Swedish Rhyme Generation

Jag sitter här och pratar med Sofie
Försöker bara räkna pi
Sedan kommer ett bi
Då får jag pollenallergi

- *av Darkleonard*

Filip Jergle Almquist och Edvin Jönsson

Text generation in 2020

ARTIFICIAL INTELLIGENCE

OpenAI's new language generator GPT-3 is shockingly good—and completely mindless

The AI is the largest language model ever created and can generate amazing human-like text on demand but won't bring us closer to true intelligence.

July 20, 2020

TECH \ ARTIFICIAL INTELLIGENCE

A college student used GPT-3 to write fake blog posts and ended up at the top of Hacker News

He says he wanted to prove the AI could pass as a human writer

By Kim Lyons | @SocialKimLy | Aug 16, 2020, 1:55pm EDT

GPT-3, explained: This new language AI is uncanny, funny — and a big deal

Computers are getting closer to passing the Turing Test.

By Kelsey Piper | Aug 13, 2020, 9:50am EDT

Goal: generate rhyming couplets
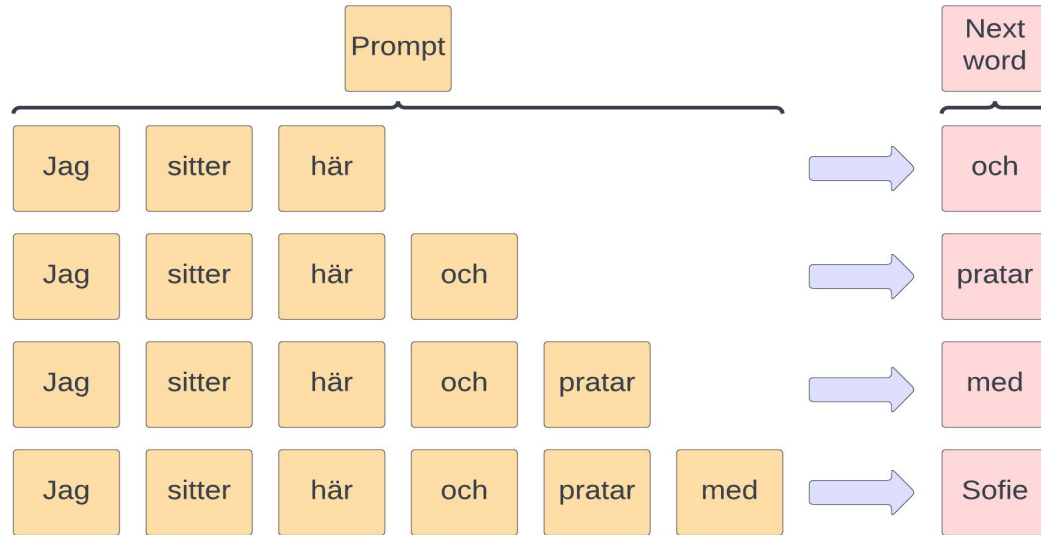
Given prompt: Sedan kommer ett bi

We generate: Då får jag pollenallergi

Rhyming is tricky

- Rhymes depend on pronunciation
- Pronunciation doesn't always follow spelling
- Single word can have multiple pronunciations depending on meaning
- planet can be stressed like
  - planét (translates to planet)
  - plánet (translates to airplane)
- Words that do not look similar can still rhyme
  - fleece
  - gris

How to generate text?

- Causal language modeling (GPT, PaLM)
- Look at previous words, guess the next

Prompt							Next word
Jag	sitter	här				→	och
Jag	sitter	här	och			→	pratar
Jag	sitter	här	och	pratar		→	med
Jag	sitter	här	och	pratar	med	→	Sofie

Ideas from other related projects

- Prompt engineering, requires more data/larger models
- Large rhyme dataset
- Does not guarantee rhymes

2-shot prompt from GPT-3 paper:

The City
BY C. P. CAVAFY
TRANSLATED BY EDMUND KEELEY
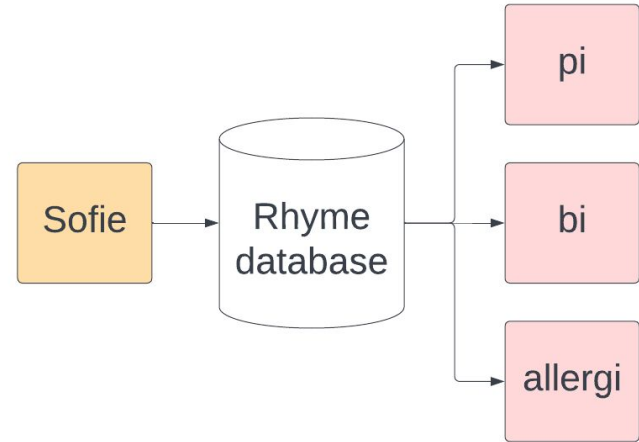[Poem text omitted]

