CS5100 Foundations of Artificial Intelligence

Module 07 Lesson 12

Applications: Natural Language Processing

Overview

- Communication And NLP
- Issues with Language
- Grammar & Syntactic Analysis
- Evolution of NLP Systems
- Machine Translation

What is NLP?

Fundamental goal: analyze and process human language, broadly, robustly, accurately…

End systems that we want to build:

- Ambitious: speech recognition, machine translation, information extraction, dialog interfaces, question answering, intelligent assistants…
- Modest: spelling correction, text categorization…

Communication

"Classical" view (pre-1953):

- language consists of sentences that are true/false

"Modern" view (post-1953):

- language is a form of action

Wittgenstein (1953) Philosophical Investigations

- Look at real language, context

Austin (1962) How to Do Things with Words

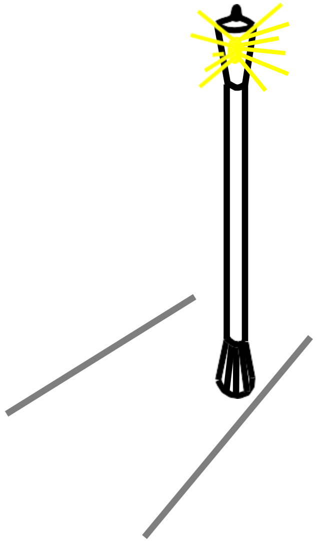
- Not just about True/False, but about getting things done

Searle (1969) Speech Acts

- Different kinds of acts

Why?

- To change the actions of other agents



Speech Acts

SITUATION

Speaker → Utterance → Hearer

Speech Acts achieve speaker's goals:

- **Inform**: “There’s a pit in front of you”
- **Query**: “Can you smell the stench?”
- **Command**: “Pick it up”
- **Promise**: “I’ll share the gold with you”
- **Acknowledge**: “OK”
- …

Speech Act Planning

Speech act planning requires knowledge of:

- Situation
- Semantic and Syntactic conventions
- Hearer’s goals, knowledge base, and rationality

Stages in Communication

Intention: Speaker S wishes to inform Hearer H that P

Generation: S selects words W to express P, in context C

Synthesis: S utters the words W

Perception: H perceives W' in context C'

Analysis: H infers possible meanings P_1, .. P_n

Disambiguation: H infers intended meaning P_i

Incorporation: H incorporates P_i into KB

Many ways all these steps can go wrong!

Communication Going Wrong

- Insincerity (S does not believe P)
- Speech wreck ignition failure
  - Ambient noise, hearing, …
- Ambiguous utterance
- Mumbling etc.
- Masks !
- Different understanding of context (C != C’)
- Vocabulary differences

Issues with Language

Real human languages provide many problems for NLP:

- ambiguity
- anaphora
- discourse structure
- metonymy
- metaphor
- non-compositionality
- Others (Vagueness…)

➔➔➔

Ambiguity

Ambiguity can be lexical (polysemy), syntactic, semantic, referential …

Consider the following:

Squad **helps** dog **bite** victim
Helicopter **powered** by human **flies**
This door is alarmed
She saw the man on the hill with a telescope
She saw (the man on the hill with a telescope)
She saw (the man on the hill) with a telescope
She saw (the man) on the hill with a telescope …

I ate spaghetti with

- meatballs
- salad
- much joy
- a fork
- a friend

Anaphora

Using pronouns to refer back to entities already introduced in the text.

Consider they, it in the following:

- After Mary proposed to John, they found a preacher and got married.
- For the honeymoon, they went to Hawaii

- Mary saw a ring through the window and asked John for it
- Mary threw a rock at the window and broke it

Metonymy

Using one noun phrase to stand for another

- I love John Le Carre
- Samsung announced record profits
- The pizza on Table 4 wants another beer

Metaphor

“Non-literal” usage of words and phrases,
often systematic:

- I’ve tried killing the process, but it won’t die. Its parent keeps it alive.

Non-compositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes

red book
red pen
red hair
red herring

More Ambiguities

Headlines:

- Enraged Cow Injures Farmer With Ax
- Hospitals Are Sued by 7 Foot Doctors
- Iraqi Head Seeks Arms
- Local HS Dropouts Cut in Half
- Juvenile Court to Try Shooting Defendant
- Stolen Painting Found by Tree
- Kids Make Nutritious Snacks

Why are these funny?

Grammar

- Grammar specifies the compositional structure of complex messages e.g., speech (linear), text (linear), music (two-dimensional)
- A formal language is a set of strings of terminal symbols
- Each string can be analyzed/generated by the grammar
- The grammar is a set of rewrite rules, e.g.,

  S → NounPhrase VerbPhrase (or S ➔ NP VP)
  Article → a | an | the

  Here S is the sentence symbol, NP and VP are non-terminals

Grammar Types

Chomsky hierarchy, 1956

Regular: nonterminal → terminal [nonterminal]

S ➔ aS

S ➔ a

Context-free: nonterminal → anything (Prog Langs, NL)

S → aSb

Context-sensitive: RHS contains at least as many non-terminals as LHS

ASB → AAaBB

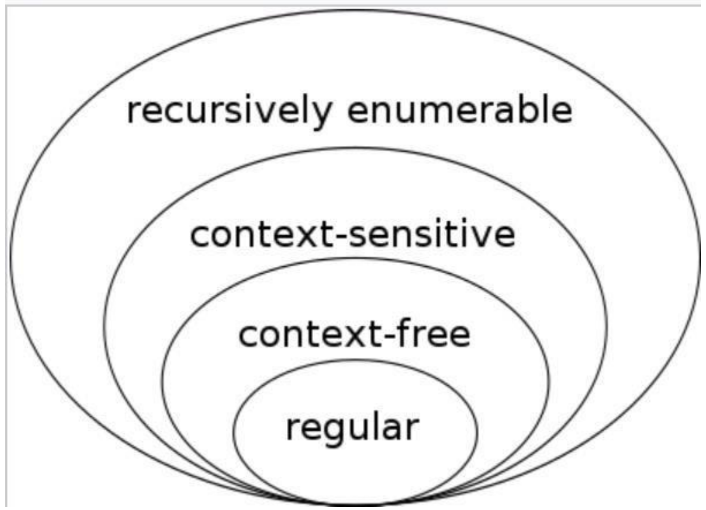
Recursively enumerable: no constraints

A B C ➔ D E etc..

More expressive power, less efficient algorithms as we go down the page

Natural languages probably context-free, parsable in real time!

