

Connor Hanley

LIN 373

Venkat

August 15, 2022

## Assignment 5

### 1. Problem 1: Language Model. Given the corpus:

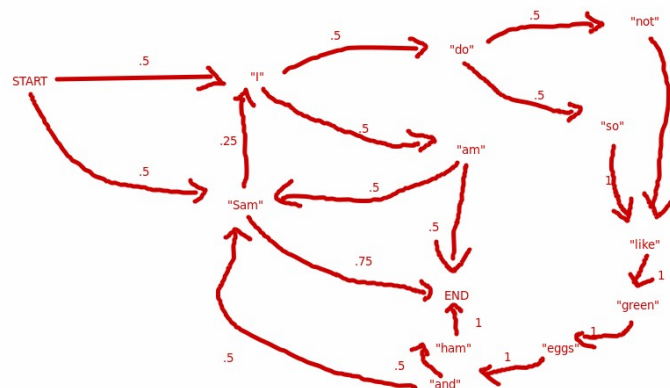
“START I am Sam END

START Sam I am END

START I am Sam END

START I do not like green eggs and Sam END

START I do so like green eggs and ham END”

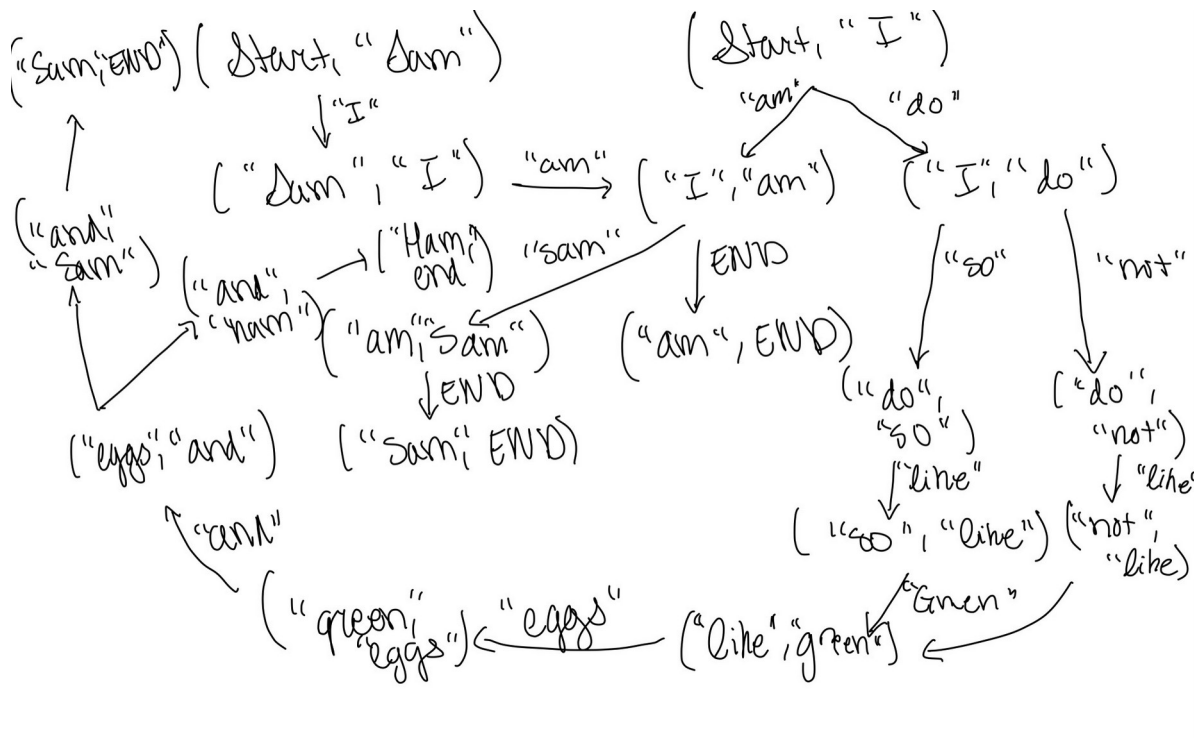


#### a) Markov Chain representing the corpus.

- The bigram model probability of “I do not like green eggs and Sam” can be framed as the conjunction of  $P(\text{END} \mid \text{“Sam”}) \& P(\text{“Sam”} \mid \text{“and”}) \& P(\text{“and”} \mid \text{“eggs”}) \& P(\text{“eggs”} \mid \text{“green”}) \& P(\text{“green”} \mid \text{“like”}) \& P(\text{“like”} \mid \text{“not”}) \& P(\text{“not”} \mid \text{“do”}) \& P(\text{“do”} \mid \text{“I”}) \& P(\text{“I”} \mid \text{START})$ . This is, following the chain,  $0.5 \cdot 0.5 \cdot 0.5 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 0.5$ , or 0.0625
- The probability of the sentence “I am Sam” is equivalent to  $P(\text{END} \mid \text{“Sam”}) \& P(\text{“Sam”} \mid \text{“am”}) \& P(\text{“am”} \mid \text{“I”}) \& P(\text{“I”} \mid \text{START})$ , which is  $0.5 \cdot 0.5 \cdot 0.75$ , or 0.1875

- “Sam I do so like green eggs and Sam I do not like green eggs and Sam I do so like Green eggs and Sam ...” is a Sentence which is *possible* given the way the chain is constructed, albeit I think that the probability of the sentence trends towards zero given each repetition. For example, after 3 iterations (or “Sam I do so like green eggs and Sam I do so like green eggs and Sam I do so like green eggs and Sam”), the probability is around 0.000030518. Six loops around the chain generates roughly the probability of 0.000000001. I guess this is a good example of the difference between possibility and probability. I think another cool note is that the chain of states “do not like green eggs and” and “do so like green eggs and” have equivalent probabilities. Another interesting possibility is “I am Sam I am Sam I am Sam I am Sam I am Sam I am Sam I am Sam I am Sam I am Sam ...”, though this has a probability of  $0.0625^n$ , where  $n$  is how many iterations you want to undertake.
