/o However,...	0.70	-0.01
—w/o In practice, ...	0.70	-0.01
—w/o In fact, ...	0.70	-0.01
—w/o First, ...	0.69	-0.02
—w/o Otherwise, ...	0.71	-
—w/o In this case, ...	0.71	-
—w/o If you care, ...	0.71	-
—w/o In general, ...	0.70	-0.01
—w/o In contrast, ...	0.70	-0.01
—w/o Finally, ...	0.70	-0.01
—w/o In short, ...	0.71	-
—w/o Then, ...	0.70	-0.01
—w/o On the other hand, ...	0.71	-
—w/o Alternatively, ...	0.71	-
—w/o Below is ...	0.71	-
—w/o In other words, ...	0.71	-
—w/o Additionally, ...	0.71	-
—w/o In addition, ...	0.70	-0.01
—w/o Furthermore, ...	0.70	-0.01
—w/o **all patterns**	0.67	-0.04

"-" in Column Δ means a change smaller than 1%