

Intro to Sequence Models and Natural Language Processing

Includes some work by undergraduate researchers:

April Volzer, Noah Madrid, Hyechan Jun, Advait Scaria, Ha-Ram Koo,
Joseph Jinn

Across-the-row exercise

- Slide will give a cue, and which half-row will participate
- Each person says one word on their turn
- Start at the middle, go towards outside.
- I'll start us off with a trigram

Who: Front row by whiteboard

A

Who: Middle row by windows

For

Who: Back row by windows

Two

In the beginning was the **Word**,

and the Word was with God,

and the Word was God.

He was with God in the beginning.

Through him all things were made;

without him nothing was made that has been made.

In him was life,

and that life was the light of all mankind.

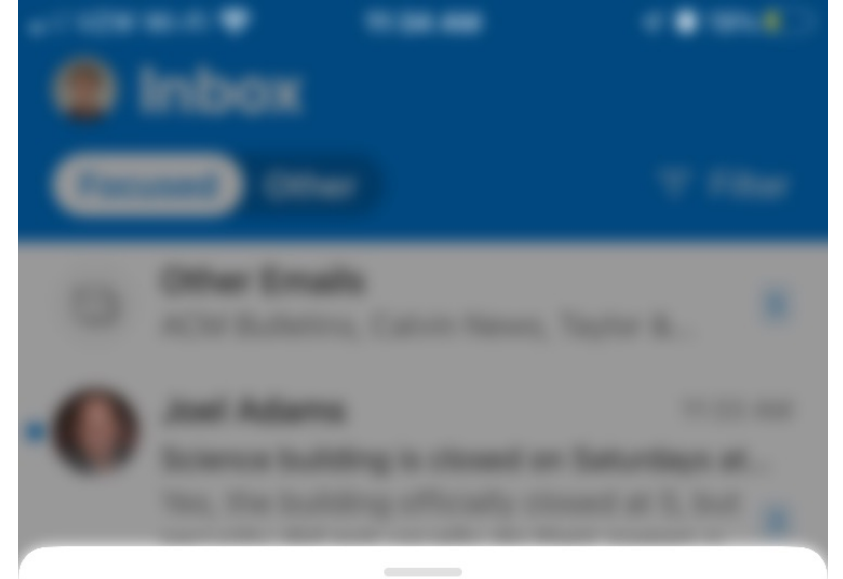
John 1 (NIV)

Logistics

- Check Moodle grades
 - Did you forget to *reply* in a discussion?
 - Did you forget to turn in a Reflection?
- Office Hours are here for you! E.g., **today 2pm**
- Don't forget:
 - Discussion 7: Wednesday
 - Homework 7: Thursday
 - Preparation 8: *postponed to Wednesday*
- Projects: more details posted, milestone coming up!
- Contributing to review materials

Today

- Predictive text shapes our writing (my dissertation and current work)
- Intro to Natural Language Processing



Text Predictions

Swipe right to autocomplete
your sentences

Subject:

Whats the best time to)meet?