SOME TREES
John Ashbery
[Poem text omitted]
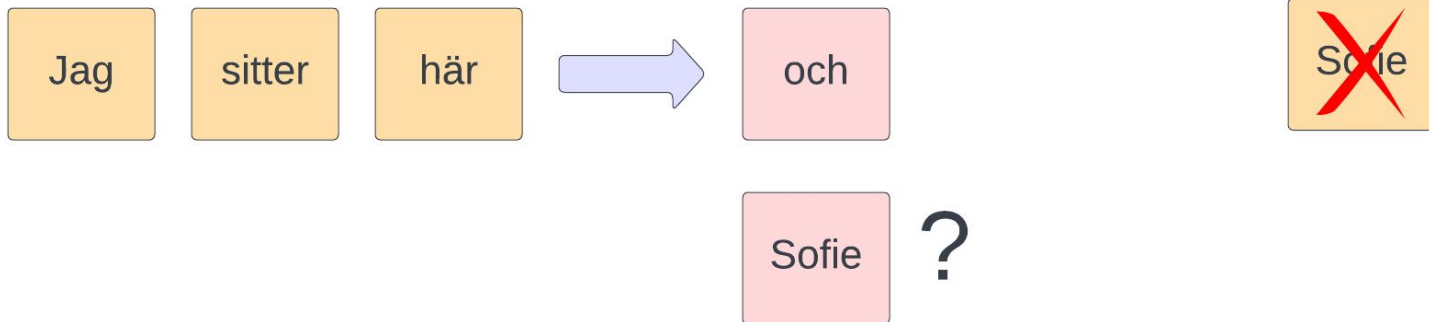
Shadows on the Way
Wallace Stevens

Use causal text generation, but force rhymes

- Use rhyme database webscraped from https://dblex.com/rimlexikon/
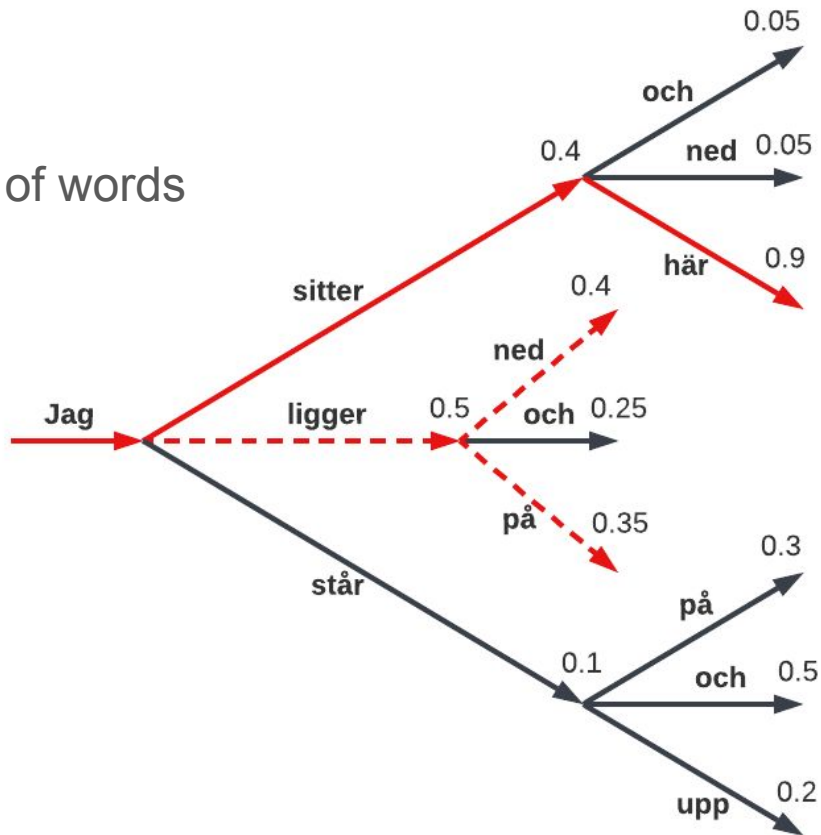- Explicitly add rhymes

Challenges with forcing rhymes

- Can't add rhyme first, because of causal LM
- When to add?
- How to keep cohesive?

