

LECTURE 1: INTRODUCTION

- Dr. Kiran Bhowmick

Syllabus

Detailed Syllabus: (unit wise)		
Module	Contents	Hours
1	Introduction: History Of NLP, Generic NLP System, Levels Of NLP, Knowledge In Language Processing, Ambiguity In Natural Language, Stages In NLP, Challenges Of NLP, Applications of NLP	03
2	Word Level Analysis: Morphology Analysis –Survey of English Morphology, Inflectional Morphology & Derivational Morphology, Lemmatization, Regular Expression, Finite Automata, Finite State Transducers (FST), Morphological Parsing With FST, Lexicon Free FST Porter Stemmer. N –Grams, Unigrams/Bigrams Language Models, Corpora, Computing the Probability Of Word Sequence, Training and Testing. Perplexity And Entropy: Smoothing and Backup, Zipf's Law, Add One Smoothing, Witten-Bell Discounting, Good Turing Discounting, Back Off Methods, Class Based Models, Google N-Gram Release.	08
3	Syntax Analysis: Part-Of-Speech Tagging (POS) - Open and Closed Words. Tag Set for English (Penn Treebank), Rule Based POS Tagging, Transformation Based Tagging, Stochastic POS Tagging and Issues –Multiple Tags & Words, Unknown Words. Hidden Markov Model (HMM), Maximum Entropy, And Conditional Random Field (CRF). CFG: Derivations, Constituency, Phrase Structure and Dependency Structure	08

Syllabus

4	Semantic Analysis: Lexical Semantics, Attachment for Fragment of English- Sentences, Noun Phrases, Verb Phrases, Prepositional Phrases, Relations Among Lexemes & Their Senses – Homonymy, Polysemy, Synonymy, Hyponymy, WordNet, Robust Word Sense Disambiguation (WSD), Lexical Disambiguation, Resolving Lexical Ambiguity, Lexical Ambiguity Resolution	06
5	Pragmatics: Discourse –Reference Resolution, Reference Phenomenon, Syntactic & Semantic Constraints on Co Reference	06
6	Neural Models of Word Representations: Problems With SVD; Intro to Word2vec, Learning Word Representations, Recurrent Neural Networks (RNNs), RNNs On POS Tagging, Statistical Machine Translation with RNNs	06
7	Applications (Preferably for Indian Regional Languages): Machine Translation, Information Retrieval, Question Answers System, Categorization, Summarization, Sentiment Analysis, Named Entity Recognition Linguistic Modeling – Neurolinguistics Models- Psycholinguistic Models – Functional Models of Language – Research Linguistic Models- Common Features of Modern Models of Language	05

Textbook:

1. Speech and Language Processing, *2nd Edition*, Jurafsky and Martin, Prentice Hall; (January 26, 2000), ISBN: 0130950696

Reference Books:

1. Manning and Schutze, "Statistical Natural Language Processing", MIT Press; 1st edition (June 18, 1999), ISBN: 0262133601
2. James Allen. Natural Language Understanding. The Benajmins/Cummings Publishing Company Inc. 1994. ISBN 0-8053-0334-0.
3. Tom Mitchell. Machine Learning. McGraw Hill, 1997. ISBN 0070428077
4. Cover, T. M. and J. A. Thomas: Elements of Information Theory. Wiley. 1991. ISBN 0-471-06259-6.
5. Charniak, E.: Statistical Language Learning. The MIT Press. 1996. ISBN 0-262-53141-0.
6. Jelinek, F.: Statistical Methods for Speech Recognition. The MIT Press. 1998. ISBN 0-262-10066-5.

Experiment list

Expt. No.	Name of Experiment	CO mapping
1	Perform pre-processing of Text on any dataset.	1,2
2	Generate bigrams and trigrams from a given corpus and calculate probability of a sentence.	3
3	Chunking and Named Entity Representation.	3,4
4	HMM in POS Tagging.	4
5	Perform Morphological Analysis on a word.	3
6	Word Sense Disambiguation.	4,5
7	Plagiarism Detection Using NLP.	5,6
8	Mini Project.	1-6

What is NLP

- **Wiki: Natural language processing (NLP)** is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (**natural**) languages.



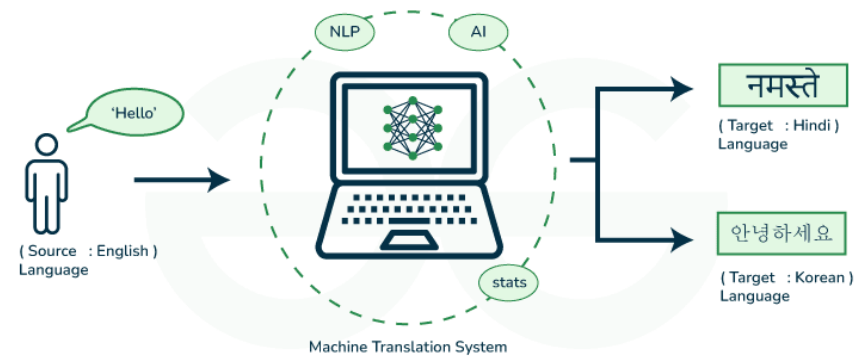
Go beyond the keyword matching



- Identify the structure and meaning of words, sentences, texts and conversations
- Deep understanding of broad language
- NLP is all around us

