Quality-Diversity Trade Off Other manual-machine text comparison score 34 Retain only mixed-case Remove repeating lines Removing upper-case, repeating lines Removing upper, lower-case, repeating lines 32 seg_edit_dist <= 0.2, edit_dist <= 0.2 10 datasets (%) seg edit dist <= 0.4, edit dist <= 0.4 seg_edit_dist <= 0.6, edit_dist <= 0.4 30 seg_edit_dist <= 0.7, edit_dist <= 0.5 Final filters 28 across WER 26 Average 22 45 50 30 35 55 60 65 40 70 Percent of Data Remaining (%)