Beam search

- Choose the most likely sequence of words
- Branch, rank, reduce, repeat
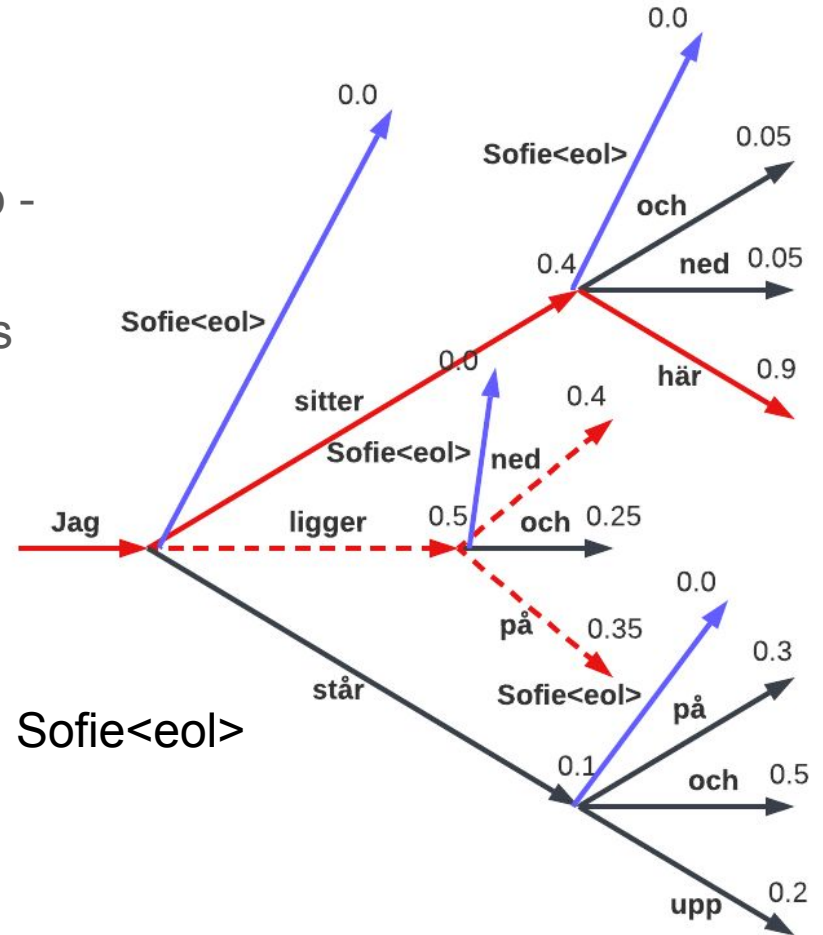
- Greedy search: 0.5 * 0.4 = 0.2

  "Jag ligger ned"

- Beam search: 0.4 * 0.9 = 0.36

  "Jag sitter här"

Constrained beam search

- Try to add constraint word at each step - extra branch
- More complicated when words consists of multiple tokens (flaggstång -> flagg, stå, ng)
- Train with special end-of-line token to make ending with rhyme most likely

Jag sitter här och pratar med Sofie<eol>
Försöker bara räkna pi<eol>
Sedan kommer ett bi<eol>
Då får jag pollenallergi<eol>

Backwards text generation

- Take a corpus, reverse the text and finetune
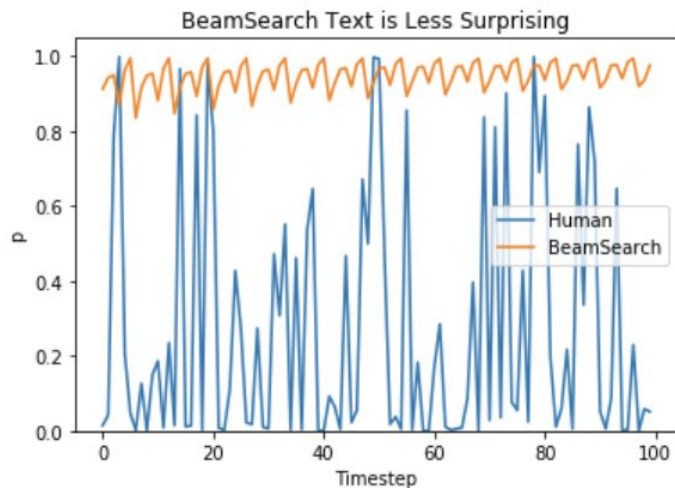- Now the rhyming word can be selected first

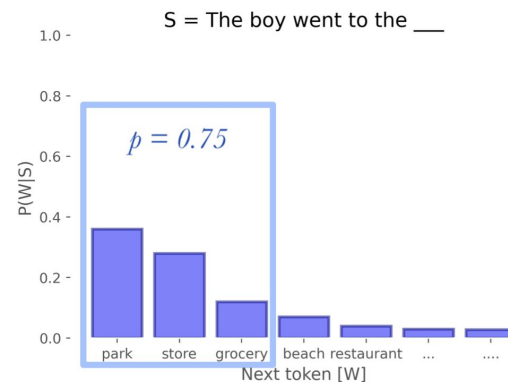Jag sitter här och pratar med **Sofie**
Försöker bara räkna **pi**