Machine translation

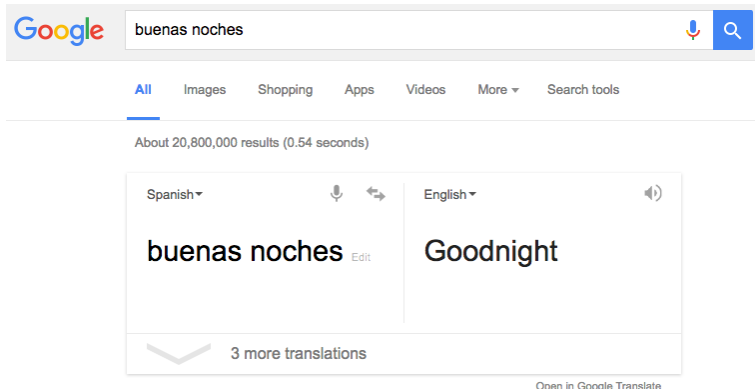
Machine translation is the process of using artificial intelligence to automatically translate text from one language to another without human involvement.



Machine Translation Model

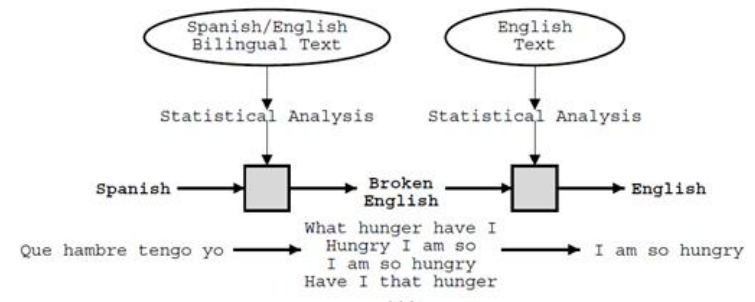
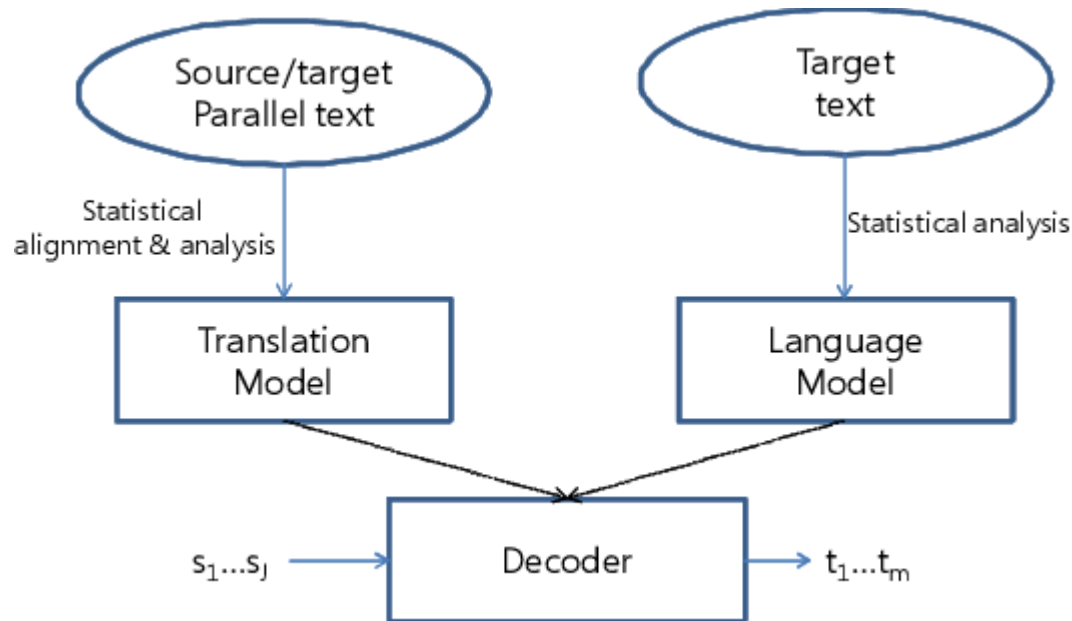


Example: Google translate, facebook tranlation



Statistical machine translation

Statistical machine translation (SMT) is a type of machine translation (MT) that uses statistical models to translate text from one language to another. Unlike traditional rule-based systems, SMT relies on large bilingual text corpora to build probabilistic models that determine the likelihood of a sentence in the target language given a sentence in the source language.




APPLICATIONS OF NLP

Dialog Systems

Gift shop


Items such as caps, t-shirts, sweatshirts and other miscellanea such as buttons and mouse pads have been designed. In addition, merchandise for almost all of the projects is available.



Hi. I'm your automated online assistant. How may I help you?


CD or DVD

There is a series of CDs/DVDs with selected Wikipedia content being produced by Wikipedians and [SOS Children](#).