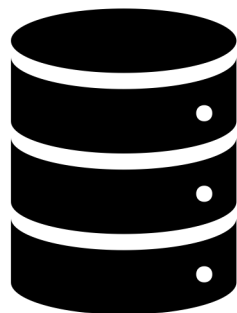


Got It

Training Data




Language Model



Next-Word
Probabilities

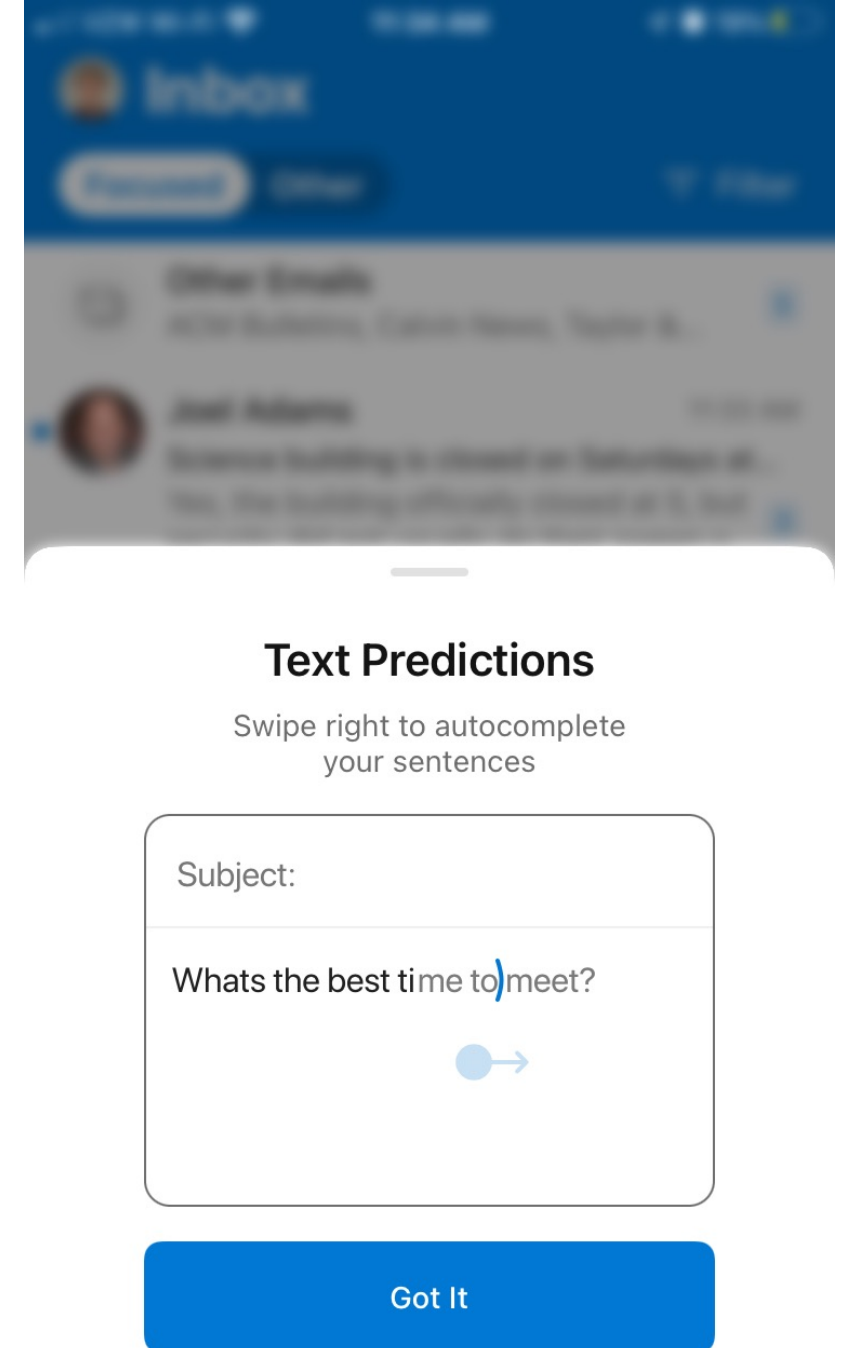
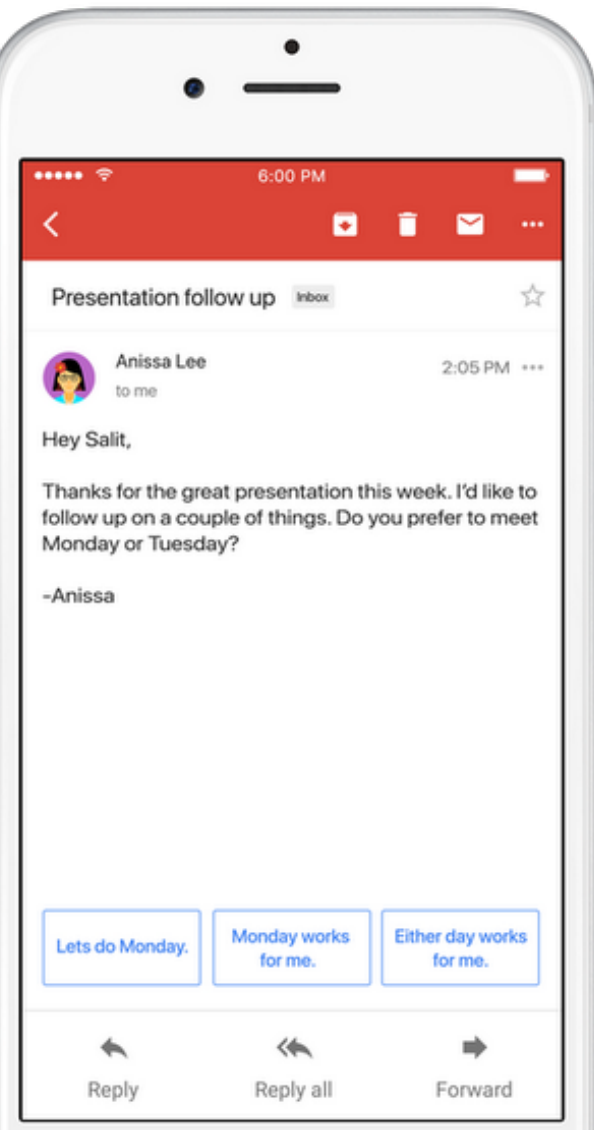
Suggestions

> I don't think you have any 

Writing in progress



Predictive Text Suggestions



How do predictive suggestions shape content?

The **text that people write**
“I have a phrase in mind and use
using predictive text entry systems
suggestions that match.”
reflects **biases** of these systems.

