Appendix A. Best Hyperparameters

The table below displays the best hyperparameters of each method and dataset, yielding the highest $F1$ on the validation set.

Models	K&H+N	BLESS	ROOT09	EVALution
NPB	$dr = 0.2$	$dr = 0.2$	$dr = 0.0$	$dr = 0.0$
NPB+Aug	$k = 5, dr = 0.2$	$k = 5, dr = 0.2$	$k = 5, dr = 0.2$	$k = 5, dr = 0.2$
LexNET	$dr = 0.4$	$dr = 0.4$	$dr = 0.2$	$dr = 0.0$
LexNET_h	$dr = 0.4$	$dr = 0.4$	$dr = 0.4$	$dr = 0.2$
LexNET+Aug	$k = 1, dr = 0.4$	$k = 5, dr = 0.2$	$k = 5, dr = 0.4$	$k = 5, dr = 0.4$
LexNET+Rep	$dr = 0.4$	$dr = 0.2$	$dr = 0.2$	$dr = 0.0$
LexNET+Aug+Rep	$k = 5, dr = 0.4$	$k = 5, dr = 0.2$	$k = 5, dr = 0.2$	$k = 3, dr = 0.2$

Table 1: Best hyperparameters of each model in each dataset.

Appendix B. Training Time per Epoch of +Aug in BLESS

The table below displays the number of processed paths and elapsed seconds per epoch during the training time in BLESS when training LexNET+Aug with a NVIDIA GeForce GTX-1080Ti. In our experiments, the training set of BLESS has 10,215 instances and 213,086 paths, including the padding paths. Thus, the number of paths processed by neural path-based methods is $213086 + 2 \times k \times 10215$.

k	the number of paths	elapsed seconds per epoch
$k = 0$	213086	335
$k = 1$	233516	402
$k = 3$	274376	441
$k = 5$	315236	484

Table 2: The number of processed paths and training times per epoch in BLESS when training LexNET+Aug.