Noam Chomsky



[Chomsky hierarchy - Wikipedia]

Wumpus Lexicon

Noun → stench | breeze | glitter | wumpus | pit| pits | nothing | gold | east | west …

Verb → is | see | smell | shoot | feel | stinks | go| grab | carry | kill | turn| shot …

Adjective → right | left | smelly …

Adverb → here | there | nearby | ahead

Pronoun → me| you | I | it

Article → a | an | the

Preposition → to | in | on | near

Conjunction → and | or | but

Wumpus Grammar

S → NP VP | S Conjunction S

I + feel a breeze

I feel a breeze + and + I smell a Wumpus

NP → Pronoun | Noun | Article Noun | Digit {Digit}* |
NP PP | NP RelClause

pits , the + Wumpus , the Wumpus + to the east,
the Wumpus + that is smelly, it, 3

VP → Verb | VP NP | VP Adjective | VP PP | VP Adverb

turn + to the east, go + ahead

PP → Preposition NP

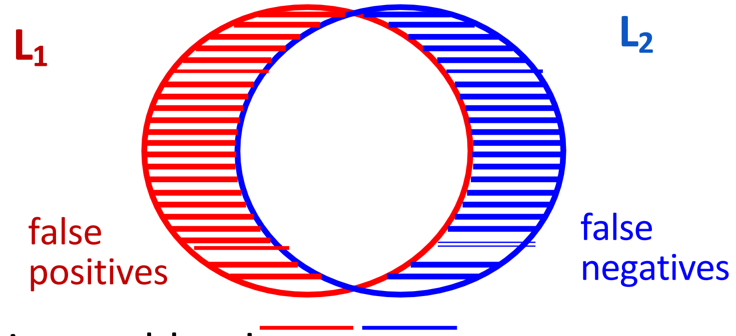
to + the east

RelClause → {that | which…} VP

that + is smelly

Grammaticality Judgments

Formal language L_1 may differ from natural language L_2



Adjusting L_1 to agree with L_2 is a learning problem!

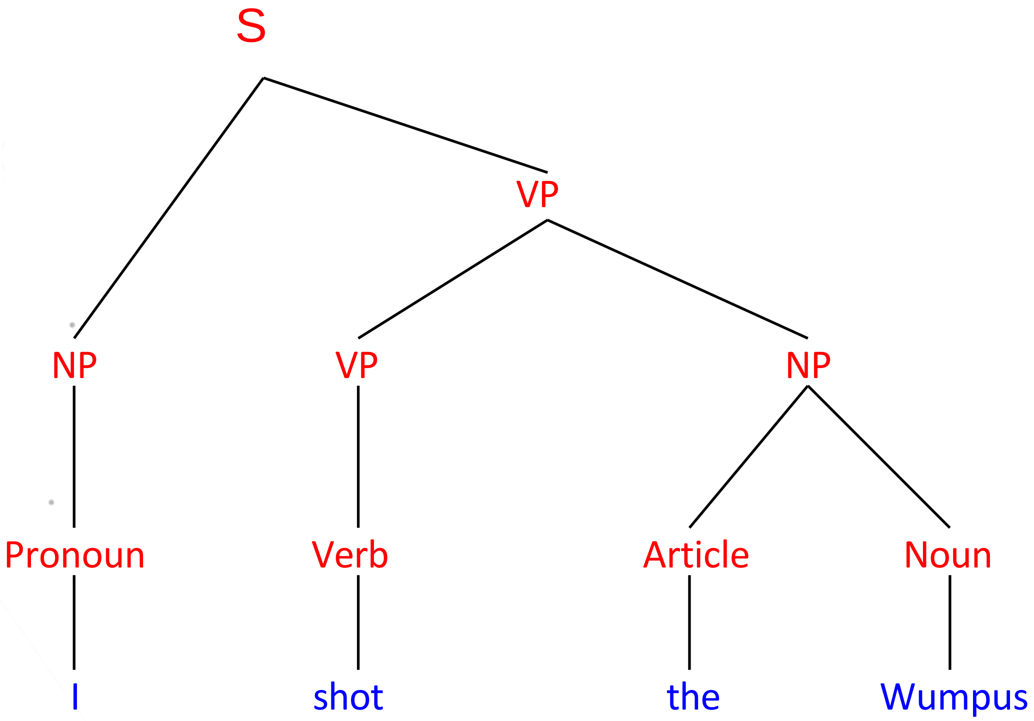
- the gold grab the Wumpus
- I smell the Wumpus the gold
- give the Wumpus the gold
- I donate the Wumpus the gold

Intersubject agreement somewhat reliable, independent of semantics!

Real grammars 10–500 pages, insufficient even for “proper” English

Contrast with colloquialisms, texting language, child language…

Parse Trees



Noun → stench | breeze | glitter | wumpus | pit | pits | nothing | gold | east | west ...

Verb → is | see | smell | shoot | feel | stinks | go | grab | carry | kill | turn | shot ...

Adjective → right | left | smelly ...

Adverb → here | there | nearby | ahead

Pronoun → me | you | I | it

Article → a | an | the

Preposition → to | in | on | near

Conjunction → and | or | but

S → NP VP | S Conjunction S
I + feel a breeze
I feel a breeze + and + I smell a Wumpus

NP → Pronoun | Noun | Article Noun | Digit {Digit}* | NP PP | NP RelClause
pits , the + Wumpus , the Wumpus + to the east, the Wumpus + that is smelly, it, 3

VP → Verb | VP NP | VP Adjective | VP PP | VP Adverb
turn + to the east, go + ahead

PP → Preposition NP
to + the east

RelClause → {that | which...} VP
that + is smelly

Syntax in NLP

Most view syntactic structure as an essential step towards meaning

"Mary hit John" != "John hit Mary"

1. "And since I was not informed—as a matter of fact, since I did not know that there were excess funds until we, ourselves, in that checkup after the whole thing blew up, and that was, if you'll remember, that was the incident in which the attorney general came to me and told me that he had seen a memo that indicated that there were no more funds."

   Ronald Reagan, 40th US President **Interview with White House Newspaper Correspondents,** *April 28, 1987* *http://www.presidency.ucsb.edu/ws/index.php?pid=34191*

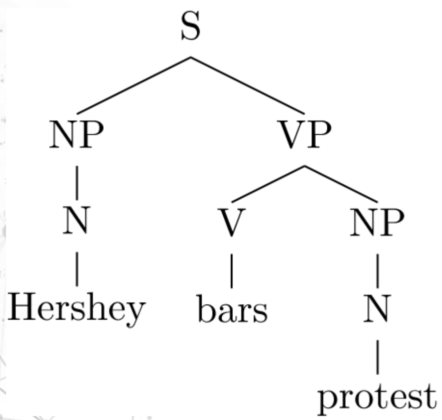
2. "Wouldn't the sentence 'I want to put a hyphen between the words Fish and And and And and Chips in my Fish-And-Chips sign' have been clearer if quotation marks had been placed before Fish, and between Fish and and, and and and And, and And and and, and and and And, and And and and, and and and Chips, as well as after Chips?"

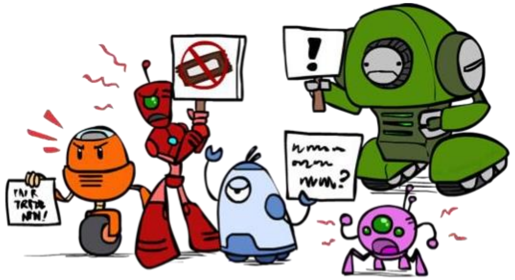
   Martin Gardner, https://en.wikipedia.org/wiki/List_of_linguistic_example_sentences

Context-free Parsing

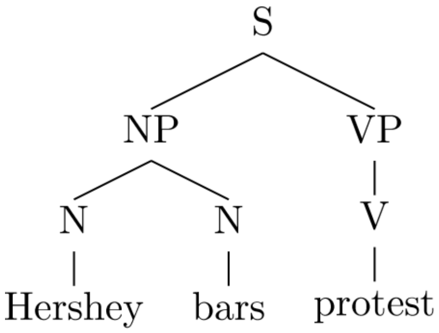
- Bottom-up parsing works by replacing any substring that matches RHS of a rule with the rule's LHS
- Efficient algorithms (e.g., chart parsing) $O(n^3)$ for context-free grammars, run at several thousand words/sec for real grammars
- Context-free parsing ≡ Boolean matrix multiplication ⇒unlikely to find faster practical algorithms