Bias *noun*. disproportionate weight in favor of or against an idea or thing

How might predictions shape us?

- Writing choices are made in-the-moment
- Suggestions are offered after *each character typed* — several per second

Prior studies *couldn't* measure this

- Transcription tasks: “Type this”
- Speed and accuracy measures

Word Suggestions Discourage the
Unexpected

Predictive Text Suggests what's Most Expected

This weekend I plan to	take go attend	I want to go for a	walk run swim	My kids are so	excited happy much
Bias against these words	visit		ride		good
	be		long		young
	do		few		scared
	make		stroll		smart
	spend		little		small
	travel		hike		different
	play		quick		proud
	write		drive		busy
	start		drink		cute
	meet		break		lucky
	have		spin		used
	return		shot		tired
	bring		beer		close
	get		bit		very
	head		jog		beautiful
	run		second		special

Suggestions favor certain kinds of words at the expense of others: a **bias by design** against the unexpected

Does writing with suggestions result in more of those kinds of words?

Study 1: Does writing *conform* to suggested words?

Vary suggestion **visibility**

- Suggestions **never** visible
- Suggestions **always** visible
- Suggestions visible when system is **confident**
 - Tuned so that full-word suggestions are shown for ~50% of words

Hypothesis: Longer, more predictable words when suggestions visible

Image Captioning Task

- **Open-ended:** suggestions can affect content (or not)
- **Repeatable:** one participant can write many times

Write the most specific and accurate description you can for the image below. After you're done, tap here:

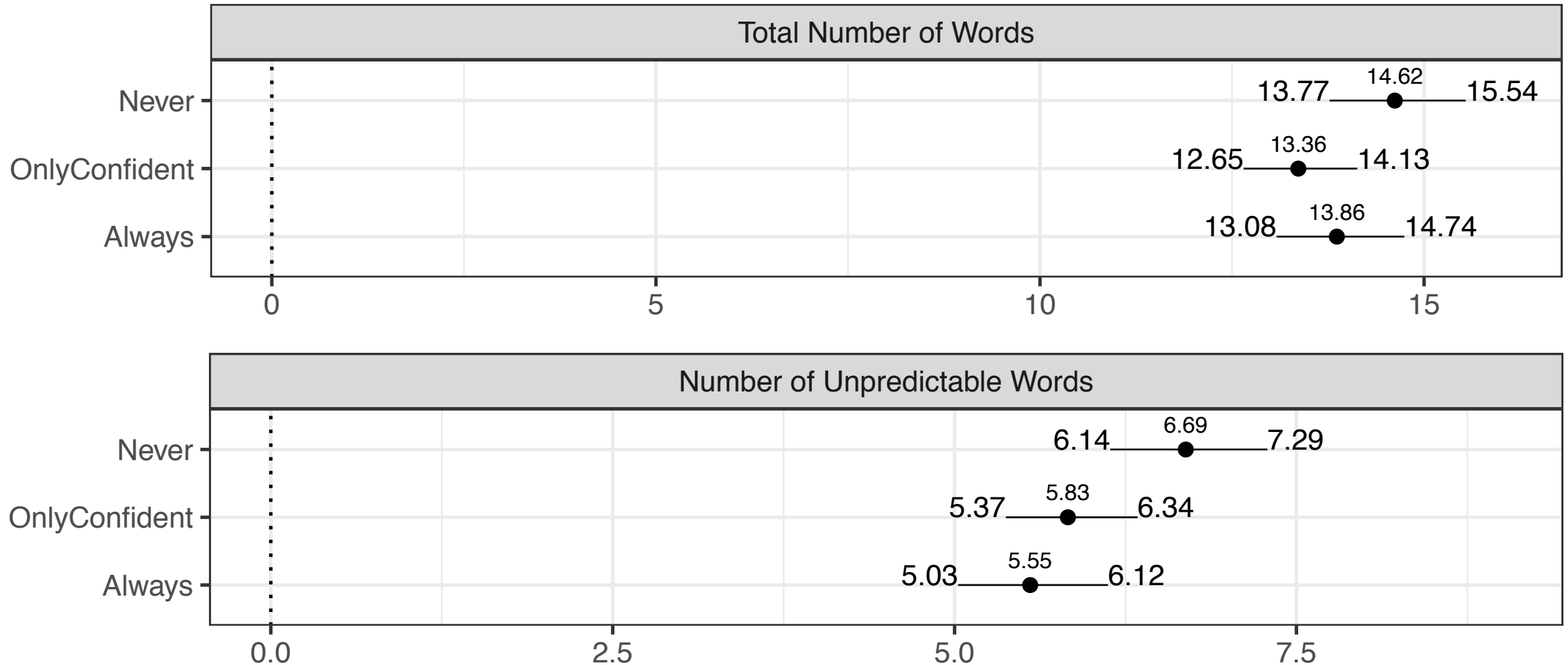
Next



a

woman			man			person			
q	w	e	r	t	y	u	i	o	p
	a	s	d	f	g	h	j	k	l
'	?	z	x	c	v	b	n	m	⌫
-	!	,	space				.	return	

