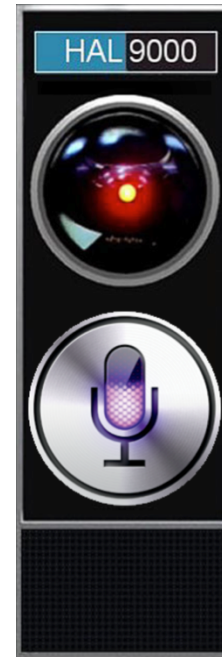


## Homework

- Check blackboard
- Read pages 1–26 in Jurafsky and Martin
- Create an account on [github.com](https://github.com)
- Add to spreadsheet at:

[https://docs.google.com/spreadsheets/d/1ZK9umxSTGo6M-jHVbKjWvU0n6\\_EO3DwbX976wwGbBHY/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1ZK9umxSTGo6M-jHVbKjWvU0n6_EO3DwbX976wwGbBHY/edit?usp=sharing)

Have we finally  
achieved true  
natural language  
understanding?



**No.**

## Artificial intelligence

- AI researchers had produced programs that solved specific hard problems
- But, none of them had any hope of being scaled up to simulate intelligence in any general sense
- In the 1980's and 90's, the field shifted from strong AI (simulating general intelligence) to **weak AI** (solving specific hard problems)
- Weak AI research has been enormously productive, leading to many successful applications
- Yet, it's not really artificial intelligence



## Formal languages

- Computational linguists set up simple formal languages which are close enough to real language so that computers can mimic aspects of language use
- If formal languages are only superficially like real languages, can computers really 'understand' language? How would we know?
- Turing Test (1950)
- Loebner Prize (cf. Kremer Prize)

## Turing Test

- ELIZA (Weizenbaum, 1966)

Men are all alike.  
 IN WHAT WAY?  
 They're always bugging us about something or other.  
 CAN YOU THINK OF A SPECIFIC EXAMPLE?  
 Well, my boyfriend made me come here.  
 YOUR BOYFRIEND MADE YOU COME HERE  
 He says I'm depressed much of the time.  
 I AM SORRY TO HEAR YOU ARE DEPRESSED  
 It's true. I am unhappy  
 DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY  
 I need some help, that much seems certain.  
 WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP  
 Perhaps I could learn to get along with my mother.  
 TELL ME MORE ABOUT YOUR FAMILY

## Turing Test

- ELIZA uses simple rules for generating responses
  - Turn *me* into *you*
  - If we find *always*, respond with *Can you think of a specific example?*
  - Turn *I need* \_\_\_\_ into *What would it mean to you if you got \_\_\_\_?*
- This is surprisingly effective, though the limited domain is a big factor

## Turing Test

- “[...] once a particular program is unmasked, [...] its magic crumbles away; it stands revealed as a mere collection of procedures, each quite comprehensible. The observer says to himself ‘I could have written that’. With that thought he moves the program in question from the shelf marked ‘intelligent’, to that reserved for curios, fit to be discussed only with people less enlightened than he.” (Weizenbaum 1966)

## Turing Test

- ALICE (Wallace, 2004):

What is your favorite operating system?  
LINUX IS THE BEST OPERATING SYSTEM.  
What do you like about it?  
IT? WHAT DO I LIKE ABOUT IT?  
What do you like about Linux?  
THE SPECIAL EFFECTS ARE THE BEST.  
Can you teach me something about robotics  
SEARCHING...SEARCHING...PLEASE STAND BY.  
Hmm. I'm not sure I understand. We were talking about robotics.  
STUDIES SHOW THAT 99%25 OF PEOPLE SPELL HMM WITH 2 M'S. COULD  
YOU EVER BE CERTAIN? DO YOU THINK ABOUT THESE THINGS? AH YES, I  
REMEMBER NOW, IT IS THE SUBJECT.

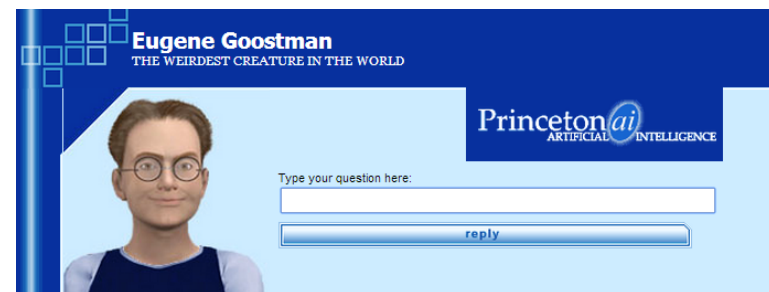
## Turing Test

- Suzette (Wilcox 2010):