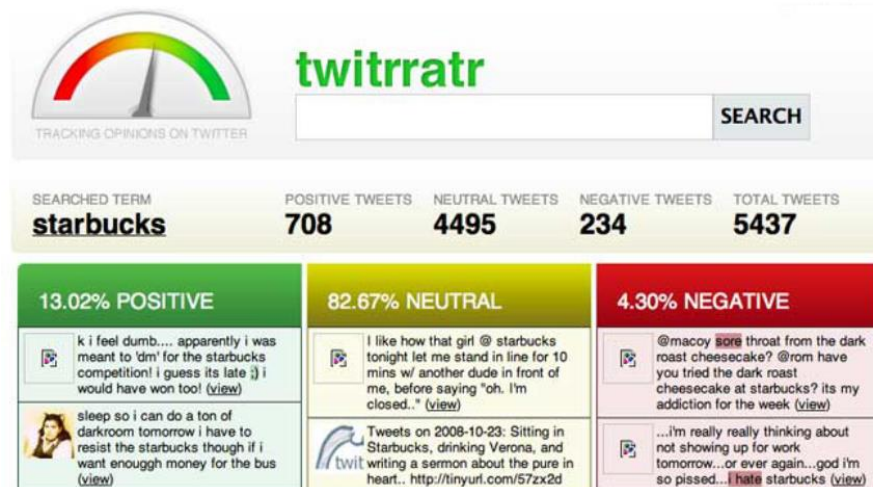
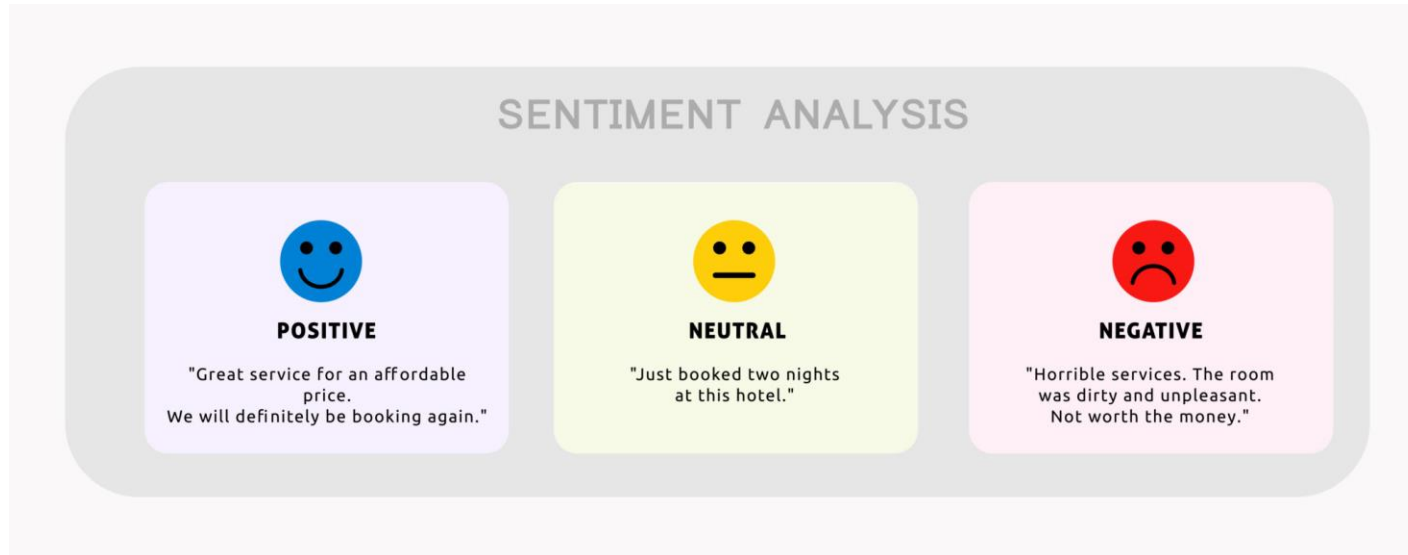


Downloading

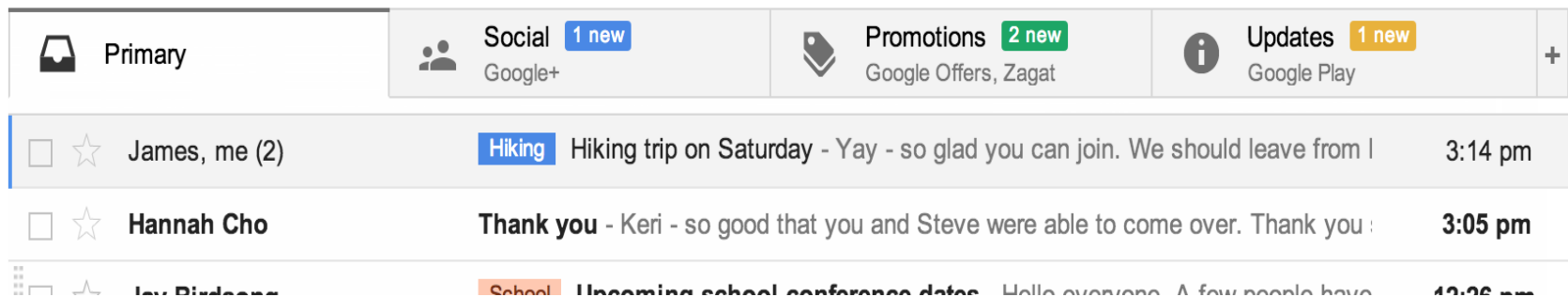
Downloading content from Wikipedia is free of charge. All text content is licensed under the [GNU Free Documentation License](#) (GFDL). Images and other files are available under [different terms](#), as detailed on