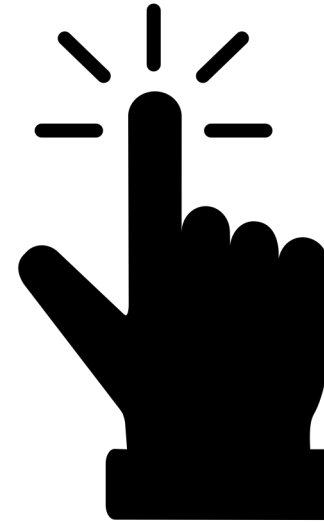
Overall: Shorter Captions, Fewer Unexpected



Suggestions may discourage thoughtfulness



Thinking about
what to write



Accepting a
suggestion

Study 2: Do positively-biased suggestions bias writing?

- N=38 participants, each named 4 restaurants (both good and bad experiences)
- Wrote reviews; system suggested sentiment-manipulated phrases

- **Positive slant**

this		
is	place	was
by far my favorite	is one of my favorite	my first time at

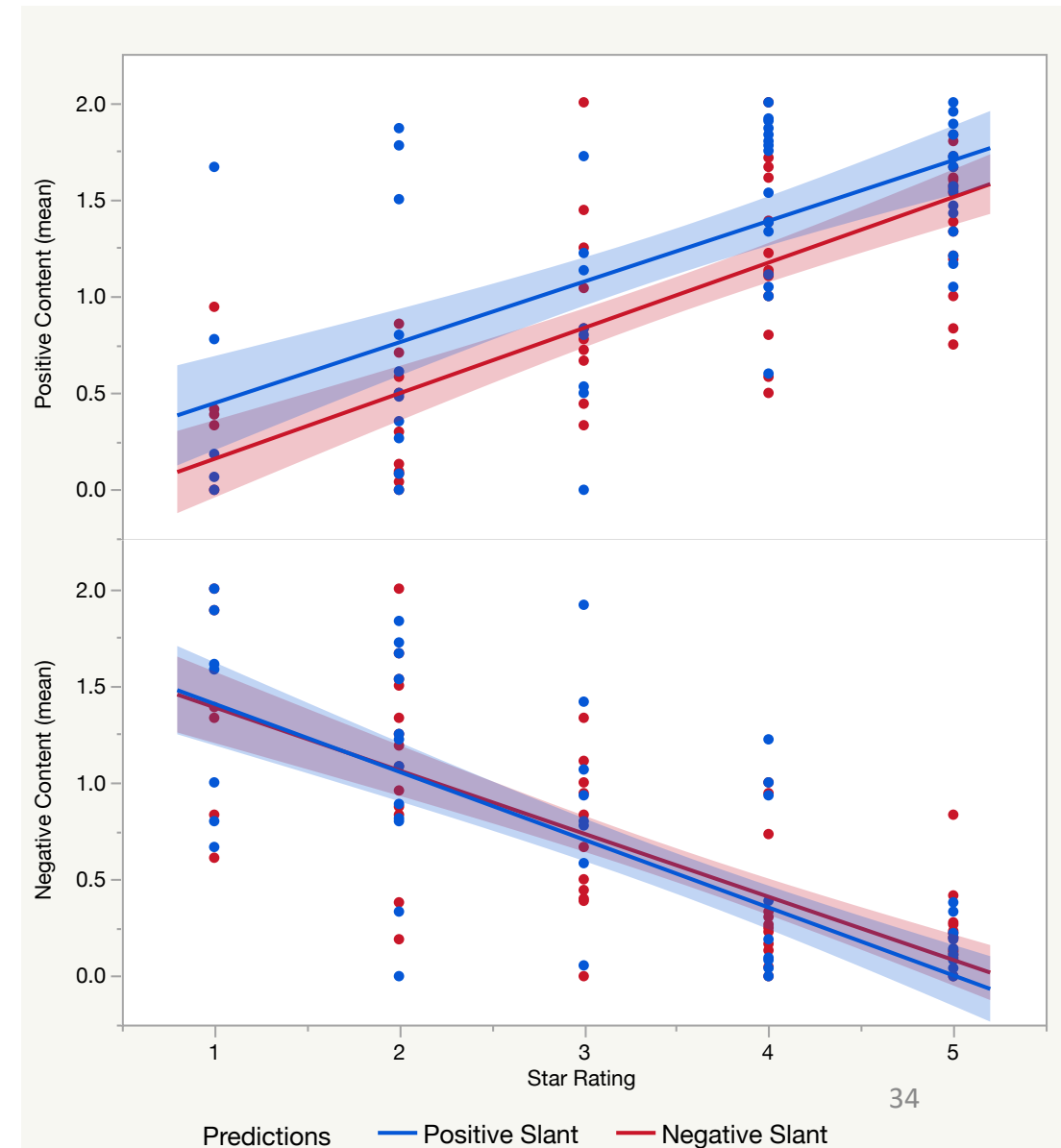
- **Negative slant**

this		
used	is	review
to be one of the	by far the worst	is for the restaurant

- **Measure:** For each sentence in each review, 3 annotators (MTurk) rate positive content, negative content