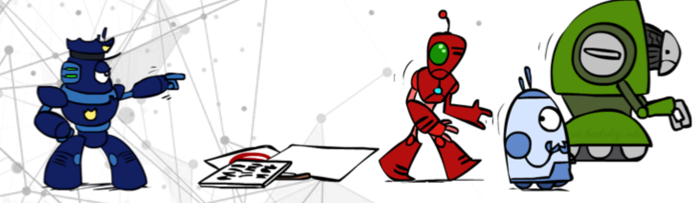
Parsing as Search

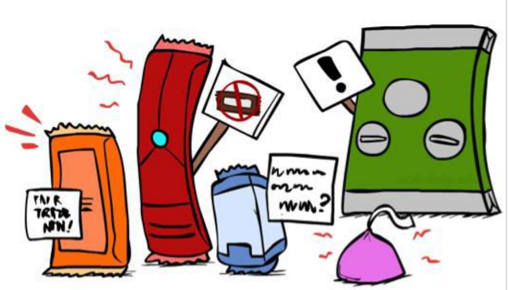




Hershey bars protest

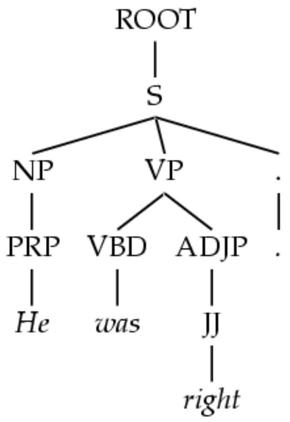






Grammar: PCFGs

- Natural language grammars are very ambiguous!
- Probabilistic CFGs (PCFGs) are a formal probabilistic model of trees
  - Each “rule” has a conditional probability
  - Tree’s probability is the product of the probabilities of all rules used
- Parsing: Given a sentence, find the best tree – search!



ROOT $\rightarrow$ S 375/420

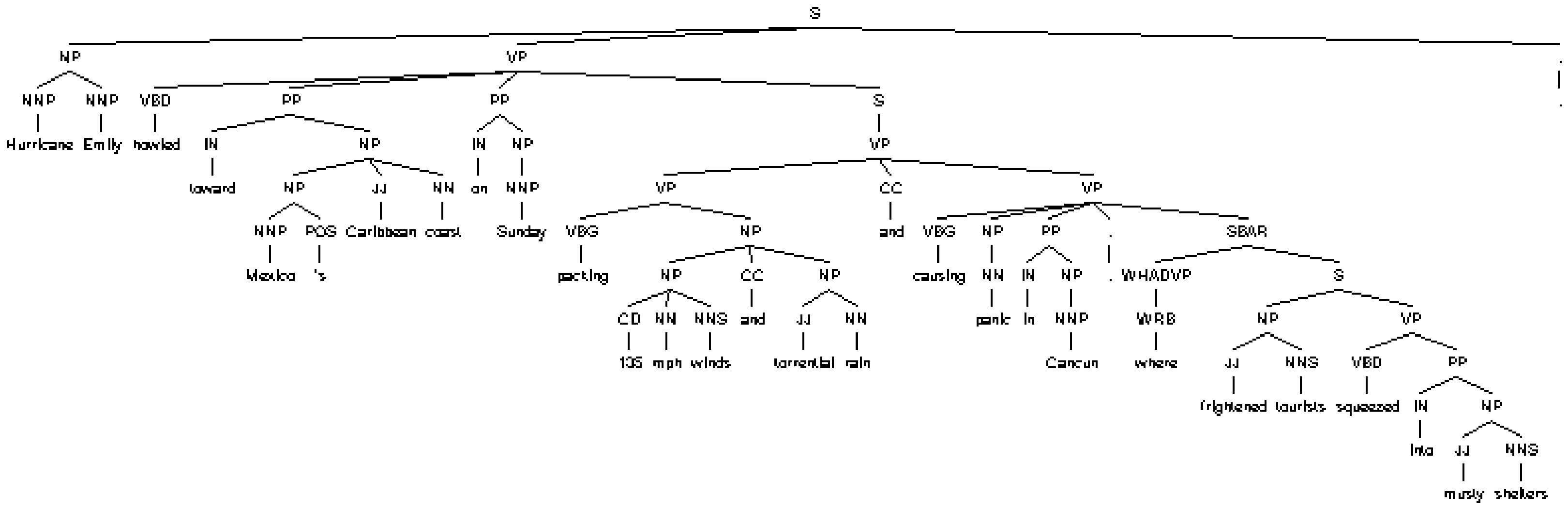
S $\rightarrow$ NP VP . 320/392

NP $\rightarrow$ PRP 127/539

VP $\rightarrow$ VBD ADJP 32/401

…..

Syntactic Analysis ..1

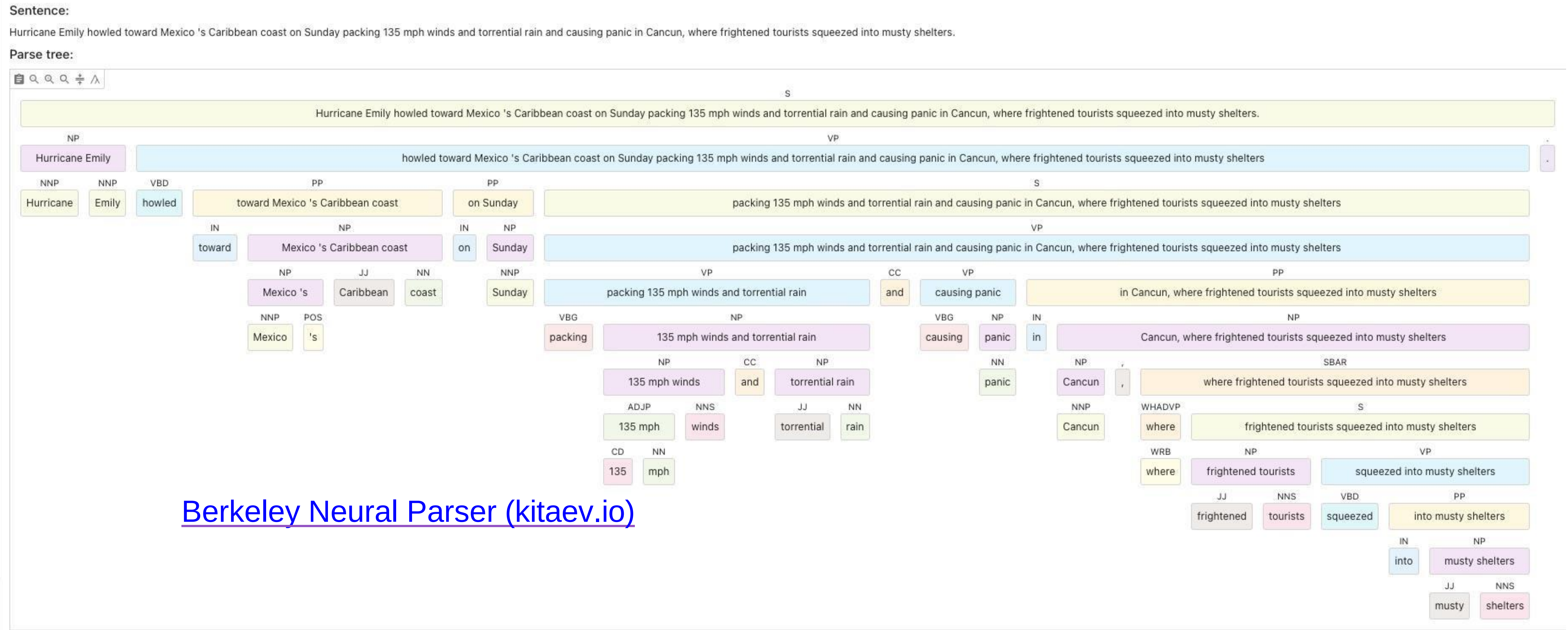


Hurricane Emily howled toward Mexico 's Caribbean coast on Sunday packing 135 mph winds and torrential rain and causing panic in Cancun, where frightened tourists squeezed into musty shelters.

