

POETWRITER: A CONSTRAINT-BASED TEXT GENERATOR

Mathieu Rolfo, Shalom Rottman-Yang, Nathan Tindall

CS221 Final Project, Fall 2014

PROBLEM DEFINITION

Our goal is to develop a text generator that produces original poetry given a body of previous work and constraints on poetic form.

CHALLENGES

- Impose rhyming and syllabic constraints on output
- Impart semantic meaning on output
- Determine rhyming pairs
- Determine syllabic structure
- Minimize search space to improve runtime

APPROACHES

Algorithm

The generation is represented as a **depth-first search** problem where search backtracks upon constraint violation, where **constraints** are enforced **within lines** (syllables) and **between lines** (rhyming).

Generation

The search problem is represented thus:

State: a (Poem, seed) tuple,

where seed is the ***n*-gram** sequence

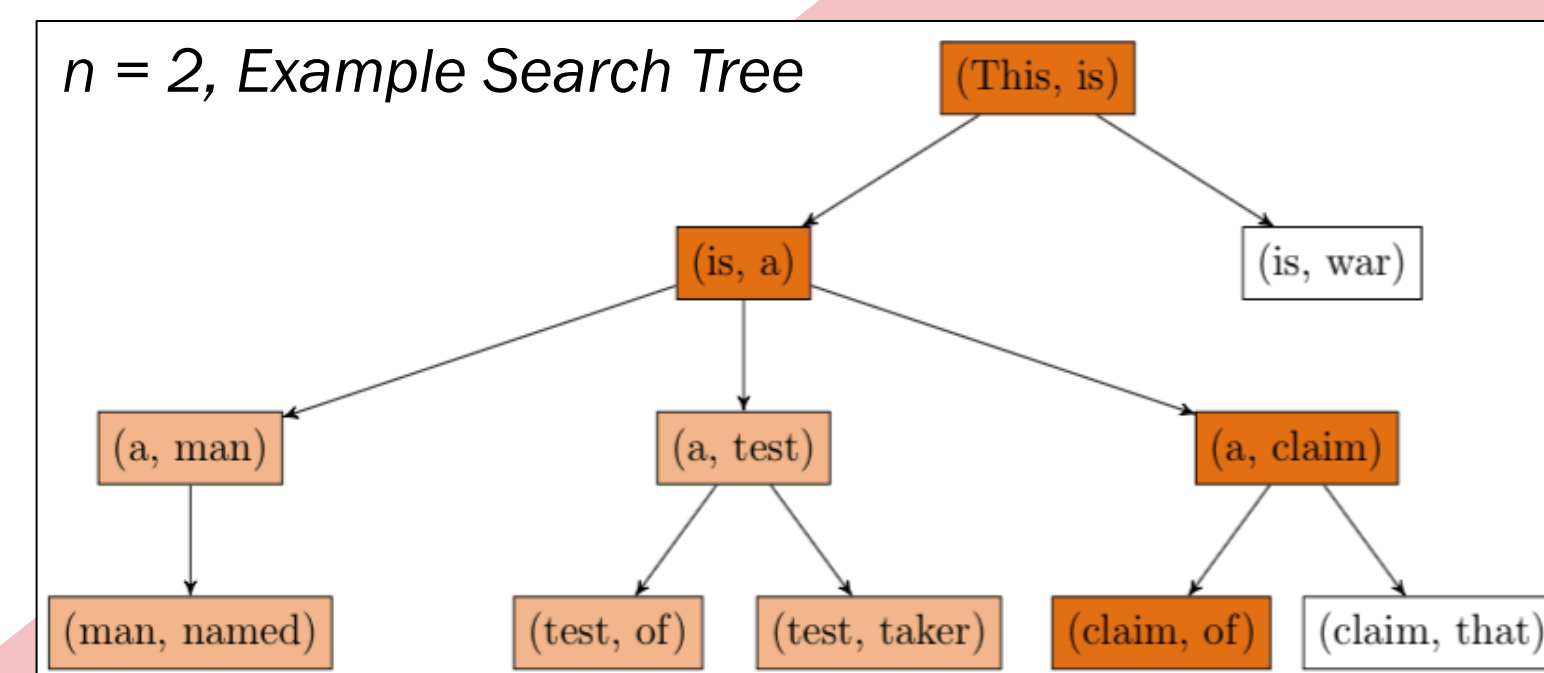
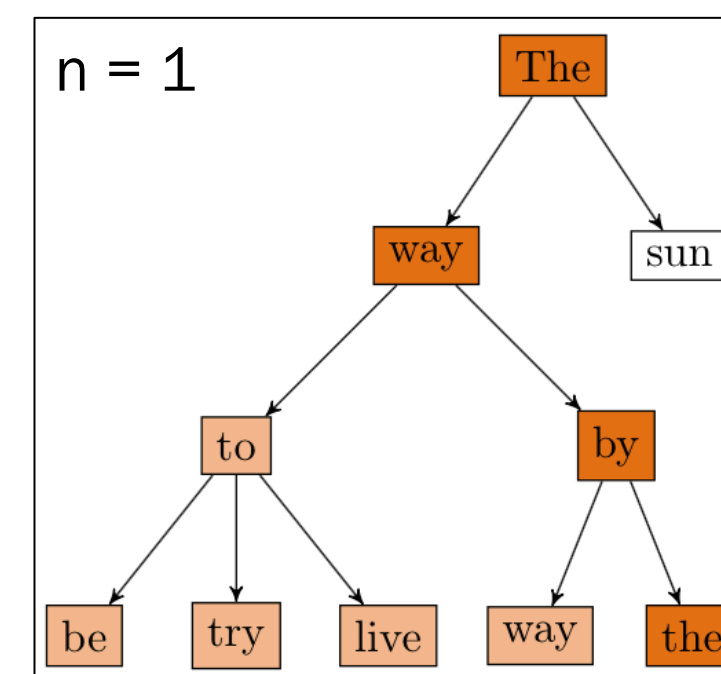
Goal state: any state with a fully completed Poem