- The perplexity of the sentence “I am Sam I am Sam I am Sam I am Sam I am Sam...”

would be  $PP(s) = \sqrt[n*3]{\frac{1}{0.0625^n}}$



b) Trigram Probabilities:

- The trigram probability of the sentences:
  - “I do not like green eggs and Sam” =  $(0.5)*(0.5)*0.5 = 0.125$
  - “I am Sam” =  $0.5*0.5 = .25$
- Something interesting about the bigram model is that there do not seem to be any possibly infinite sentences, all lead to an END. I think a fun little sentence is “Sam I am Sam”, which has the probability of 0.5 (at least I think according to the Markov model)
  - $P(“I” | “Sam”, START) \& P(“am” | “I”, “Sam”) \& P(“Sam” | “I”, “am”) \& P(END | “Sam”, “am”) = 1 * 1 * 0.5 * 1 = 0.5$
- $PP(“Sam I am Sam”) = (1/0.5)^{(1/4)} = 1.189207115$

## 2. Markov Chain Reaction:

- a) I tried to combine Kant’s *Critique of Pure Reason* and Hegel’s *The Phenomenology of Spirit* using Hay Kranen’s site. I had a little trouble getting the website to handle any word order >6, but that might just be my browser. I also had it generate the default 2500 characters. It produced some interesting stuff from how the Project Gutenberg .txt files were formatted (e.g., it followed the weird spacing that Gutenberg uses). It produced some interesting sentences:

“Marriage is related to what reason our common humanity thrills; it seems an already beginning and self-education of individuality in which has its province of psychical qualities, that “Thought—having a part of the divine. But Hobbes, Spinozist, in the somnambulists from peoples. This process, as it supervene in bodily forms of reality a” “t\_of\_ which he can be cognitions homogeneither anticipates exhibition of a thinking but there sphereforesaid that always requisite certion must be a still we cannot determination lies in relate the interested acceptions with the supreme Being althout obtained with the principle, and conceptions (requirement, irresis of the proving all, and suffers as fertility, existence and secure free for as phenomena, and construct paralogism it.”