[Demo: Berkeley NLP Group Parser http://tomato.banatao.berkeley.edu:8080/parser/parser.html
No longer available]

Syntactic Analysis ..2

Hurricane Emily howled toward Mexico 's Caribbean coast on Sunday packing 135 mph winds and torrential rain and causing panic in Cancun, where frightened tourists squeezed into musty shelters.



[Berkeley Neural Parser (kitaev.io)](https://parser.kitaev.io)

Evolution of NLP Systems
Turing's Imitation Game

- **Alan Turing** proposed [1950] :
    - Machine interacts with human observer using NL conversation via screen & keyboard
    - Machine "intelligent" if it could convince observer that it was human
    - "Turing Test"

    - ➔ Demonstrate understanding and use of language

ELIZA: Imitating a Psychotherapist

Produce human-like responses to user input [1966]

Joseph Weizenbaum, MIT

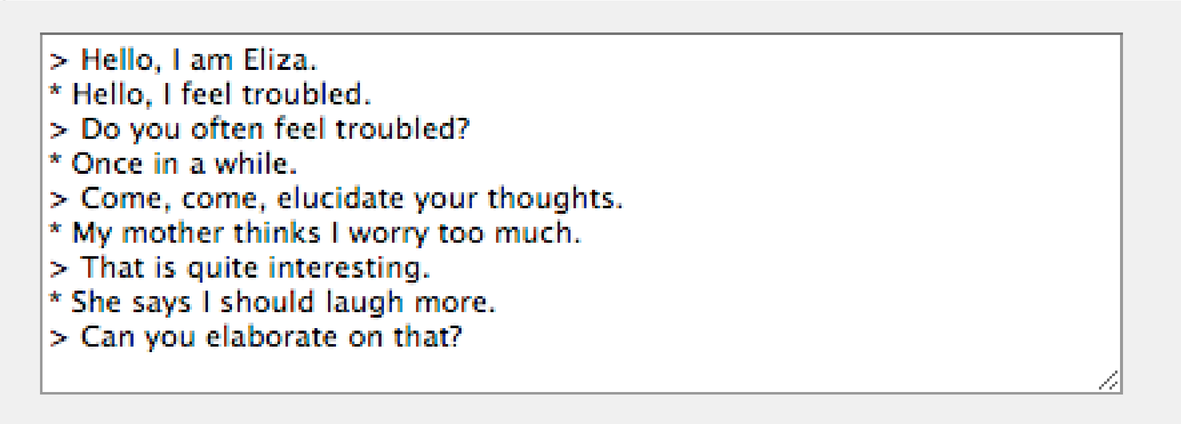
How does it work?

- Parses input
- Uses templates
  - e.g. * mother * ➔ "Tell me more about your family"
- Makes simple substitutions
  - e.g. I ➔ you, you ➔ me, my ➔ your…
- Uses canned phrases
  - e.g. "Can you elaborate on that?"

ELIZA: Demo ++

Demo at: http://www.simonebaldassarri.com/eliza/eliza.html

- Very limited



Profound impact on people, e.g., Weizenbaum’s secretary

Led to a lot of related work

What came from work on ELIZA

- Weizenbaum authored influential book “Computer Power and Human Reason” [1976]
- “… while Artificial Intelligence may be possible, we should never allow computers to make important decisions because computers will always lack human qualities such as compassion and wisdom. …
  Deciding is a computational activity, something that can ultimately be programmed. Choice, however, is the product of judgment, not calculation.

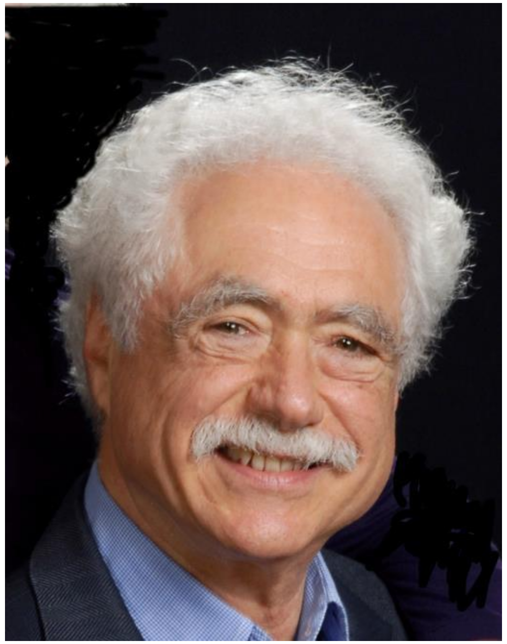
  It is the capacity to choose that ultimately makes us human. Comprehensive human judgment is able to include non-mathematical factors, such as emotions. ”

https://en.wikipedia.org/wiki/Joseph_Weizenbaum

SHRDLU

- Operated in simplified blocks world
- **Terry Winograd**, MIT (now at Stanford)
- Most common characters in English used to be…E T A O I N S H R D L U …

- User could converse with SHRDLU to name blocks, move them, ask about the state of the blocks
  - Limited vocabulary, simple domain
  - Name objects/collections, arrange them, answer questions, recreate arrangements
  - Some limited memory, e.g., “Put the blue block on the red block”
  - Some “physics”: e.g.. cannot stack block on top of a cone



http://hci.stanford.edu/~winograd/

SHRDLU: Sample Interaction

Code + demo at http://hci.stanford.edu/~winograd/shrdlu/

- Person: PICK UP A BIG RED BLOCK.
- Computer: OK. (does it)
- Person: GRASP THE PYRAMID.
- Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
- Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.
- Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
- Computer: OK. (does it)
- Person: WHAT DOES THE BOX CONTAIN?
- Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
- Person: WHAT IS THE PYRAMID SUPPORTED BY?
- Computer: THE BOX.

- Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
- Computer: FOUR OF THEM.
- Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?
- Computer: YES, THE RED CUBE.
- Person: IS IT SUPPORTED?
- Computer: YES, BY THE TABLE.
- Person: CAN THE TABLE PICK UP BLOCKS?
- Computer: NO.
- Person: CAN A PYRAMID BE SUPPORTED BY A BLOCK?
- Computer: YES.
- …

START

- Answers NL questions
  about its knowledge base
- **Boris Katz** and his team
  (1993) from … MIT !
- START: SynTactic Analysis using Reversible Transformations
- Uses “annotations” (sentences/phrases to describe information segments)
- Data parsed into info segments; answers generated from info segments
- Input questions matched via annotations to info segments
- Demo at at http://start.csail.mit.edu/

  http://start.csail.mit.edu/answer.php?query=What+is+the+population+of+Bellevue%3F

AskMSR .. 1

- Answers NL questions using info from the Web
- Michele Banko, Eric Brill, Susan Dumais, Jimmy Lin [2002]
  -- work done at Microsoft Research, Redmond!
- How it worked:
  - Transformed user query into web search query
  - “Who is Bill Gates married to?” ➔ “Bill Gates is married to”
  - Query is sent to a search engine, results analyzed
  - From result summaries, substrings extracted based on query types
    - i.e. is it a *Who* question or *What* question etc..
  - Answers assembled from frequent substrings.




AskMSR ..2

- Depends on information redundancy on the Web
  - no means to check the validity of the answers it returns.
- But if many instances of “Bill Gates is married to his work” are seen,
  the answer to “Who is Bill Gates married to?”
  may very well come out as “His work” !

Watson: Overview

- Question-answering system [2010]
  IBM, with CMU, MIT, UMass, USC, Univ of Texas, Univ of Trento;
  David Ferrucci, Principal Investigator, DeepQA/Watson project
- Designed to play the game of Jeopardy!
- How it works:
  - Sophisticated NLP: deep analysis of questions, noisy matching of questions to potential answers
  - Lots of data: information from a huge collection of documents (e.g. Wikipedia, etc..), exploits redundancy
  - Lots of computation: 90+ servers

Watson: Jeopardy!

- Jeopardy: 3 rounds, 3 contestants
  - 5 questions, 6 categories for first 2 rounds
  - 1 question in 3rd round
- Clue is an assertion with missing information;
  response has to be in the form of a question
  e.g. For clue: *"This harvest festival is celebrated by the Tamil people of India in January",*
  a valid response would be *"What is Pongal?"*
- Need to choose category & clue, buzz-in fast if confident of answer, and have a good wagering strategy

Watson: Metrics

- Metrics:
  - Percent-answered: % of questions system chooses to answer
  - Precision: % correct, of those it answers
- Humans answer 40-50% with 85-95% precision;
  champion Ken Jennings answered 62% with 92% precision.

  ➔ Ken's metrics a goal for Watson!

Watson: How it Works

1. Analyze question
   Category (puzzle? pun ?); Answer type;
   Is there a relation between entities; question decomposable?
2. Generate Hypotheses
   Generate 100s of responses, filtered later
3. Soft filter
   Use question classification, answer type to reduce to ~ 100 candidate answers
4. Hypothesis & Evidence Scoring
   Gather all evidence for each candidate, get combined score for each
5. Merge equivalent answers
   Dr. King, Martin Luther King Jr., Rev. King are equivalent
6. Use ML to rank different candidates, and come up with best answer and associated confidence value

Watson
Performance & Learnings

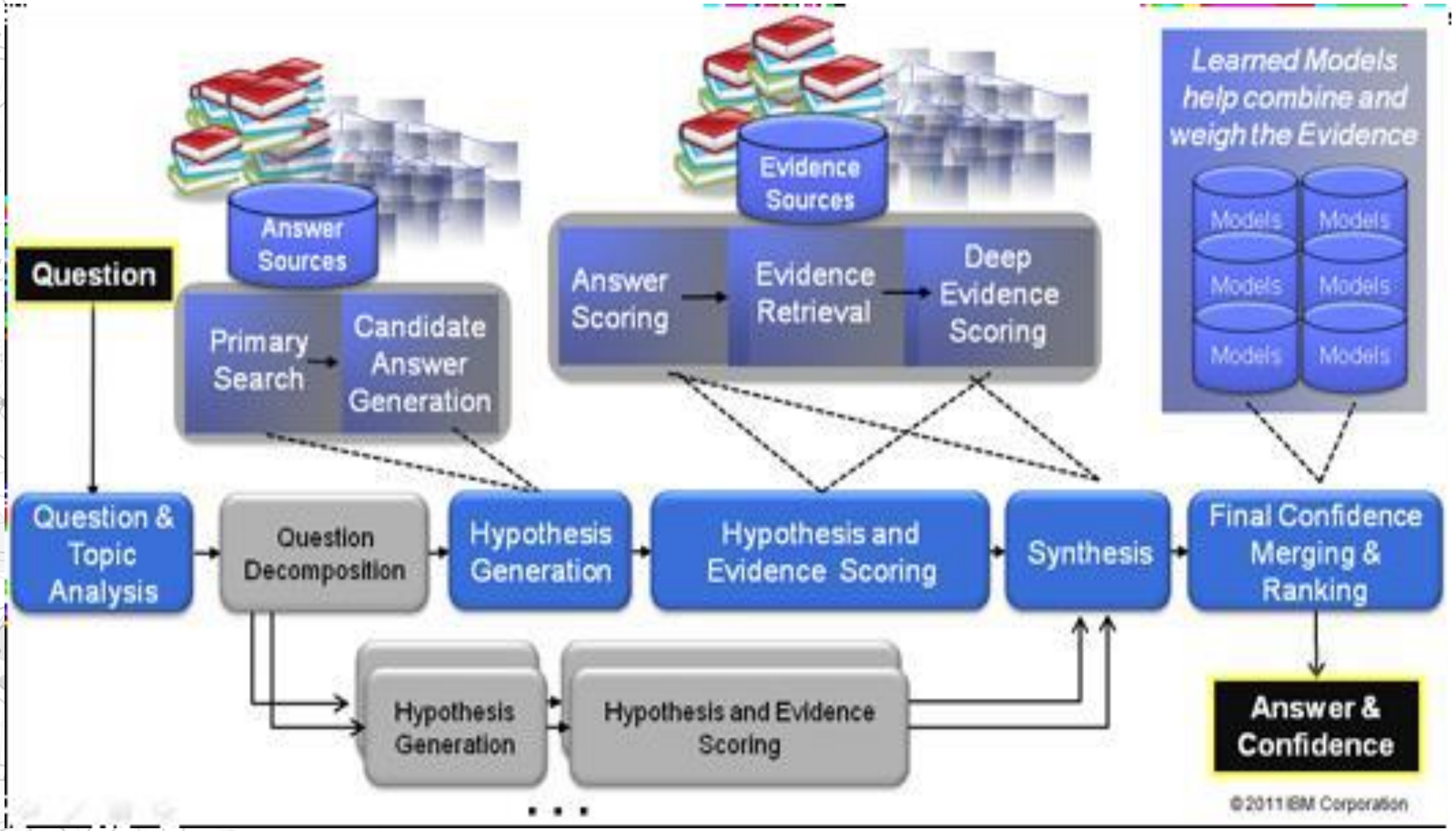
Performance in matches, broadcast Feb 14-16, 2011

- First match
  - Watson: $35, 734 ; Brad Rutter: $10, 400; Ken Jennings: $4,800
- Second match
  - Watson: $77, 147 ; Ken Jennings: $24,000; Brad Rutter: $21, 600
    Prizes: $1m, $0.3m, $0.2m

Learnings:

- Showed system could perform at championship level
- Good to have flexible architecture
- Generate & Test was good
- General CS smarts used to improve speed

DeepQA High-Level Architecture



http://researcher.ibm.com/researcher/view_project_subpage.php?id=2159

Is NLP a solved problem?