Action: adding another word to the Poem

Successors: all states with a legitimate *n*-gram-

generated word added to the Poem's current Line

Cost: zero



Model

Our **Poetry object** stores:

A list of **Line objects**, which store:

- Number of completed syllables in the line
- Other lines it will rhyme with

The current Line

The number of completed Lines

MOTIVATION

Poetry is one of the most expressive and notoriously difficult ways to use language. It is an interesting challenge to produce poetry that matches the **qualitative constraints** of style and the **formal constraints** of rhyme and meter.

RESULTS

The corpora for these poems were the CS221 lecture note text outlines and Kanye West's lyrics, scraped from rapgenius.com. Both utilize 3-grams.

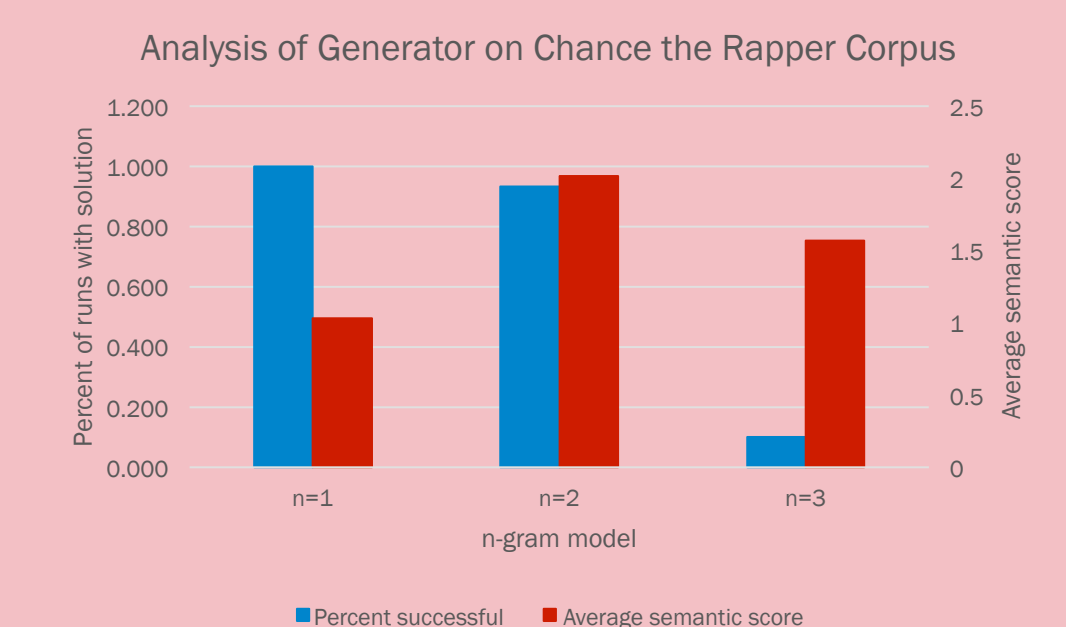
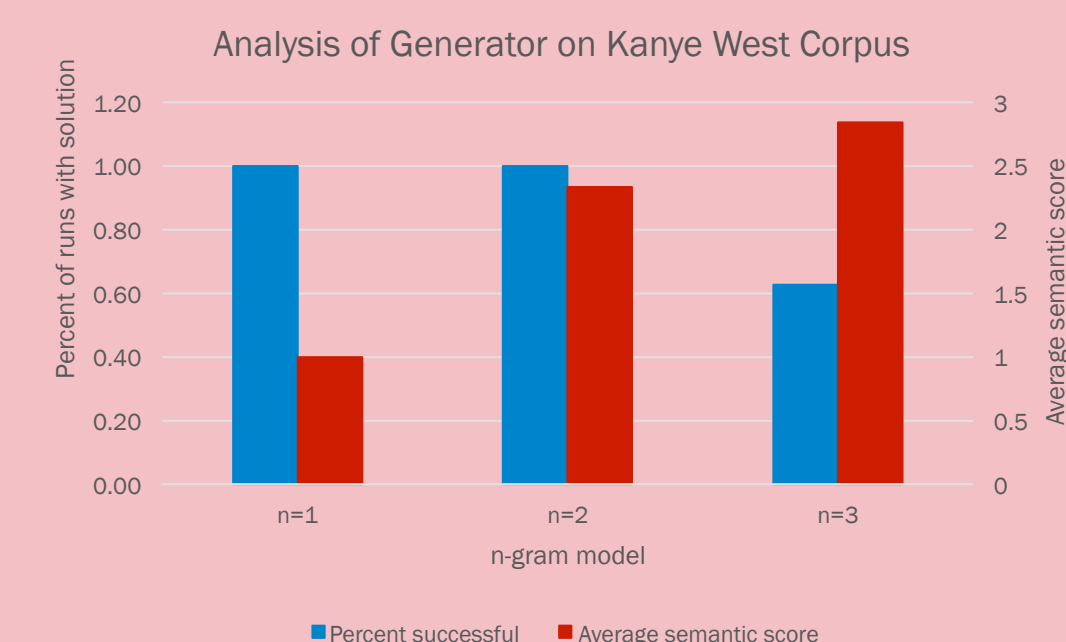
RESULT
source text: percy_221_lecture_1.txt

to make optimal choices in the last letter of ii we can proceed to the owner ca of the above problems and continuous optimization is a huge chasm between the first step is to express the quantity that we wish to compute the derivative gradient if is too small then we output negative where do the weights come from we can insert the last letters of and the first step is to write a

RESULT
source text: eminem.txt

sucked a lot of love hangin out with the thorn bush that you used to say i got a brand new horn to be tooted to sum it up quick so you can wake up the next day you thought i was the man of the house a thousand weeks and still countin even when i was three plus when i was in the back in a while i just shine on this microphone i will eat your ass up for the first time i went to fill

ANALYSIS



Notes: Poems had 8 lines, 8 syllables, every two lines rhyme. Used simplified heuristics for rhyme and syllable count. Kanye has 47273 lines; Chance has 8678 lines. Scraped from RapGenius. Percent completed is percent w/ solution found. Semantic score was from 1-5. 1 is "no semantic meaning whatsoever". 5 is "fully semantically meaningful". Scores evaluated by current Stanford students.