

Course logistics:

↳ Piazza for discussion

↳ Gradescope for submission

↳ course website

↳ cs.umd.edu/~umiyer/cmsc723

↳ course email

↳ cmsc723instructors@gmail.com

TAs:

↳ Maharshi } office hours
↳ Yapei / Yekyung } TBD

Mohit : Fridays 9-10am

↳ office hours start next week

Course materials

- ↳ no textbooks
- ↳ readings are PDFs / blogposts linked on the course page

Things to expect:

- ↳ basic machine learning
- ↳ programming (Python)
- ↳ math: linear algebra
probability
calculus

Grading:

- ↳ 25% exam 1 ^{→ 10/19}
 - ↳ 25% exam 2 ^{→ 12/19}
 - ↳ practice exams: Resources tab in Piazza
 - ↳ 20% problem sets
 - ↳ 30% final project
- } in-class multiple choice

↳ 10% proposal

↳ 20th report (12/18)

↳ groups of 4-5 people

↳ any NLP related topic

AI policy

↳ AI allowed on HW

↳ AI not allowed on exams

↳ AI partially allowed on project reports

↳ can use for coding, brainstorming, analysis, writing feedback on drafts

↳ cannot use AI for writing submitted text

↳ this policy is enforced

↳ submissions must obtain a "100% human-written" score on Pangram

↳ autograder script allow for
a few submissions, tell you score
each time

upcoming due dates

↳ Hw0 : 9/16

↳ form project groups : 9/14

↳ project proposals : 10/5

A Small oblong

A ✓ c h ✓ ✓ ✓ ✓ s ✓ s ✓ ✓ ✓

reading lamp

✓ c ✓ c c ✓ ✓ ✓ ✓ b o ✓ ✓

on the desk

✓ i ✓ ✓ ✓ ✓ ✓ ✓ + ✓ ✓ ✓ ✓

shed glow on

i ✓ i l ✓ i ✓ ✓ ✓ s b ✓ ✓

polished wood

+ i r e c ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓

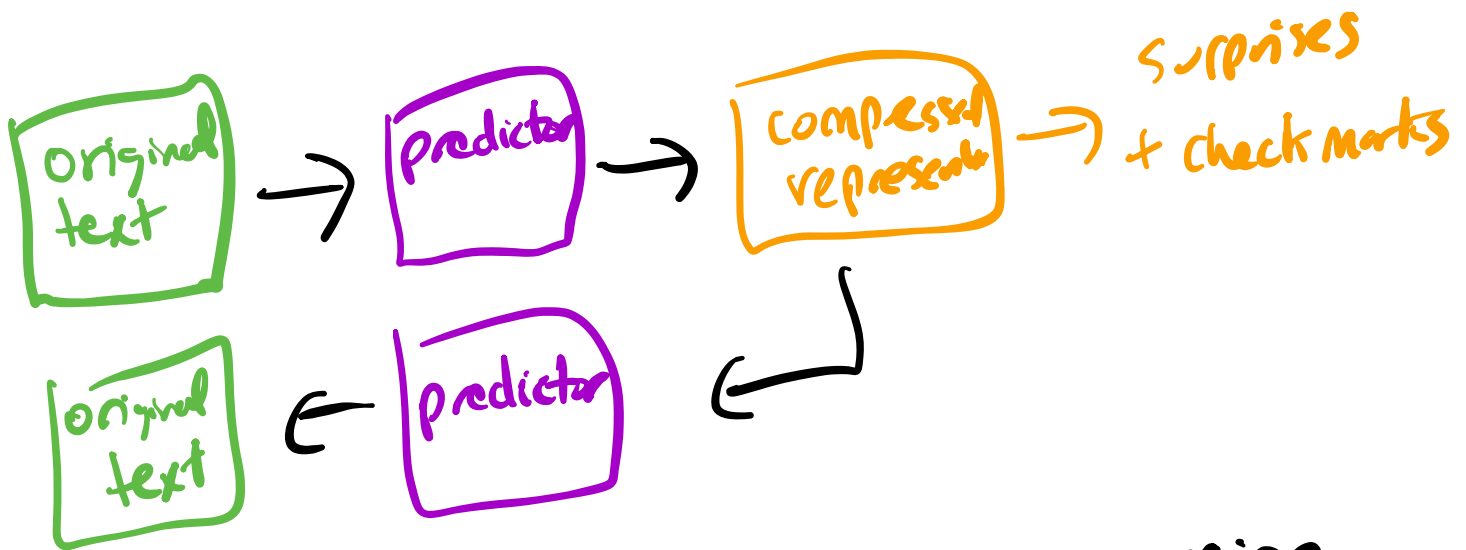
↳ Claude Shannon, 1951

↳ 70% correct

↳ random seq : $1/27$, $\sim 4\%$
acc

↳ conclusion : English is predictable

↳ anyone who predicts exactly like
you can recover the original sentence
& just from the mistakes + checkmarks
↳ "surprises"



intuition : prediction $\Rightarrow$ compression

↳ better predictor $\Rightarrow$ fewer surprises

↳ what if we never compressed?

S1: Colorless green ideas sleep furiously.

S2: Furiously sleep ideas green colorless.

Chomsky (1957): Under ^{text} predictors of that era,

$$P(S1) = P(S2) = 0$$

"probability of a sentence is useless..."

1957 - 1988: "symbolic era"

↳ building large databases of rules, facts, grammars

↳ did not generalize

1988 - 2000s: statistical NLP

↳ IBM machine translation

English $\rightarrow$ French translation

$e \rightarrow \left\{ \begin{smallmatrix} \text{noisy} \\ \text{channel} \end{smallmatrix} \right\} \rightarrow f \Rightarrow \text{Shannon}$

$$e^* = \underset{e}{\operatorname{argmax}} P(f|e) P(e)$$

$\hookrightarrow$ translation model, dictionary
 $\hookrightarrow$ language model
 $\downarrow$
only need English data

estimate from parallel data } big data

$\hookrightarrow$ no more large sets of rules

$\hookrightarrow$ shift to "feature engineering"

$\hookrightarrow$ hard to generalize!

$\hookrightarrow$ statistical NLP $\Rightarrow$ neural NLP

$\hookrightarrow$ neural language models

$\hookrightarrow$ 2003: Bengio "a neural probabilistic LM"