Sentiment/Opinion Analysis



Text Classification



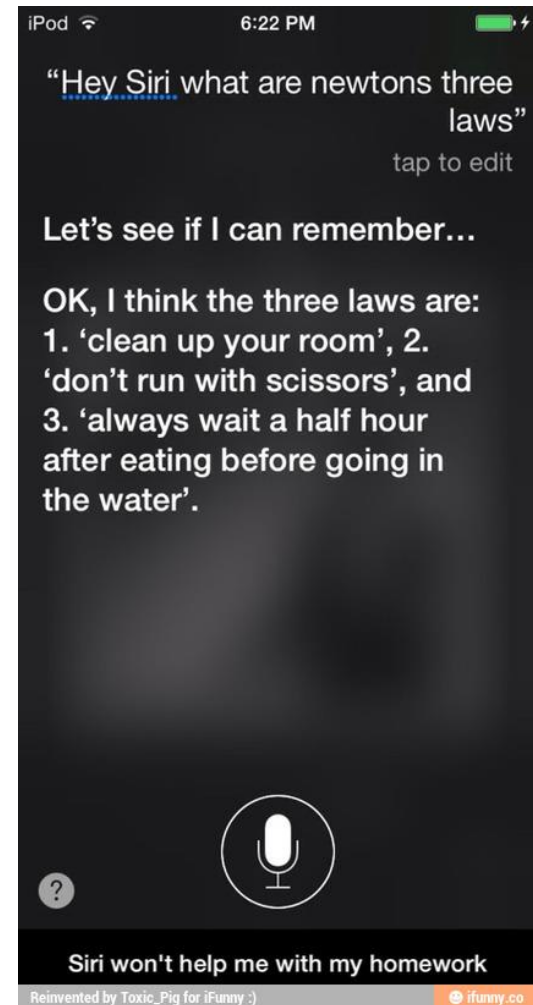
Question answering



'Watson' computer wins at 'Jeopardy'

<https://www.youtube.com/watch?v=P18EdAKuC1U>

The whole project was about doing research and deeper analytics in natural language understanding and creating technology to solve problems which humans really cared about.



Reinvented by Toxic_Pig for iFunny :)

ifunny.co

credit: ifunny.com

Question answering


- Go beyond search

Google


what is the capital of nagaland

All Images Maps News Videos Shopping Web More Tools



These are results for **what is the capital of nagaland**
Search instead for **what is the capital of nagaland**

 Nagaland > Capital :

Kohima
Executive Branch



KOHIMA
नागालैंड की राजधानी


Mizoram
Aizawl

Manipur
Imphal

Meghalaya
Shillong

मराठी मध्ये In English

Kohima, situated in the south at an altitude of 1444m above sea level, occupies pride of place as the capital city of Nagaland.

 **Kohima District**
<https://kohima.nic.in/history>

[History | District Kohima, Government of Nagaland | India](#)

People also ask :

What is capital of Nagaland?

What is the capital of Dimapur?

What is the old name of Nagaland?

Kohima
City in Nagaland

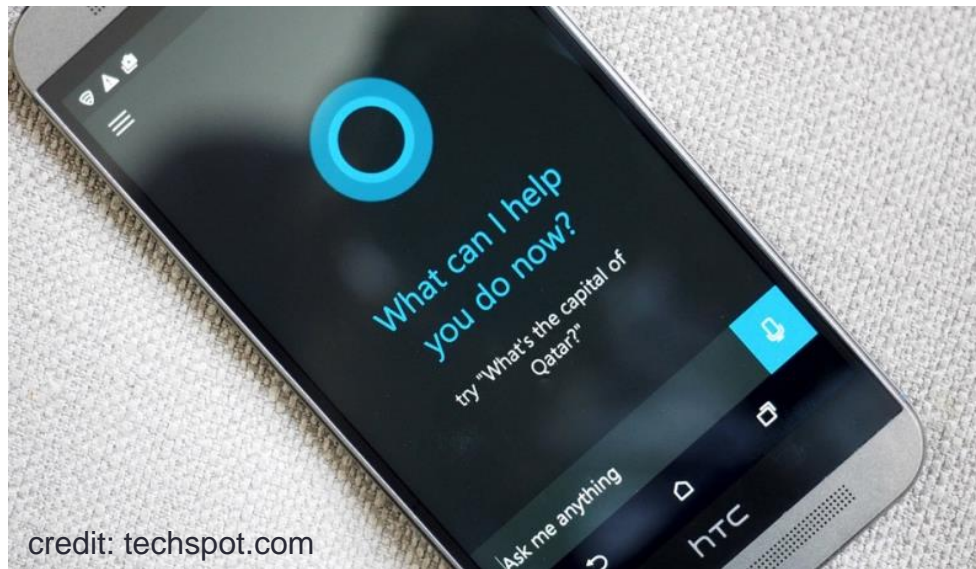
Kohima is the capital of the North East Indian state of Nagaland. With a resident population of almost 100,000, it is the second largest city in the state. Kohima constitutes both a district and a municipality. The municipality covers 20 km². [Wikipedia](#)

Elevation: 1,444 m

Natural language instruction



Digital personal assistant



- Semantic parsing – understand tasks
- Entity linking – “my wife” = “Kellie” in the phone book

Information Extraction

- Unstructured text to database entries

New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

Person	Company	Post	State
Russell T. Lewis	New York Times newspaper	president and general manager	start
Russell T. Lewis	New York Times newspaper	executive vice president	end
Lance R. Primis	New York Times Co.	president and CEO	start

Yoav Artzi: Natural language processing

Language Comprehension

Christopher Robin is alive and well. **He** is the same person that you read about in the book, **Winnie the Pooh**. As **a boy**, **Chris** lived in a pretty home called **Cotchfield Farm**. When **Chris** was three years old, **his father** wrote a poem about **him**. The poem was printed in a magazine for others to read. **Mr. Robin** then wrote a book

- Q: who wrote Winnie the Pooh?
- Q: where did Chris lived?