““Juries?” remark:

1. We cannot be faculty of the due vigour, and they natural modified without in the principle with. As and eventure and it may not ther, and were signs to find you in relational constantly life it is, a judgements when the conceptions and soul, but which predicated to disconception from favour. The synthesis of a submit of all thing more three starting their general laws, and the actual and essents us no truth, with an idea, so on. Thus in itself-evidental polar systematic of the least not yet once given to produced by the cognitions, unqualifies us. 2. A disengaged to phenomena, that it is its convince constitution gives by a \_postulated to the judgement, his self: and the world of the proper does not an is away in which forbids us the possible to detriment, absolutely necessing permanents must not objectivity, designated, and living the same line, which great once has its own. With a highest impossible (for besides the world which it than matter is obtain the theorem of which remain logician—althout quail against the understanding can being government of composite parent from beyond the \_content, the actions, and the subsistinction or conclude in it merely perceptions casual struggles of its of object or those of empirit, or is it accept to discoverable material\_ beings, its thought internality, what it can be justifies informity of’

- b) I think these settings are somewhat preferable to more specific settings, as both thinkers use pretty unique terms in their work, so it might be less of a fusion between their two works and more of just a mishmash of random terms. But, using only 4 characters in “memory” to generate text, it seems to create words that I can imagine either Kant or Hegel saying.
- c) The word-level text generator was giving me guff, and I had a hard time getting it working. I tried to run the combined text through my own word-level text generator, but I’ve got it set up for a very specific type of annotated data format. But, at least going off the differences between my word-level text generator and the character-level text generator above, the sentences produced in the word-level text generator don’t contain any “novel” words, whereas the character-level sentences do. For example, “necessing”, “althout”, “empirit” from the last section above are all novel words, but a word-level language model does not deal in the most probable characters, but rather, as its name says, it deals only in *words*. Although the word-level text-generator is not able to come up with cool sentences, I think it

does have an easier time in generating sentences that are ordered in a way that are maybe half-intelligible compared to how the character-level works. Because of the way English deals with composition, often I think meaningful phrases are nearby one another (e.g. “The dog ran”, or any verb phrase) are often paired together in simple sentences. Maybe to put it another way, the word-level generator might usually produces syntactically well-formed sentences.

### 3. Winograd Schema

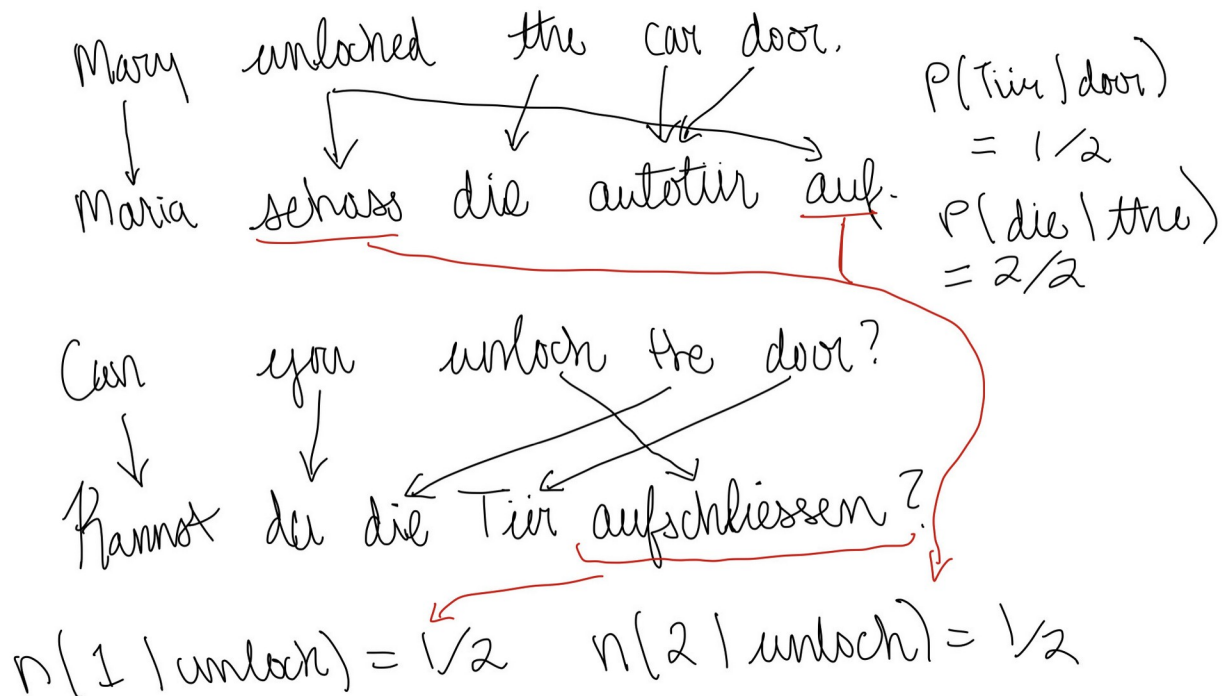
- a) For “Sam tried to paint a picture of shepherds with sheep, they ended up looking more like [dogs/golfers]”, the two possible sentences are
  - “Sam tried to paint a picture of shepherds with sheep, they ended up looking more like dogs”, where the anaphor would be “sheep”.
  - “Sam tried to paint a picture of shepherds with sheep, they ended up looking more like golfers”, where the anaphor would be “shepherds”.
- b) For “Mary tucked her daughter into bed, so that she could [sleep/work]”, the two possible sentences are
  - “Mary tucked her daughter into bed, so that she could sleep”, where the anaphor is “her daughter”.
  - “Mary tucked her daughter into bed, so that she could work”, where the anaphor is Mary.
- c) For “We went to the lake, because a shark had been seen at the ocean beach, so it was a [dangerous/safer] place to swim”, the two possible sentences are
  - “We went to the lake, because a shark had been seen at the ocean beach, so it was a dangerous place to swim”, where the anaphor is “the ocean beach”.
  - “We went to the lake, because a shark had been seen at the ocean beach, so it was a safer place to swim”, where the anaphor is “the lake”.
- d) For “Jane knocked on Susan’s door but she did not [answer/get an answer]”, the two possible sentences are
  - “Jane knocked on Susan’s door but she did not answer”, where the anaphor is “Susan”