Results: Positive-Slanted Suggestions → Positive-Slanted Reviews

- Effect size: about 1/3 star rating
- No measurable effect on negative content
- **Subjective preference towards positive slant system**
 - Negative slant viewed as less relevant



Excluding 3 participants who wrote without tapping keys at all

Predictive text biases writers.

- Using predictive text leads to fewer unpredictable words IUI 2020
- Phrases have stronger impact than words UIST 2016
- Biases extend to meaning, not just low-level details GI 2018

Influence may be subconscious.

This is all by design.

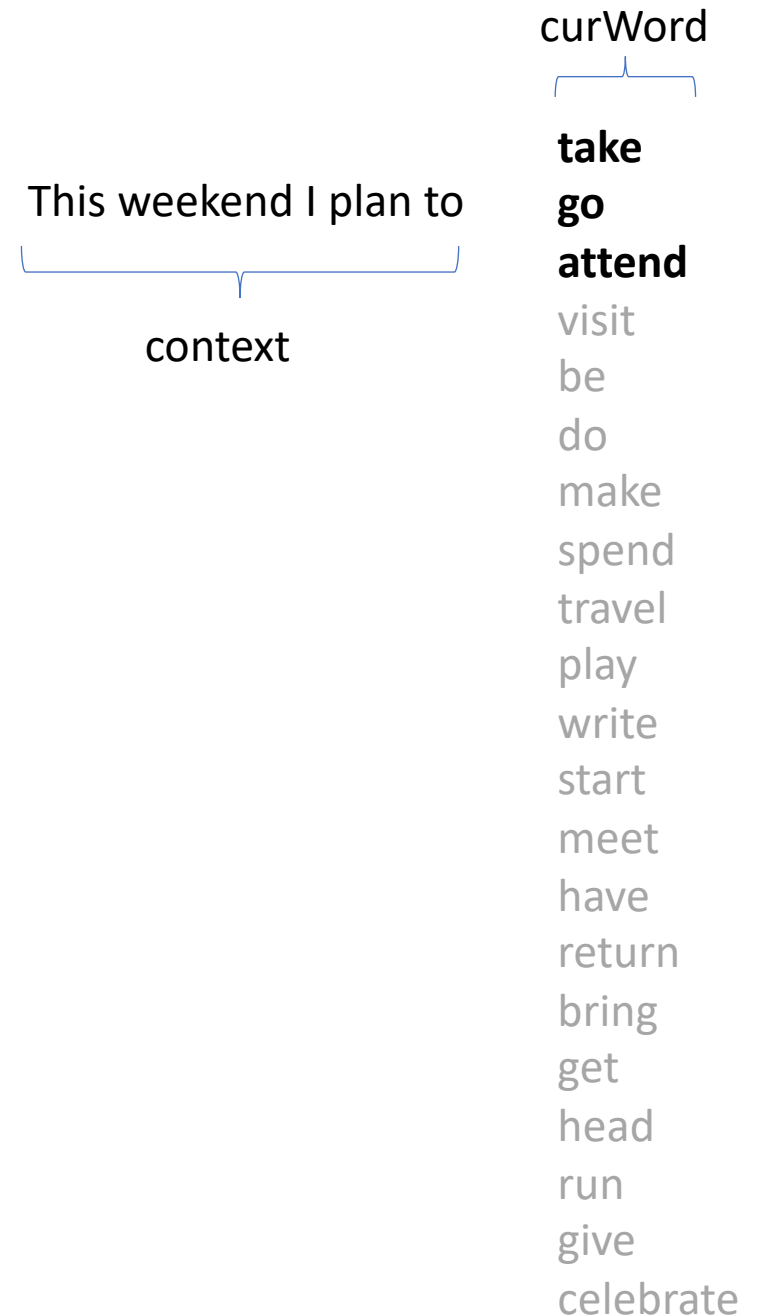
What does the system do?

1. **Compute** probability of *next word*:

$$P(\text{curWord} \mid \text{context})$$

2. **Suggest** *most likely* word:

$$\text{argmax}_{\text{curWord}} P(\text{curWord} \mid \text{context})$$



Predict-the-next-word as *classification*

- **Technical name:** *causal language modeling*
 - Don't get to look at "future" words or backtrack
- **Task:** given what came *so far*, predict *the next thing*
- Next **character**: # possibilities (classes) = _____
- Next **word**: # possibilities = _____
- What else could we use as "next thing"?