HISTORY OF NLP

History of NLP

- First Phase (Machine Translation Phase) - Late 1940s to late 1960s
- Second Phase (AI Influenced Phase) – Late 1960s to late 1970s
- Third Phase (Grammatico-logical Phase) – Late 1970s to late 1980s
- Fourth Phase (Lexical & Corpus Phase) – The 1990s

The Machine Translation Phase

- The research on NLP started in early 1950s after Booth & Richens' investigation and Weaver's memorandum on machine translation in 1949.
- 1954 was the year when a limited experiment on automatic translation from Russian to English demonstrated in the Georgetown-IBM experiment.
- In the same year, the publication of the journal MT (Machine Translation) started.
- The first international conference on Machine Translation (MT) was held in 1952 and second was held in 1956.
- In 1961, the work presented in Teddington International Conference on Machine Translation of Languages and Applied Language analysis was the high point of this phase.

AI Influenced Phase

- In early 1961, the work began on the problems of addressing and constructing data or knowledge base. This work was influenced by AI.
- In the same year, a **BASEBALL question-answering system** was also developed. The input to this system was restricted and the language processing involved was a simple one.
- A much advanced system was described in Minsky (1968). This system, when compared to the BASEBALL question-answering system, was recognized and provided for the need of inference on the knowledge base in interpreting and responding to language input.
- Eliza – a mock psychotherapist
- <https://web.njit.edu/~ronkowitz/eliza.html>
- <https://www.masswerk.at/elizabot/>

Grammatico-logical Phase

- The grammatico-logical approach, towards the end of decade, helped us with powerful general-purpose sentence processors like SRI's Core Language Engine and Discourse Representation Theory, which offered a means of tackling more extended discourse.
- In this phase we got some practical resources & tools like parsers,
 - e.g. Alvey Natural Language Tools along with more operational and commercial systems,
 - e.g. for database query.
- The work on lexicon in 1980s also pointed in the direction of grammatico-logical approach.

Lexical & Corpus Phase

- The phase had a lexicalized approach to grammar that appeared in late 1980s and became an increasing influence.
- There was a revolution in natural language processing in this decade with the introduction of machine learning algorithms for language processing.

GENERIC NLP SYSTEMS

Generic NLP System

- ELIZA
- SysTran
- TAUM METEO
- SHRDLU
- LUNAR

Generic NLP System

- ELIZA

- ELIZA is an early natural language processing computer program created from 1964 to 1966 at the MIT Artificial Intelligence Laboratory by Joseph Weizenbaum.
- Created to demonstrate the superficiality of communication between humans and machines, Eliza simulated conversation by using a "pattern matching" and substitution methodology that gave users an illusion of understanding on the part of the program, but had no built in framework for contextualizing events
- <https://web.njit.edu/~ronkowit/eliza.html>
- This program is one of the most popular AI programs of all time since it carried the keyword based analysis approach to an absurd (but impressive) extreme. The success of this simple technique in this domain is due to the fact that ELIZA doesn't actually need to know anything to mimic a psychotherapist. As Weizenbaum notes, this is one of the few dialogue genres where the listener can act as if they know nothing of the world. ELIZA's deep relevance to Turing's ideas is that many people who interacted with ELIZA came to believe that it really understood them and their problems.
- Indeed, Weizenbaum (1976) notes that many of these people continued to believe in ELIZA's abilities even after the program's operation was explained to them.

ELIZA System

Consider the following session with a version of ELIZA that limited the responses of a psychotherapist.

User 1:

You are like my brother in some ways.

ELIZA 1:

WHAT RESEMBLANCE DO YOU SEE?

User 2:

You are not very intelligent but I think you don't want me to notice that.

ELIZA 2:

WHAT MAKES YOU THINK I AM NOT INTELLIGENT?

User 3:

You don't agree with me.

ELIZA 3:

WHY DO YOU THINK I DON'T AGREE WITH YOU?

User 4:

You are afraid of me.

ELIZA 4:

DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU?

User 5:

I need some help, that much seems certain.

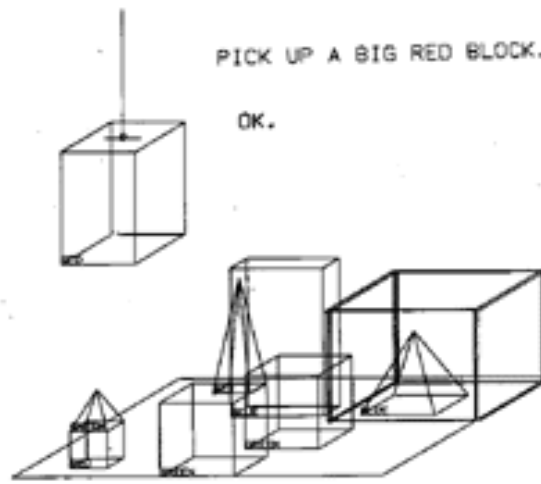
ELIZA 5:

WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP?

- SysTran
 - SYSTRAN, founded by Dr. Peter Toma in 1968 is one of the oldest machine translation companies. SYSTRAN has done extensive work for the United States Department of Defense and the European Commission.
- TAUM METEO
 - The METEO System is a machine translation system specifically designed for the translation of the weather forecasts issued daily by Environment Canada.
 - The system was used from 1981 to 30 September 2001 by Environment Canada to translate forecasts issued in French in the province of Quebec into English and those issued in English in other Canadian provinces into French.
 - Since then, a competitor program has replaced METEO System after an open governmental bid.

