5-gram Overlap versus Token Length for simpy