What those computations look like

- $P(\text{word} \mid \text{context}) = \text{softmax}(\text{wordLogits}(\text{word}, \text{context}))$
- $\text{wordLogits}(\text{word}, \text{context}) = \text{dot}(\text{vec}_{\text{word}}, \text{vec}_{\text{context}})$
- vec_{word} : look up in a (learnable) table: **embedding**
- $\text{vec}_{\text{context}}$: computed by a neural network

Do you recognize this structure?

This week: Natural Language Processing

- What **tasks** might we want to do with natural language?
- How do we turn words into numbers and back?
- How might we use neural nets for these tasks?

One architecture for everything?

- **Transformers** (Vaswani+ 2007) have transformed language processing
 - ... and most other machine learning also
- Radical *convergence*: a single model architecture for all kinds of data and tasks. Just *make everything look like a sequence*
 - **Language**: sequence of words
 - **Vision**: sequence of image patches
 - **Speech**: sequence of audio segments
 - **Behavior**: sequence of actions and consequences
 - ...!

So we'll switch materials

- **Fastai**: strong on vision, but weak on Transformers
- So: switching to Hugging Face Transformers
 - Also built on PyTorch
 - Also built by some of the same people
 - Much better documented (phew!)
 - Course is starting to look good!
 - <https://huggingface.co/course>



Your job: do the HuggingFace course.

- We'll work through exercises together
- We'll fill in conceptual gaps together

Deep Neural Networks

- Composition of 3 kinds of functions:
 - *Similarity* (dot product) with learnable *prototypes* (matrices)
 - *Choice* (softmax)
 - *Bend* (rectifier or other nonlinearity)
- Training by gradient descent
 - How to adjust each prototype to boost the score of the right answer?
 - Each word can help tune every prototype that was involved
- All functions are differentiable -> use *chain rule* to differentiate

Transformer Architecture

- Attention mechanism: programmable information flow
 - Same softmax mechanism determines which other word to pay attention to
- Also learned by gradient descent: boost the score of the word that would give the most helpful information

Self-Attention networks are powering an AI explosion

- AlphaFold: predicting folded structure of proteins
- Translation, summarization
- CLIP / DALL-E: image generation
- ViT – using these architectures for image recognition

It's By Design

- Models trained to mimic writers (huge computational effort!)
- Interaction designed to suggest **most likely next word** (least surprising, least informative)
- ... like someone who's eerily good at finishing your sentence.

Creation: Genesis 1

In the beginning God created the heavens and the earth. Now the earth was formless and empty, darkness was over the surface of the deep, and the Spirit of God was hovering over the waters.
And God said, “Let there be light,” and there was light. God saw that the light was good, and he separated the light from the darkness.
And God said, “Let there be a vault between the waters to separate water from water.” So God made the vault and separated the waters above the vault from the waters below.
And it was so. God called the vault “sky.” And there was evening, and there was morning—the second day.
And God said, “Let the water under the sky be gathered to one place, and let dry ground appear.” And it was so. God called the dry ground “land,” and the gathered waters he called “seas.”
Then God said, “Let the land produce vegetation: seed-bearing plants and trees on the land that bear fruit with seed in it, according to their kinds.” And it was so.
And it was so. The land produced vegetation: plants bearing seed according to their kinds and trees bearing fruit with seed in it, each according to its kind.
And God said, “Let there be lights in the vault of the sky to separate the day from the night, and let them serve as signs to mark seasons and days and years.” And it was so.
And it was so. God made two great lights—the greater light to govern the day and the lesser light to govern the night. He also made the stars.
And God said, “Let the water teem with living creatures, and let birds fly above the earth across the vault of the sky.” So God created the large sea creatures and every living creature that moves and swims in the water, and every bird of every kind.
And God said, “Let the land produce living creatures according to their kinds: the livestock, the creatures that move along the ground, and wild animals.”

Then God said, “Let us make mankind in our image, in our likeness, so that they may rule over the fish in the sea and the birds in the sky, over the livestock and all the wild animals, and over all the creatures that move along the ground.
So God created mankind in his own image,
in the image of God he created them;
male and female he created them.

Sin and Predictive Text

- Automated lazy communication (“Thank you so much!”)
- Self-centeredness (what’s easy for *me* to say?) vs thinking of *reader*
- Dehumanization
 - Success = lots of suggestions accepted?

Jesus

- *Speaks* grace and truth
 - Will judge our selfishness... and our dehumanizing technology
 - Gives the Spirit: wisdom, humility
 - Will redeem *us* (and has already started)
-
- ...and will redeem our *technology*
 - ...and has already started.