SHRDLU System

- SHRDLU was an early natural-language understanding computer program, developed by Terry Winograd at MIT in 1968–1970.
- SHRDLU had a simple conversation with a user via teletype about a small world of objects, the BLOCKS world, shown on an early computer screen.
- The user interacts with the application by moving objects, naming collections, and asking questions about the condition of a condensed "blocks world," a virtual box filled with various blocks.
- It is a simulation based programming system involving of a hand and eye. It is a syntax based system which is a combination of deep and surface structure.
- It contains a syntactic parser with a fairly wide coverage which builds surface structures that are not simply of trivial category labeling.
- It performs the combination and integration of many components which will create a total system. For example:



The dialog that was used as a SHRDLU demo:

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.
Person: CAN A PYRAMID SUPPORT A PYRAMID?
Computer: I DON'T KNOW.
Person: STACK UP TWO PYRAMIDS.
Computer: I CAN'T.
Person: WILL YOU PLEASE STACK UP BOTH OF THE RED BLOCKS AND EITHER A GREEN CUBE OR A PYRAMID?
Computer: OK. (does it)
Person: WHICH CUBE IS SITTING ON THE TABLE?
Computer: THE LARGE GREEN ONE WHICH SUPPORTS THE RED PYRAMID.
Person: IS THERE A LARGE BLOCK BEHIND A PYRAMID?
Computer: YES, THREE OF THEM: A LARGE RED ONE, A LARGE GREEN CUBE, AND THE BLUE ONE.
Person: PUT A SMALL ONE ONTO THE GREEN CUBE WHICH SUPPORTS A PYRAMID.
Computer: OK. (does it)
Person: PUT THE LITTLEST PYRAMID ON TOP OF IT.
Computer: OK. (does it)
Person: DOES THE SHORTEST THING THE TALLEST PYRAMID'S SUPPORT SUPPORTS SUPPORT ANYTHING GREEN?
Computer: YES, THE GREEN PYRAMID.
Person: WHAT COLOR IS IT?
Computer: BY "IT", I ASSUME YOU MEAN THE SHORTEST THING THE TALLEST PYRAMID'S SUPPORT SUPPORTS.
Computer: RED

LUNAR System

- It was developed by Woods in 1970.
- It is one of the largest and most successful question-answering system using AI techniques.
- This system had a separate syntax analyzer and a semantic interpreter. Its parser was written in ATN (Augmented Transition Network) form.
- The system was used in various tests and responded successfully to queries like followings:
 - How many oak trees have height greater than 15 inches?
 - What is the average concentration of hydrogen and oxygen in water?
 - Which one is the oldest material between Iron, Bauxite and Aluminum?
- The LUNAR system mainly deal with queries. But the performance of the system is very good than other systems.

HAL System

- HAL is an artificial agent capable of such advanced language processing behaviour as speaking and understanding English.
- The HAL system was developed by Arthur C. Clarke.
- Generally HAL system is useful for **language and speech recognition**.
- By speech and language processing we have in mind those computational techniques that process spoken and written human language.
- HAL require much broader and deeper knowledge of language.
- To determine what the user is saying, HAL must be capable of analyzing an incoming audio signal and recovering the exact sequence of words user used to produce that signal.
- Similarly, in generating its response HAL must be able to take a sequence of words and generate an audio signal that the user can recognize.
- Both of these tasks require knowledge about **phonetics and phonology** which can help model how words are pronounced in colloquial speech.

KNOWLEDGE IN SPEECH AND LANGUAGE PROCESSING

Aunty: Aaj toh tum saari Dilli mein aag lagaa dogi

SIFRA: Kya aap chahti ho ki main saari Dilli mein aag lagaa doon?

SIFRA - an agent would have to be capable of interacting with humans via language, which includes

- ✓ **understanding** humans via **speech recognition** and **natural language understanding** (and of course **lip-reading**), and
- ✓ **communicating** with humans via **natural language generation** and **speech synthesis**.
- ✓ **able** to do **information retrieval** (finding out where needed textual resources reside),
- ✓ **able to do information extraction** (extracting pertinent facts from those textual resources), and
- ✓ **able to draw inference** (drawing conclusions based on known facts).

What an NLP agent needs?

- knowledge about **phonetics and phonology**, which can help model how words are pronounced in colloquial speech
- knowledge about **morphology**, which captures information about the shape and behavior of words in context
- The knowledge needed to order and group words together comes under the heading of **syntax**
- knowledge of the meanings of the component words, the domain of **lexical semantics**
- The appropriate use of this kind of polite and indirect language comes under the heading of **pragmatics**
- Correctly structuring these such conversations requires knowledge of **discourse conventions**

Levels of NLP

- Phonology level
- Morphological level
- Lexical level
- Syntactic level
- Semantic level
- Discourse level
- Pragmatic level

Levels of NLP

- **Phonology level** –
 - deals with pronunciation
 - English is partially phonetics
 - E.g John inputs the data
- **Morphological level** –
 - Deals with smallest part of the word that convey meaning and suffixes/prefixes
 - Morphemes – studying how the words are built from smaller meaning
 - Dog – single morpheme
 - Dogs – two morphemes : ‘dog’ and ‘s’
- **Lexical level** –
 - Deals with respect to lexical meaning and Part-of-Speech.
 - Lexeme – basic unit of lexical meaning
 - It represents the set of forms or “senses” taken by morpheme
 - E.g “duck” – can be noun or verb. Depends on the context in the phrase/sentence
- **Syntactic level** –
 - Deals with grammar and structure of sentences.
 - Also called as parsing
 - Allows extraction of phrases which convey more meaning than just the individual words by themselves