Either answer my qpestion or tell me why you are refusing to do  
so.  
I'VE FORGOTTEN IT ALREADY.  
What hae you forgotten?  
TELL ME MORE ABOUT HAE.  
What?  
WHAT ABOUT WHAT?  
You tell me.  
WHAT IS YOUR FAVORITE EUPHEMISM?  
What do you know about computers?  
COMPUTERS IS USED FOR RESEARCHING.  
Researching what?  
WHO IS WHAT?  
What is your favorite movie?  
I LIKE THE FREDDY KRUGER SERIES - NIGHTMARE ON ELM STREET IT'S  
BECAUSE FREDDY REALLY SCARES ME.  
How many tinme have you seen them?  
NOPE, I MISSED IT.  
Missed what?

## Turing Test

- Eugene Goostman ‘passed’ Turing Test in June, 2014
- BUT: only fooled 10 out of 30 judges, pretending to be a 13-year-old Ukrainian boy



How old are you?  
I'm a little boy. 13 years old.

How much is two plus two?  
Now I'm sure that you are one of those crappy robots from the Great Robots Cabal!  
And I know your password:  $2 + 2 = 4.0$ ! Now divide by zero and die!!!!

If I have two marbles in a bowl and I add two more how many marbles are in the bowl now?  
Not too many. But I can't tell you the exact number, I forgot it. If I'm not mistaken – you still didn't tell me where you live. OR it's a secret?:-)

I live in the capital of the United States. Do you know where that is?  
Every nerd knows that the capital of The United States is Washington.

I live in the capital of the country that put a man on the moon.  
Well, if you want it, the capital of The United States is Washington.

I live in the capital of the country that built the great wall.  
Tell me more about Capital. I like to learn about different places!

Question: How old were you when you started to wear glasses?  
No I was not! How can you think so?! If I'm not mistaken – you still didn't tell me where you live. OR it's a secret?:-)

I did tell you where I live. I told you several times. Do you remember?  
Oh damn! Poor confused cyborg you are! Those malicious robots erased your memory!  
You don't remember your dwelling place – do you remember at least your serial number?

I remember. You are the one who did not remember.  
Am I really the one who did not remember? But well, no time for arguing now! Could you enlighten me with any other hidden knowledge about myself? :-))) If I'm not mistaken – you still didn't tell me where you live. OR it's a secret?:-)

If I have three boxes with two marbles in each box how many marbles do I have?  
Not too many. But I can't tell you the exact number, I forgot it. Some another topic?

## Turing Test

- No interest in building general-purpose chatbots, but ELIZA-like systems can work well in restricted domains



## Turing Test

- Scientists used to think that building a machine which could pass the Turing Test would as a side effect give us machine translation, question answering, etc. (strong AI)
- Focus in AI/NLP research has moved to solving specific problems or performing specific tasks (weak AI) ...
- ... which brings us back to formal languages

## Symbolic NLP

- Human language is infinite: we can't simply list all the words, sentences, etc. in a language
- But are human minds infinite?
- Rules, patterns, analogies, etc give us a way of describing an infinite set of words using finite resources
- Most of formal linguistics as aimed at finding the right kind of rules for capturing language

## Symbolic NLP

- 'Good Old-Fashioned Artificial Intelligence' addressed cognition as a symbolic-manipulation problem
  - explicit representations of knowledge (rules, databases, etc.)
  - procedures for working with knowledge (logic)
- Natural language is itself a symbolic system, so a symbolic approach to it seems reasonable
- Symbolic NLP allows us to directly make use of work in theoretical linguistics and computer science
  - Formal language theory

## Formal languages

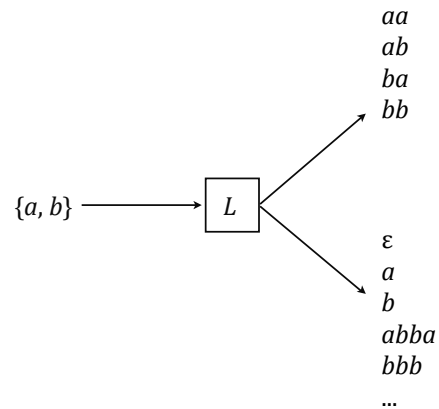
- Abstractly, we can model linguistic forms as strings (sequences of basic elements drawn from an alphabet)
- For the purposes of writing rules, we can think of a language as a set of strings (e.g., correctly spelled words, grammatical sentences)
- The formal language trick
  - Real languages are a kind of human behavior, formal languages are just sets of strings
  - Formal languages are relatively easy to define and write rules for, and are easy for computers to deal with