“Let Your Kingdom Come”

- Dreams of the renewed Creation
- “Glory of the nations” (Revelation 21) includes our *technology*
- (Re)shaping technology *now*

“Swords into plowshares” (Isaiah 2:3-5)

- Repurpose existing technology for better use
- Leverage pretrained language models
 - initially trained for mimicry
 - repurposed for empowerment

Two student projects

- Sentence reorganization (April Volzer, Noah Madrid, Joseph Jinn)
- AI Interviewer (Hyechan Jun, Ha-Ram Koo, and Advait Scaria)

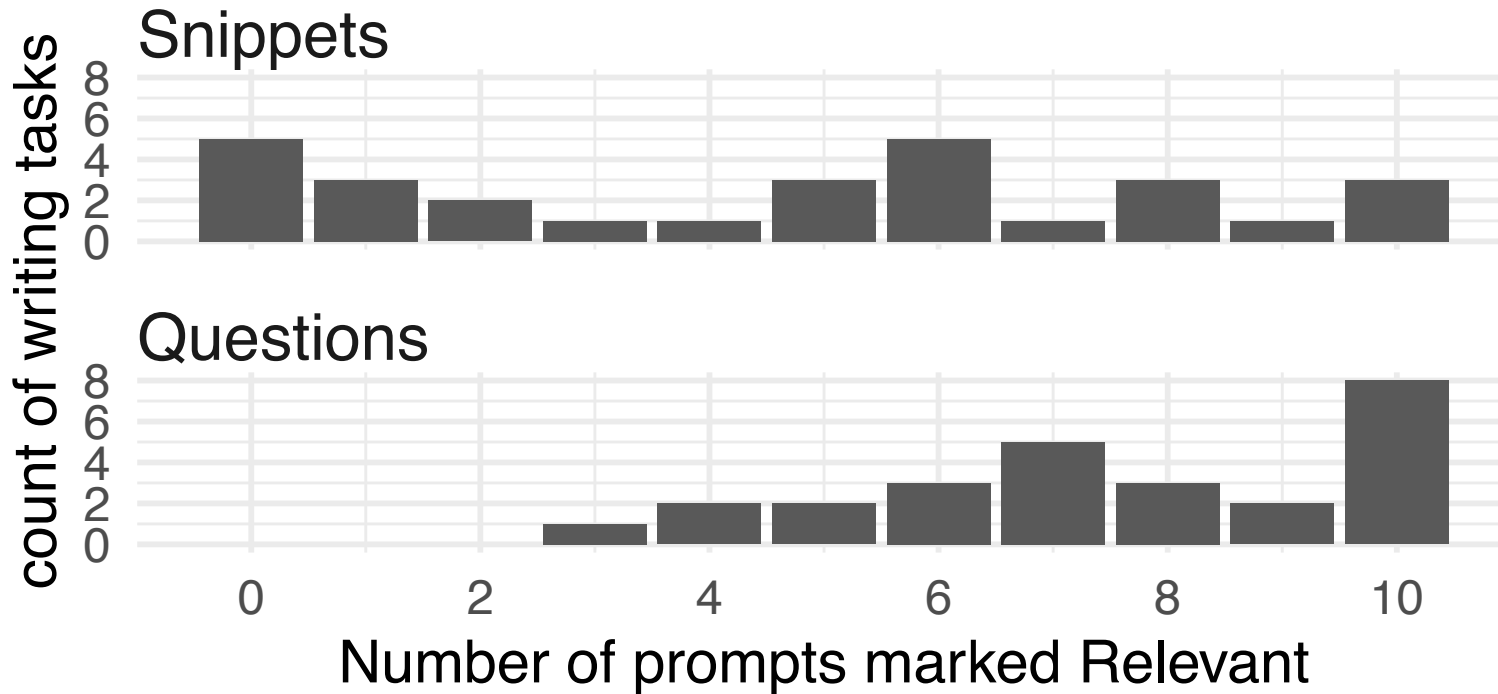
Can Questions help writers with Drafting?

Suppose you're writing a Wikipedia article about a sci-fi drama.