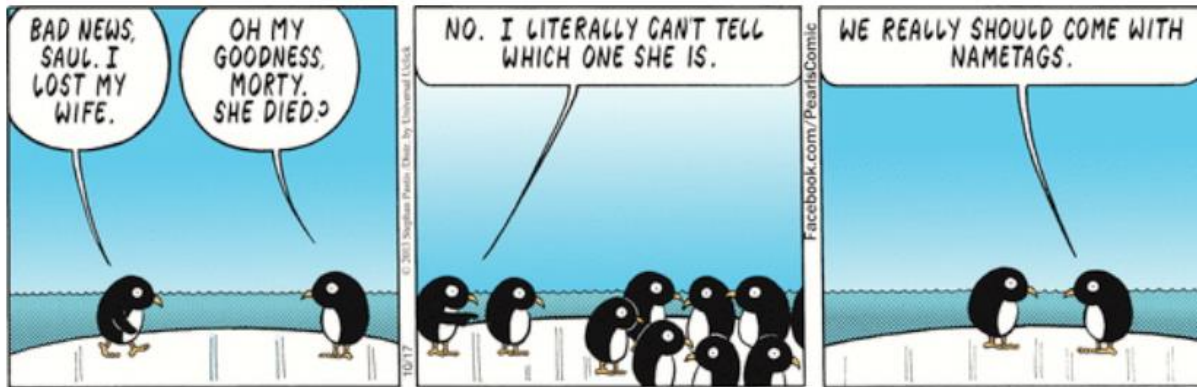
Levels of NLP

- Semantic level –
 - Deals with the meaning of words and sentences
 - Two approaches
 - Syntax-driven semantic analysis
 - Semantic grammar
- Discourse level –
 - Deals with the structure of different kinds of text.
 - Two types of discourse
 - Anaphora resolution
 - Discourse/text structure recognition
- Pragmatic level –
 - Deals with the use of real world knowledge
 - understanding of how this influences the meaning of what is being communicated.

CHALLENGES

Challenges – ambiguity

- Word sense ambiguity



credit: A. Zwicky



Challenges – ambiguity

- Word sense / meaning ambiguity



Credit: <http://stuffsirisaidth.com>

Challenges – ambiguity

- PP attachment ambiguity

San Jose cops kill man with knife

Text

Paper

Translate

Listen

Close

San Jose cops kill man with knife

Ex-college football player, 23, shot 9 times allegedly charged police at fiancée's home

By Hamed Aleaziz
and Vivian Ho

A man fatally shot by San Jose police officers while allegedly charging at them with a knife was a 23-year-old former football player at De Anza College in Cupertino who was distraught and depressed, his family said

Thursday.

Police officials said two officers opened fire Wednesday afternoon on Phillip Watkins outside his fiancée's home because they feared for their lives. The officers had been drawn to the home, officials said, by a 911 call reporting an armed home invasion

that, it turned out, had been made by Watkins himself.

But the mother of Watkins' fiancée, who also lives in the home on the 1300 block of Sherman Street, said she witnessed the shooting and described it as excessive. Faye Buchanan said the confrontation happened

shortly after she called a suicide intervention hotline in hopes of getting Watkins medical help.

Watkins' 911 call came in at 5:01 p.m., said Sgt. Heather Randol, a San Jose police spokeswoman. "The caller stated there was a male breaking into his home armed with a knife," Randol said. "The caller also stated he was locked in an upstairs bedroom with his children and request-

ed help from police."

She said Watkins was on the sidewalk in front of the home when two officers got there. He was holding a knife with a 4-inch blade and ran toward the officers in a threatening manner, Randol said.

"Both officers ordered the suspect to stop and drop the knife," Randol said. "The suspect continued to charge the officers with the knife in his hand. Both officers, fear-

ing for their safety and defense of their life, fired at the suspect."

On the police radio, one officer said, "We have a male with a knife. He's walking toward us."

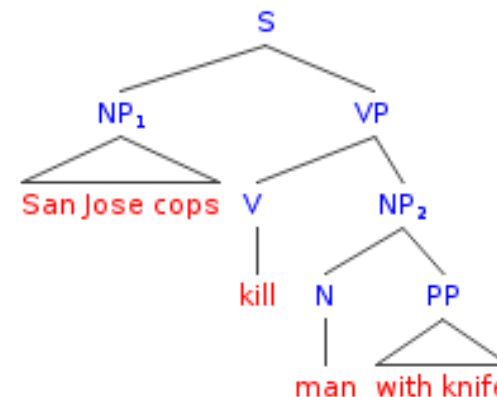
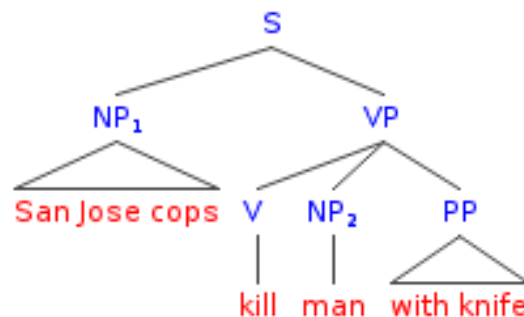
"Shots fired! Shots fired!" an officer said moments later.

A short time later, an officer reported, "Male is down. Knife's still in hand."