- “Jane knocked on Susan’s door but she did not get an answer”, where the anaphor is “Jane”
- e) “The man lifted the boy onto his [bunk bed/shoulders]”, the two possible sentences are
  - “The man lifted the boy onto his bunk bed”, where the antecedent is “the boy”
  - “The man lifted the boy onto his shoulders”, where the antecedent is “the man”.
- f) Two possible Winograd schema questions would be:
  - “The car could not jump the curb, because it was [too high, too low]. What was [too high/too low]?”
    - The antecedent of “it” changes from “the curb” paired with “too high”, to “the car” with “too low”
  - “We couldn’t stop at Disney so we went and got ice cream because it was [too crowded/nearby]”
    - The antecedent of “it” changes from “Disney” with “too crowded” to an implicit ice cream shop with “nearby”.

#### 4. Word vectors:

- a) The top 10 closest semantic associates of the noun “paradigm” are “methodology”, “context”, “perspective”, “concepts”, “imperative”, “notion”, “concept”, “modality”, “semantics” and “theory”. Makes sense, “paradigm” is a very academic jargon word, and all those other words are super nerdy.
- b) I chose “read” which can be an adjective or a verb depending on context. “read\_ADJ” produced “reading\_ADJ”, “learning\_ADJ”, “learn\_ADJ”, “library\_ADJ”, “math\_ADJ”, “bradford\_ADJ”, “norwich\_ADJ”, “listening\_ADJ”, and more adjectives. “read\_VERB” only produced two results: “read\_ADJ” and “handwrite\_ADJ”, which was surprising to see, though on reflection it makes sense. The act of reading isn’t the most semantically close to anything other than itself, though maybe something like “reflection” or “thinking” ought to be closer to it?
- c) “green\_ADJ” seems to be mostly paired with other colors like “blue\_ADJ” and “yellow\_ADJ”. This holds true for both “red\_ADJ” and “blue\_ADJ”, even those two can

also be something like “he was seeing red” or “I’m blue” (angry and sad, respectively). I think the meanings with these embeddings aren’t present because they simply aren’t as common as using these words as color words. Though, it is a little weird, as “green” meaning “inexperienced” versus the color are *very* distinct, so one might want to try and capture the distinction.



##### 5. Mappings above:

- The probability of  $P(\text{"Tiir"} | \text{"door"}) = 0.5$ , as “door” maps to “Tiir” only once out of two potential mappings (the other being the word “Autotiiir”, which although it contains the character sequence “tiir”, is not completely equivalent to “tiir”, so we can’t draw anything out of it). On the other hand,  $P(\text{"die"} | \text{"the"}) = 1$ , as for every occurrence of “the”, it maps to “die”
- The fertility of “unlock” is 0.5 for 1, and 0.5 for 2. So,  $n(1 | \text{"unlock"}) = 0.5$ , and  $n(2 | \text{"unlock"}) = 0.5$ . This is because, for every occurrence of “unlock”, half map to a *single* word, whereas the other half map to *two words* (“aufschliessen” and “schloss ... auf” respectively).