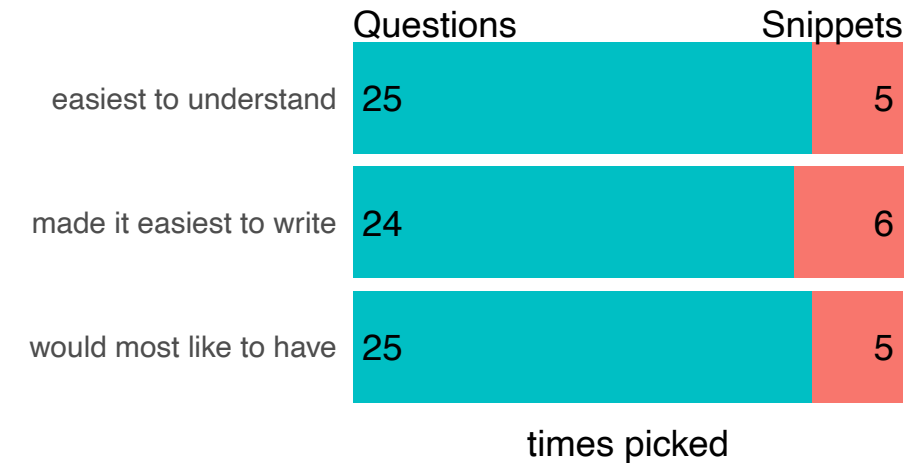
- **inspirational Example:** "Blade Runner" initially underperformed in North American theaters and polarized critics
- **vs abstracted Question:** How did it initially perform? How did critics react? What aspects did critics praise?
- Research question: do writers find *questions* as useful as *examples*?
- **Experiment:** sentence-level informational writing
- **Study Design:** 30 MTurk participants wrote 10 sentences for each of 3 different prompt **presentations**
 - Snippets (of Featured Articles)
 - Questions (corresponding to those snippets)
 - No prompts

Writers strongly preferred Questions

While writing, more of the Questions were relevant.



After writing, writers subjectively favored Questions



Could writing feel more like conversation?



“Interesting, tell me more!”



“What were they doing before that happened?”



“Have you thought about this perspective?”

Question generation can be
more **forgiving** of error than next-word prediction
natural, like getting interviewed

Question-**asking** is a ripe opportunity for intelligent systems

Exploring alternative wordings is difficult

- **Cognitive:** fixation on what's already written
- **Mechanical:** non-local edits
 - **tedious:** cut/paste, retype, cursor navigation within a sentence
 - **error-prone:** forgetting updates
- Goal: easy to **explore alternative wordings** of a sentence
- Co-creative interaction: once writer selects a sentence...
 1. system recommends alternative wordings
 2. writer refines by semantic editing operations

System recommends high-quality diverse paraphrases

Yellowstone National Park was established by the US government in 1972 as the world's first legislated effort at nature conservation.

- + Yellowstone National Park was established by the U.S. government in 1972 as the world's first legislative effort for nature conservation.
- + In 1972, the Yellowstone National Park was established by the U.S. government as the world's first legislative effort for nature conservation.
- + The US government established Yellowstone National Park in 1972 as the world's first legislative effort for nature conservation.
- + By the US government, the Yellowstone National Park was established in 1972 as the world's first legislative effort for the conservation of nature.
- + The world's first legislated effort for the conservation of nature was established by the U.S. government in 1972 in Yellowstone National Park. +

Writer refines by semantic edits

e.g., drag and drop phrases

It was earlier demonstrated by researchers that heart attacks can be caused by high stress.

select It was earlier demonstrated by researchers that heart attacks can be caused by high stress.

drag earlier demonstrated researchers high stress heart attacks caused

drop It was earlier demonstrated by researchers that high stress can cause heart attacks.

Approach: Paraphrase by round-trip translation

- To **recommend reorderings**, we vary sentence openings
 - Identify **candidate sentence-openers** using heuristics on dependency tree
 - **Beam-search for completions** of each candidate opener
 - **Filter bad paraphrases** (e.g., content words dropped)
 - **Re-rank** by sentence likelihood
- Interpret reordering **edits** as partial ordering on words
 - use constrained beam search (Hu et al. 2019) (work in progress)
- **Other editing operations**: specifying what aspect to emphasize, replacing phrases with synonyms, etc. (work in progress)

AI Interviewer

- A skilled interviewer brings out the best in their guests
- Could an AI likewise inspire writers?

CS Senior Project (Hyechan Jun, Ha-Ram Koo, and Advait Scaria)

- **Dataset:** NPR interview transcripts
- **Task:** Mimic NPR interviewers
- **Approach:** “fine-tune” language models to generate host’s utterance given previous guest utterance
 - Two base models: GPT-3 (OpenAI / Microsoft), BART (Facebook)

Guest

It's not at all unexpected. If anything, investors and some people in the auto industry had been begging **Ford to make this step**. General Motors Corp. earlier this year did something quite similar. [...]

AI Response

Well, why did **Ford take this step** now?

NPR Interviewer

The **UAW** traditionally has been one of the biggest, most powerful labor unions in America. Is the union, because of **Ford's** woes, being forced to take whatever the company offers?

Sometimes it works great...

Welcome to the BART Interactive Interview AI!

Tell us an interesting story, or type 'END' to stop.

> I have a cat. His name is Jonathan. He is seven years old and has black fur.

AI: And how did you decide to adopt Jonathan?

And sometimes it, well...

Welcome to the GPT-3 Interactive Interview AI!

Tell us an interesting story, or type 'END' to stop.

> Hello. My name is Hyechan. I'm a senior at Calvin University studying computer science.

AI: And how do you translate these skills in things like software testing, which you've mentioned, to the business world? Do you see your career path in either one?