- The strategy used for Jeopardy! is not generic
- Playing Jeopardy! is a very specific task

- Lots more to do!

Intelligent Assistants

Alexa, Cortana, Google Home

1. Good microphone array
2. Speech to text
3. Intent Understanding, probably
   via templates and heuristics
4. Great backend search +
   links to home automation, music apps etc..
5. Text to Speech Synthesis
6. Decent loudspeakers

Machine Translation

Translate text from one language to another

Recombines fragments of example translations

Challenges:

- What fragments? [learning to translate]
- How to make it efficient? [fast translation search]

"Il est impossible aux journalistes de rentrer dans les régions tibétaines"

Bruno Philip, correspondant du "Monde" en Chine, estime que les journalistes de l'AFP qui ont été expulsés de la province tibétaine du Qinghai "n'étaient pas dans l'illégalité".

Les faits Le dalaï-lama dénonce l'"enfer" imposé au Tibet depuis sa fuite, en 1959

Vidéo Anniversaire de la rébellion

"It is impossible for journalists to enter Tibetan areas"

Philip Bruno, correspondent for "World" in China, said that journalists of the AFP who have been deported from the Tibetan province of Qinghai "were not illegal."

Facts The Dalai Lama denounces the "hell" imposed since he fled Tibet in 1959

Video Anniversary of the Tibetan rebellion: China on guard

Problem with Dictionary Lookups

顶部	/**top**/roof/
顶端	/summit/peak/**top**/apex/
顶头	/coming directly towards one/**top**/end/
盖	/lid/**top**/cover/canopy/build/Gai/
盖帽	/surpass/**top**/
极	/extremely/pole/utmost/**top**/collect/receive/
尖峰	/peak/**top**/
面	/fade/side/surface/aspect/**top**/face/flour/
摘心	/**top**/topping/

Example from Douglas Hofstadter

MT
60 Years in 60 Seconds

Warren Weaver

When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

John Pierce

"Machine Translation" presumably means going by algorithm from machine-readable source text to useful target text… In this context, there has been no machine translation…

MT is the "first" non-numeral compute task

Berkeley's first MT grant

ALPAC report deems MT bad

Statistical data-driven approach introduced

Statistical MT thrives

Deep MT

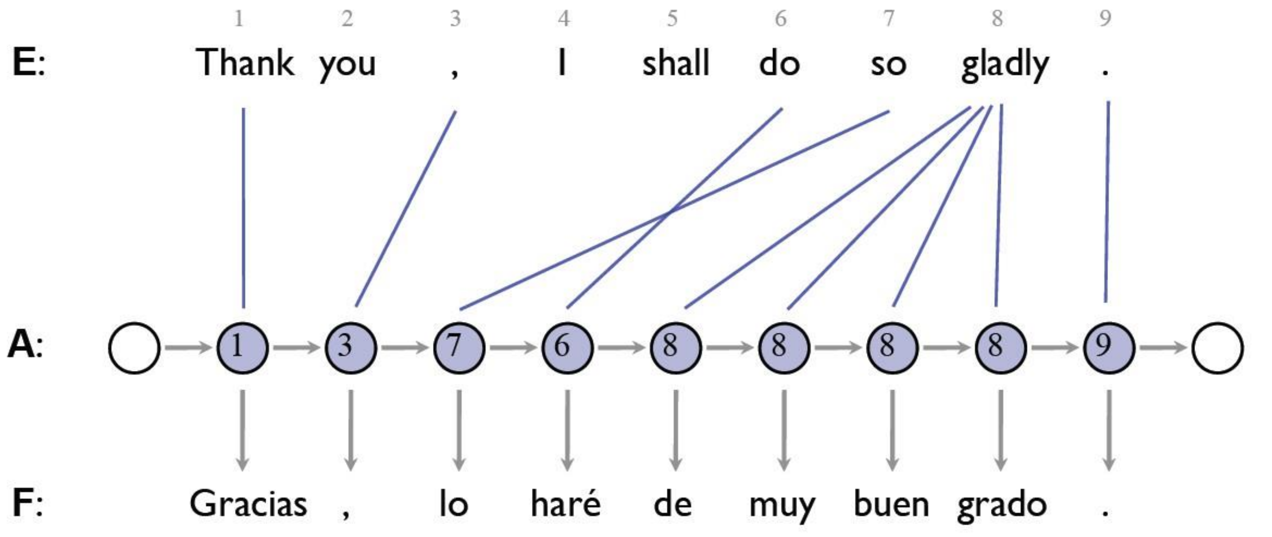
'47 '58 '66 '90's '00's

Learning to Translate

	No.	CLASSIC SOUPS	Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲呑	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲呑湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

Example from Adam Lopez

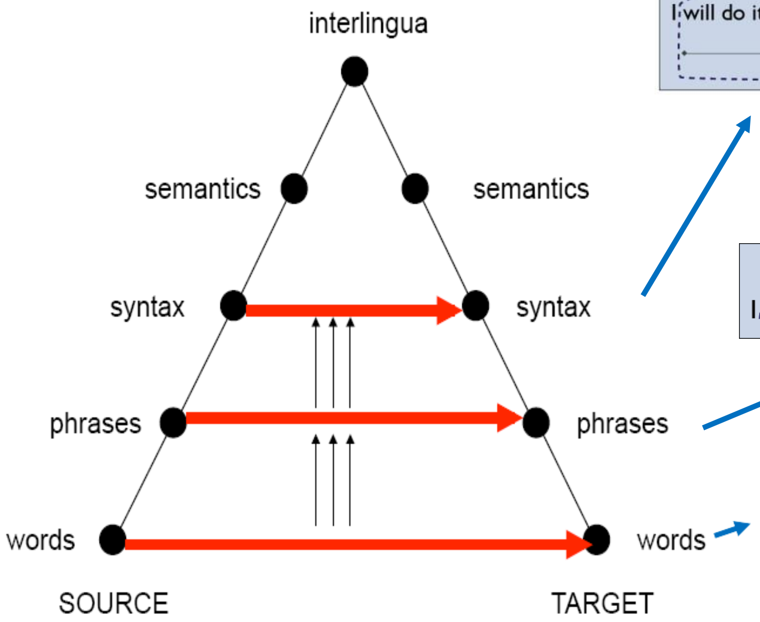
An HMM Translation Model

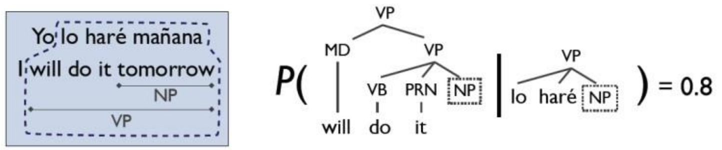


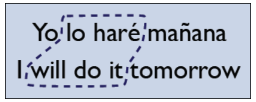
Model Parameters

Emissions: $P(F_1 = \text{Gracias} \mid E_{A_1} = \text{Thank})$ *Transitions*: $P(A_2 = 3 \mid A_1 = 1)$

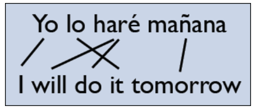
Levels of Translation
The Vauquois triangle (1968)