↓

Sofie med pratar och här sitter Jag
pi räkna bara Försöker

Backwards text generation

- Can use any sampling, more surprising
- Top-k, top-p, typical sampling

BeamSearch Text is Less Surprising

S = The boy went to the ___
$k = 4$

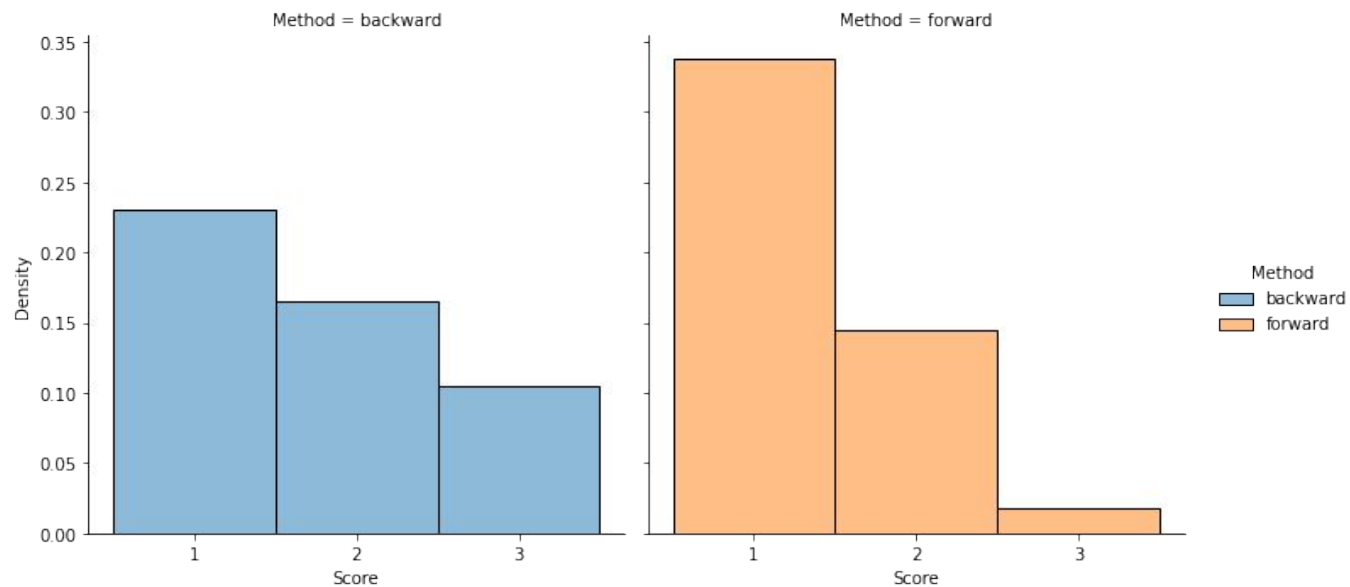S = The boy went to the ___
$p = 0.75$

Implementation details

- Pretrained 117M parameter GPT-2 from Hugging Face
- Swedish Wikipedia used for pretraining
- Created dataset of ~2600 couplets from Runeberg and Stagnelius poem collections, used for finetuning
- Trained with crossentropy loss on the predicted words

Results

- Qualitative analysis of 200 couplets, 100 forwards (constrained beam search), 100 backwards
- Evaluated on validation part from collected dataset
- Gave a score of 1, 2, or 3 to each generated line
- Cohen's Kappa, (inter annotator agreement) did we agree due to chance?

Combined scores from 2 annotators, higher is better

Results

- Average score forwards: 1.34 & 1.38
- Average score backwards: 1.87 & 1.63

- Cohen's Kappa score forwards: 0.197
- Cohen's Kappa score backwards: 0.409
- Poetry is subjective

Cohen's Kappa	Interpretation
0	No agreement
0.10 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Near perfect agreement
1	Perfect agreement

Examples

~translations

vid lekar och sång

(with play and song)

hör sköna vänner natten lång

(hear sweet friends all night long)

nu ur lundens sköte och ur blommans knopp

(now from the womb of the grove and the flowers bud)

och blomsterängen gungar opp

(and the flower meadow swings up)

satt med isad glaven

(sat with an iced glaive)

då varenda rustad steg ut ur graven

(when every armored rose from the grave)

Conclusion (very punchy)

Det är svårt att rimma

Och orden ihop limma

Nu kan maskinen dig (nästan) hjälpa

Att dina dåliga rim stjälpa

~translation

It is hard to rhyme

And the words together bind

Now the machine rhymes can (maybe) churn

And your bad rhymes overturn