

Generative Text

David Ye

August 17, 2023

1 Examples

The below 50 word examples were generated with maximum 4-gram.

1. **Romeo and Juliet**

Generated Text: Thy tempest-tossed body. How oddly thou weep'st not mercy. Heaven is here Where Juliet lives, and every tongue shall smooth thy name, When but weeps and weeps; And now falls on her bed, and mistress. I am so he for the numbers that Petrarch flowed in. Laura, to his lady, fie,

2. **The Odyssey**

Generated Text: Sing the suitors some one of imposing presence and superhuman strength, whereas he turns out to be able to string such a fine son as you will, and listen to our lands, neither here nor elsewhere, till she has made her choice and married some one or other of us."

3. **Peter Pan**

Generated Text: moon was riding in a jug for Hook felt curiously warm. "Smee," he said, "this seat is hot." He jumped up. "Odds bobs, hammer and tongs I'm very sorry, but how could I know you were in the city to the ground with an arrow in her breast.

4. **Love Story (Taylor Swift)**

Generated Text: love story, baby, just say, "Yes" So I sneak out to the garden to see the party, the ball gowns See you make your way through the crowd And say, "Yes" Oh, oh I got tired of waiting Wonderin' if you were ever comin' around My faith in you was fading

5. **NPR: NYC officials are trying to stop the deadly social media trend of 'subway surfing'**

Generated Text: Subway Surfing kills. We need everyone to be a part of ending this dangerous threat," Adams tweeted on X, the social media platform formerly known as Twitter. In response to the boy's death, TikTok spokesperson Michael Cortez. Kemper also said there have been "dozens of apprehensions and over 70

6. **CNN: Walmart is up. Target is down. Here's why**

Generated Text: Walmart for groceries and essentials. Walmart's sales at stores open for at least a year grew 6.7

7. **Song Mix (Lemon Tree, Somewhere Only We Know , Fade Into You, Die 4 Me, Live or Die, Summertime Sadness)**

Generated Text: love it on the low And you don't have to say I'm crazy 'cause I want to take the breath that's true I look to you, and I see nothing I Got you tonight Have you ever seen a red flower? Yeah Have you ever seen a red flower? Yeah Have

8. **Lady Gaga (Poker Face, Bad Romance, Shallow)**

Generated Text: love and all your lover's revenge You and me could write a bad romance (oh-oh-oh-oh-oh) I want your love Love, love, love, I want your love (love, love, love) (I want your love) You know that I want you And you know that I want you And you know that I

9. **170+ Pickup Lines**

Generated Text: I lost my teddy bear. Can I have in common? We both want to be part of your world. I thought this was a bar, but you owe me a drink (pause), because when I saw you, my heart stopped. If I were a cat, I'd spend all nine of my

10. **100+ Funny Short Jokes**

Generated Text: Knock knock. Who's there? A little old lady. A little old lady who? Hey, you can yodel! What do you do if you see a spaceman? Park in it, man. Velcro. What a ripoff. What do you call a parade of rabbits hopping backwards? A receding hare-line. What is sticky hair?

2 Analysis

1. **Corpus Size vs. Sentence Quality:** The larger the corpus, the more 'original' the sentences are. However, these 'original' sentences usually do not have the best punctuation, interpretation, or flow. Moreover, consecutive sentences do not feel very connected to each other. On the other hand, smaller corpuses tend to generate less 'original' sentences, and directly copy the corpus a lot more. And because the sentences generated tend to match up with the corpus, they usually have better quality (punctuation, interpretation, flow, etc.) than sentences generated from larger corpuses.
2. **Maximum n -gram length vs. Sentence Quality:** As the maximum n -gram length increases, the sentence quality increases. However, at the same time the sentence originality decreases (even with back-off smoothing). When the maximum n -gram is small, the sentences do not make much sense, however are very original. A comparison of sentences generated with different maximum n -gram length can be found in table [1](#).

n -gram	Text Generated
2	Thy tears are womanish, thy wild acts denote The unreasonable fury of mine ancient ears. Lo here upon you for some private place, And go, Get you to Juliet ere you to Juliet is alive, For this so sudden business. LADY CAPULET. That is no slander, sir, which is a
4	Thy form cries out thou art. Thy tears are womanish, thy wild acts denote The unreasonable fury of a sigh, Speak but lately dead. There art thou hurt? MERCUTIO. Ay, ay, a scratch, a scratch. Marry, 'tis enough. Where for this once.—What, ho!— They are all forth: well, thou art
8	Thy old groans yet ring in mine ancient ears. Lo here upon thy cheek the stain doth sit Of an old tear that lies she, too desperate, would have thought it? Romeo! JULIET. What devil art thou, that dost torment me thus? This torture should be roar'd in dismal hell.
16	Thy noble shape is but a form of wax, Digressing from the valour of a man; Thy dear love sworn but hollow perjury, Killing that love which thou hast vow'd to cherish; Thy wit, that ornament to shape and love, Misshapen in the conduct of them both, Like powder in a
32	Thy old groans yet ring in mine ancient ears. Lo here upon thy cheek the stain doth sit Of an old tear that is not wash'd off yet. If ere thou wast thyself, and these woes thine, Thou and these woes were all for Rosaline, And art thou chang'd? Pronounce this

Table 1: Table comparing 50 word texts generated from Romeo and Juliet with various max n -grams.

3 Originality

From before, we've seen that when we train on a large corpus with a small max n -gram, we generate pretty original sentences. However, if we train on a small corpus with a large max n -gram, we generate sentences that likely already exist in the corpus. Furthermore, the higher the backoff probability the more original the sentences. And likewise the lower the backoff probability the less original the sentences are. Some examples of sentences generated from The Odyssey (max n -gram of 16) with various backoff probability are shown in table 2.

Backoff Probability	Text Generated
$\frac{1}{1+f_1+\dots+f_l}$	Sail night and day, and keep your ship well away from the islands; the god who watches over you and protects you will send you a fair wind. As soon as you get to Ithaca send your ship and men on to the town, but yourself go straight to the swineherd
$\frac{2}{2+f_1+\dots+f_l}$	Sail night and day, and keep your ship well away from the islands; the god who watches over you and protects you will send you a fair wind. As soon as you get to Ithaca send your ship and men on to the town, but yourself go straight to the swineherd
$\frac{4}{4+f_1+\dots+f_l}$	Sail night and day, and keep your ship well away from the islands; the god who watches over you and protects you will send you a fair wind. As soon as you get to Ithaca send your ship and men on to the town, but yourself go straight to the swineherd
$\frac{8}{8+f_1+\dots+f_l}$	Sail night and day, and keep your ship well away from the islands; the god who watches over you and protects you will send you a fair wind. “And I saw them their spoil—even they, and wrecked it. Nestor and think of many a word of all those Danaans tore
$\frac{12}{12+f_1+\dots+f_l}$	Sail night and day, and keep your ship well away from the islands; the god who watches over which they say such a desert island of him, you or no thought to the suitors, who comes begging about which I saw that heaven meant me to get inside that was the

Table 2: Table comparing 50 word texts generated from The Odyssey (max n-gram of 16) with various backoff probability.