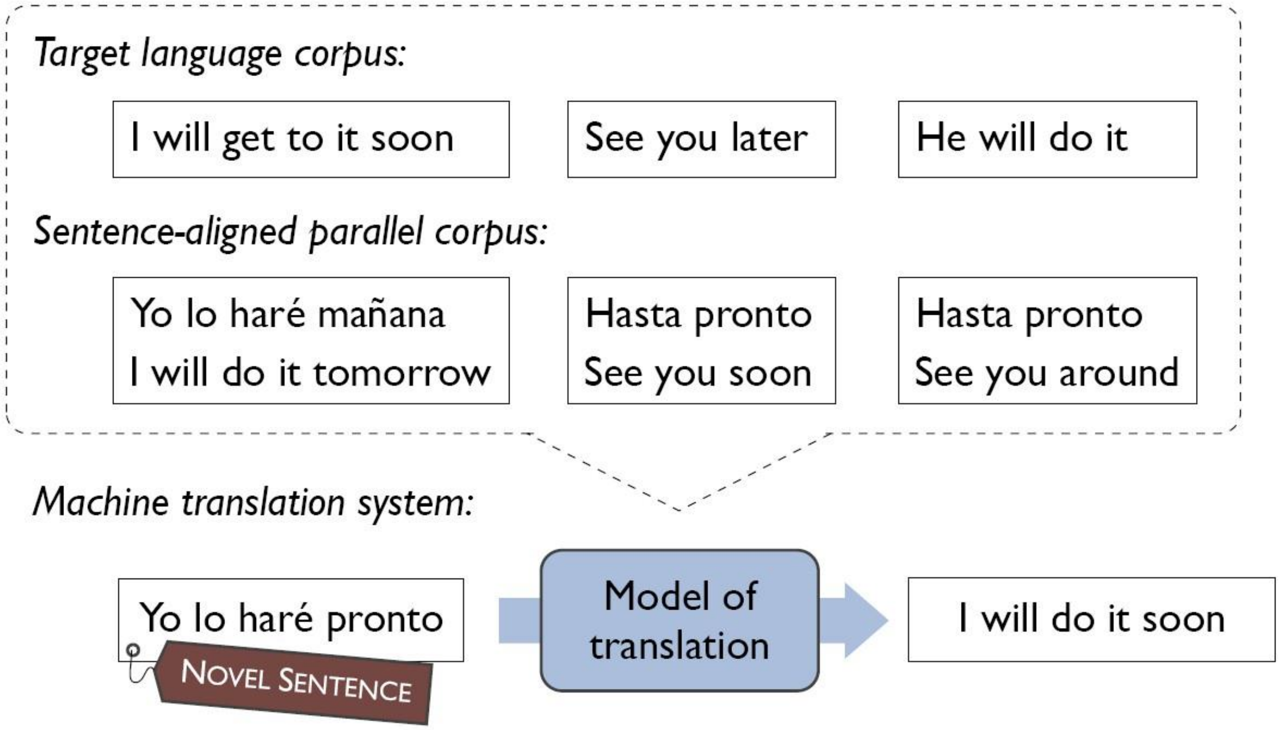


English (E)	*P*(E | lo haré)
will do it	0.8
will do so	0.2



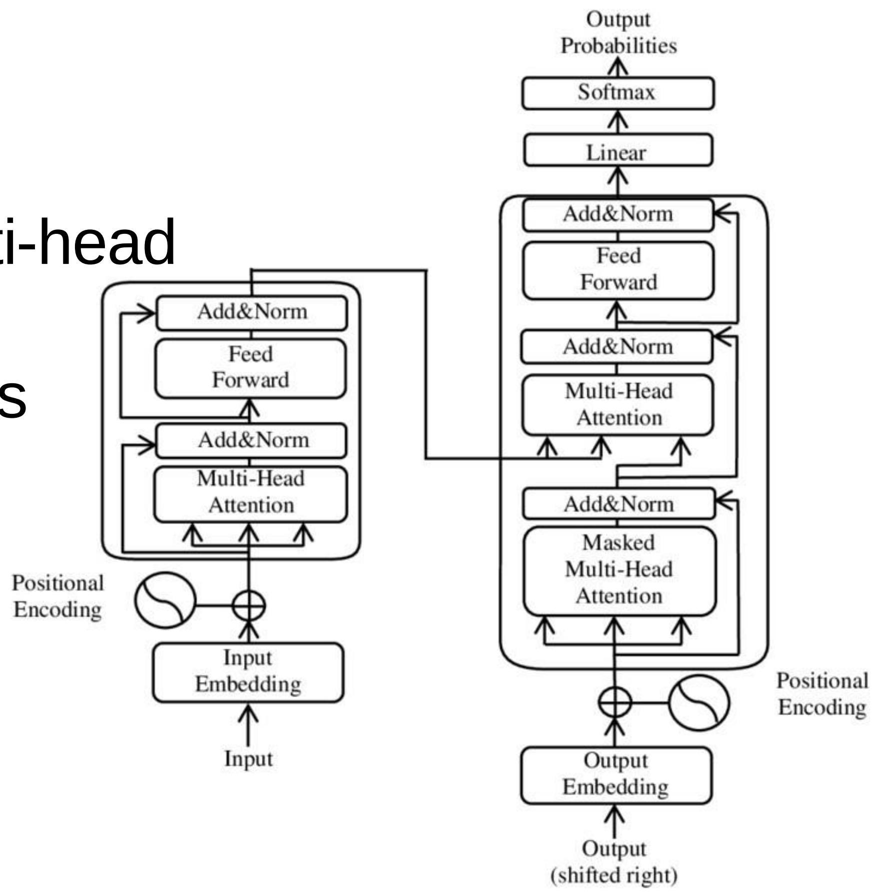
English (E)	*P*(E | mañana)
tomorrow	0.7
morning	0.3

Data-Driven Machine Translation



Deep Learning: Transformer Technology

- Developed by Google 2017
  - “Attention is All You Need”
  - Vaswani et. al.
- Tokens are “contextualized” via “parallel multi-head attention mechanism”
- Allows for key tokens to be amplified and less important ones to be dimished



Deep Learning: Transformer Technology

- GPT is a Transformer-based architecture and training procedure for natural language processing tasks
- Training follows a two-stage procedure
  - First, a language modeling objective is used on the unlabeled data to learn the initial parameters of a neural network model
  - Subsequently, these parameters are adapted to a target task using the corresponding supervised objective.

