Perplexity vs Quantity of Gold Standard Fine-tuning Data Fully Trained GPT-2 with Synthetic and Gold-Standard Fine-Tuning Data Frisian Occitan Cebuano 2×10^{3} 10^{3} 10^2 10^{3} Perplexity 6×10^2 6×10^{1} 6×10^{2} 4×10^2 4×10^{1} 3×10^2 3×10^{1}

Fine-tune Data Quantity

(1000s of gold-standard sentences)

0

100

200 Pure

Hybrid

 4×10^{2}

200

100

0