DESIGNING TEXT

I want to pick up a comment or rather link that Sam left on the Blather site entry – to Eric Benson's blog.

Archaeological text generation – Markov chains

… The official definition of a markov chain is: A model of sequences of events where the probability of an event occurring depends upon the fact that a preceding event occurred. The way Stephen has implemented it is through a script that takes a collection of text and creates a mapping of words dependant on the relationship each word has to those that precede it. For example, in this sentence, the word "example" is preceded by "for" and "preceded" is preceded by "is" (twice). In this example, it's only looking one word back, but you can also look multiple words back to see that "sentence" is preceeded by "in this" in the example sentence. Using this information, you can rebuild new sentences. Basically, you just give it a starting word, like "For" and it can look for all words that came after "For" and slowly add on to the sentence in a way that will feel organic. The further you looked back when building the chains, the more natural the sentences will feel because they will be build with

blocks that had more context in them. There’s also a greater chance that they will end up building sentences that actually existed in the original text. So you have to find that balance where you’re building new sentences, and they feel natural, and yet they didn’t exist in the original text.