Deep Learning: Transformer Technology

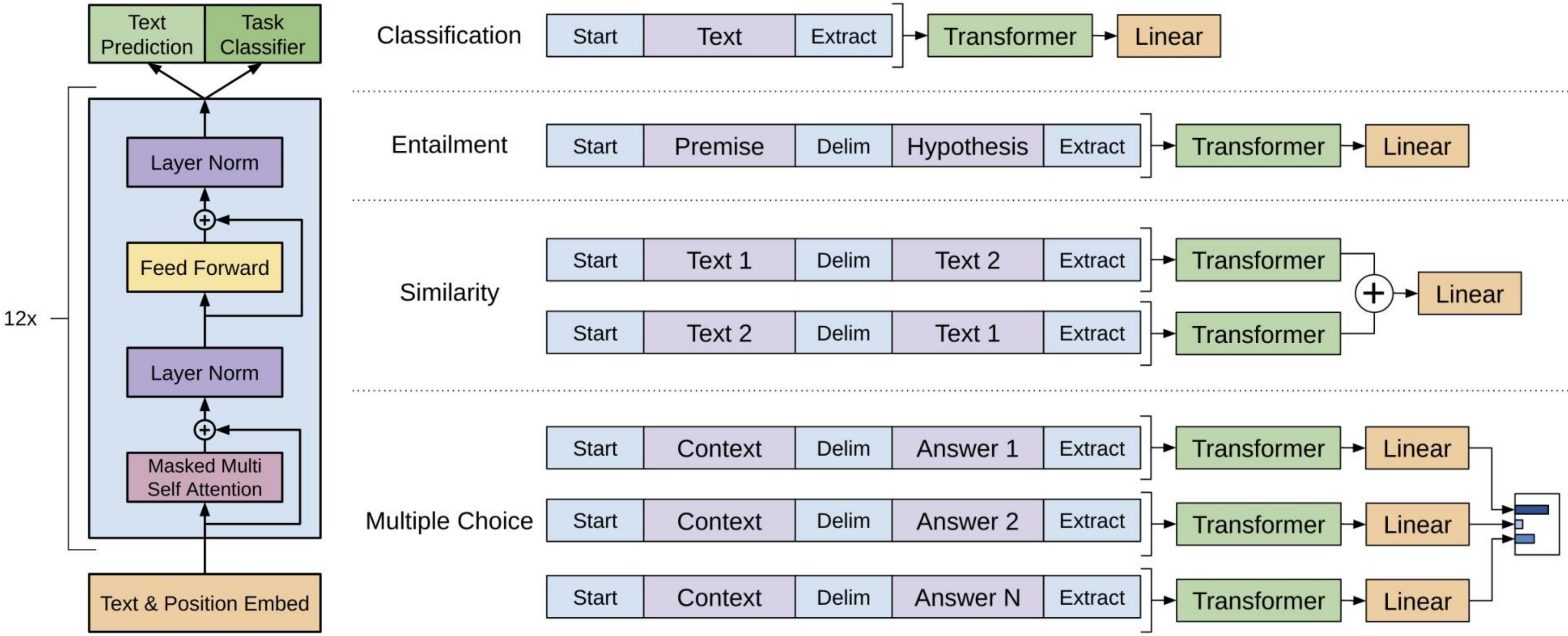
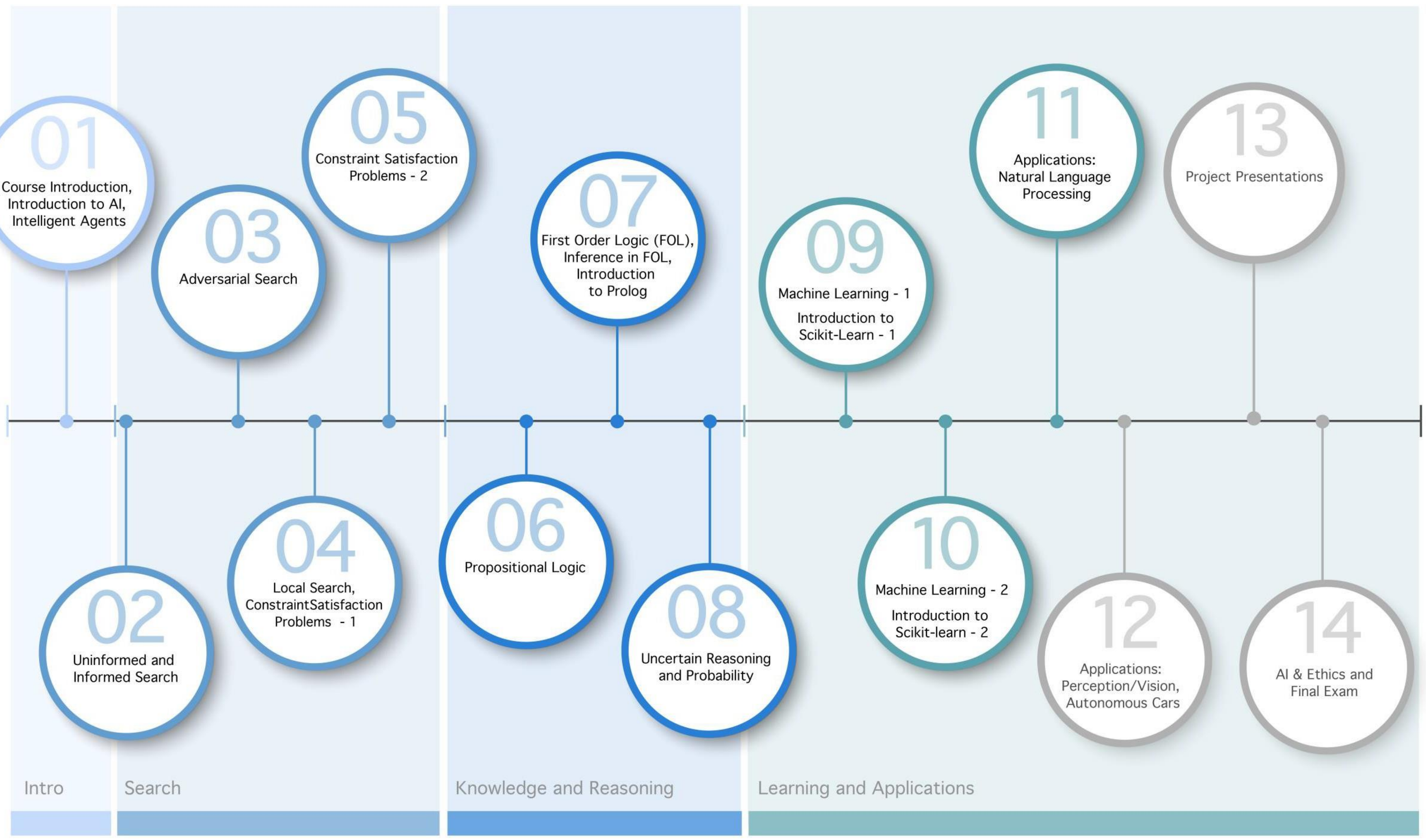


Figure 1: **(left)** Transformer architecture and training objectives used in this work. **(right)** Input transformations for fine-tuning on different tasks. We convert all structured inputs into token sequences to be processed by our pre-trained model, followed by a linear+softmax layer.

https://sebastianraschka.com/pdf/lecture-notes/stat453ss21/L19_seq2seq_rnn-transformers slides.pdf

How ChatGPT Works

https://www.youtube.com/watch?v=-4Oso9-9KTQ


01
Course Introduction, Introduction to AI, Intelligent Agents
02
Uninformed and Informed Search
03
Adversarial Search
04
Local Search, ConstraintSatisfaction Problems - 1
05
Constraint Satisfaction Problems - 2
06
Propositional Logic
07
First Order Logic (FOL), Inference in FOL, Introduction to Prolog
08
Uncertain Reasoning and Probability
09
Machine Learning - 1
Introduction to Scikit-Learn - 1
10
Machine Learning - 2
Introduction to Scikit-learn - 2
11
Applications: Natural Language Processing
12
Applications: Perception/Vision, Autonomous Cars
13
Project Presentations
14
AI & Ethics and Final Exam
Intro
Search
Knowledge and Reasoning
Learning and Applications