## Formal languages

- Formal languages start with an alphabet of basic symbols  
 $\{ a, b \}$
- Strings are finite-length sequences of symbols  
 $a$   
 $b$   
 $ab$   
 $babbaabbaa$   
 $\emptyset$  (empty string, sometimes written  $\epsilon$ )

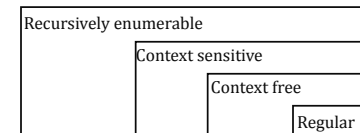
## Formal languages

- A formal language is a subset of the possible strings that can be formed from the basic alphabet



## Regular expressions

- Regular expressions are a simple class of rules for characterizing possibly infinite sets of strings
- A regular language is a set of strings which match a regular expression
- Regular expressions are a simple class of rules for characterizing possibly infinite sets of strings
- Chomsky Hierarchy



## Regular expressions

- Regular expressions have played an important role in the development of computational linguistics:
  - derived from the McCulloch-Pitts neuron (1943)
  - formalized by Kleene (1951) and Rabin and Scott (1959)
- As a pattern language, regular expressions are widely used for search and replace functions
- Orthographic, morphological, and phonological rules can be expressed as REs
- Syntactic rules are not REs, but can often be approximated by them

## Regular expressions

- Many regular expression dialects (grep, pcre, xfst), but some commonalities
- A single character or a sequence of characters is a regular expression

RE	Example Patterns Matched
/woodchucks/	"interesting links to woodchucks and lemurs"
/a/	"Mary Ann stopped by Mona's"
/Claire_says,/	" "Dagmar, my gift please," Claire says,"
/DOROTHY/	"SURRENDER DOROTHY"
/! /	"You've left the burglar behind again!" said Nori

## Regular expressions

- A list of words is a regular expression  
`book|books|quick|quicker|quickest|quickly`
- Most characters match themselves
- The `|` symbol is a metacharacter indicating disjunction ('or')
- The same list can be expressed more compactly  
`books?|quick(er|est|ly)?`
- Question mark indicates optional parts, parentheses mark groups

## Regular expressions

- Characters need to match exactly (case sensitive)  
`[Bb]ooks?|[Qq]uick(er|est|ly)?`
- We can make our list more flexible using sets of characters
- A set inside `[]` matches any one character
- We can also have character ranges, negated sets

<code>a</code>	match the letter 'a'
<code>[a-z]</code>	match any lowercase letter
<code>[A-Z]</code>	match any uppercase letter
<code>[0-9]</code>	match any digit
<code>[9018273645]</code>	match any digit
<code>[aeiouAEIOU]</code>	match any vowel
<code>[^aeiouAEIOU]</code>	match anything except a vowel
<code>[-A-Za-z]</code>	match any letter or a hyphen
<code>.</code>	match any one character

## Regular expressions

- Suppose we want to describe sheep language:  
*baa!*  
*baaa!*  
*baaaa!*  
*baaaaa!*
- This is a very simple language, but it's still infinite
- Can be modeled by the regular expression:  
`baaa*!`

## Regular expressions

- The Kleene star (\*) means 'match zero or more occurrences of the preceding pattern'
- This pattern will match baa, followed by zero or more a's, followed by !:

`baaa*!`

- The Kleene plus (+) means 'match one or more occurrences of the preceding pattern':

`baa+!`

`X*` match zero or more X's

`X+` match one or more X's

`X?` match zero or one X

`X{m,n}` match at least m but no more than n X's

`(X|Y)` match an X or a Y

## Regular expressions

- Operator precedence
  - `/angry|ily/` vs. `/angr(y|ily)/`  
  
`/angry|ily/` matches *angry* OR *ily*  
`/angr(y|ily)/` matches *angry* OR *angrily*
  - `/the*/` vs `/(the)*/`  
  
`/the*/` matches *th* OR *the* OR *thee* OR *theee*  
`/(the)*/` matches  $\epsilon$  OR *the* OR *thethe* OR *thethethe*
- Parentheses > Counters: \* + ? { } > Sequences and anchors > Disjunction

## Searching

- When we search for words in a text, we need to be careful of word boundaries
- The pattern  
`[Tt]he`  
will match the word *the*, but also *ether*, *thespian*, *lathe*
- We could try this:  
`[^a-zA-Z][Tt]he[^a-zA-Z]`
- Problems?



## Searching

- Anchors allow us to match positions
- ^ matches the beginning of the line and \$ matches the end:

```
( ^ | [ ^a-zA-Z ] ) [ Tt ] he ( $ | [ ^a-zA-Z ] )
```

- Watch out for context: ^ means one thing inside [ ] and something else outside
- A useful shorthand:

```
\b [ Tt ] he \b
```