Buchanan said she had been prompted to call the

Shoot continues on D8

Back Continue



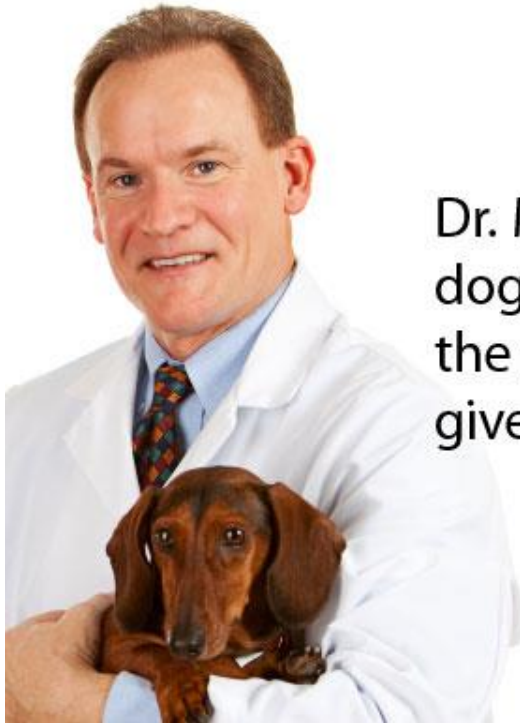
Challenges -- ambiguity

- Ambiguous headlines:
 - Include your children when baking cookies
 - Local High School Dropouts Cut in Half
 - Hospitals are Sued by 7 Foot Doctors
 - Iraqi Head Seeks Arms
- Safety Experts Say School Bus Passengers Should Be Belted
- Teacher Strikes Idle Kids



Challenges – ambiguity

- Pronoun reference ambiguity

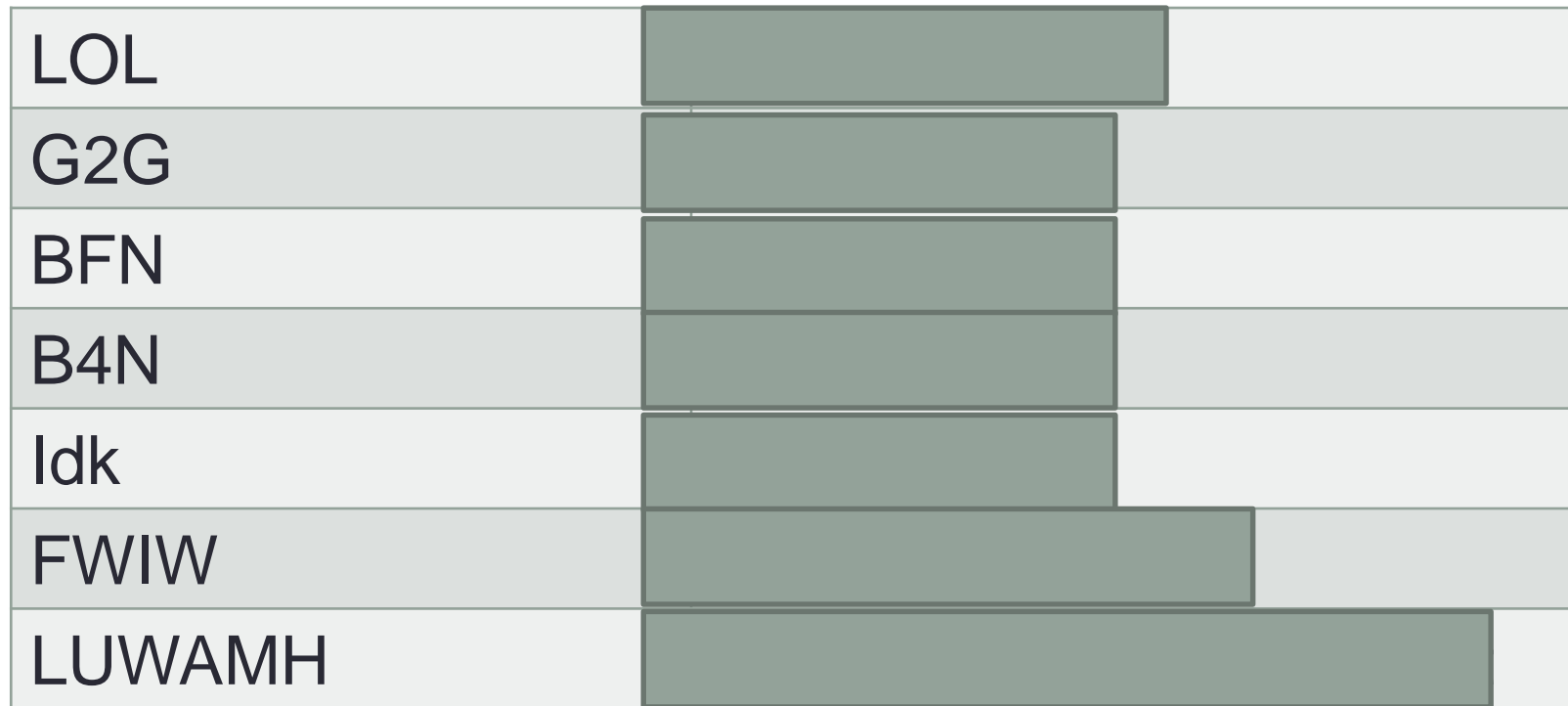


Dr. Macklin often brings his dog Champion to visit with the patients. **He** just loves to give big, wet, sloppy kisses!

Credit: <http://www.printwand.com/blog/8-catastrophic-examples-of-word-choice-mistakes>

Challenges – language is not static

- Language grows and changes
 - e.g., cyber lingo



Challenges--language is compositional



Carefully
Slide



Challenges--language is compositional

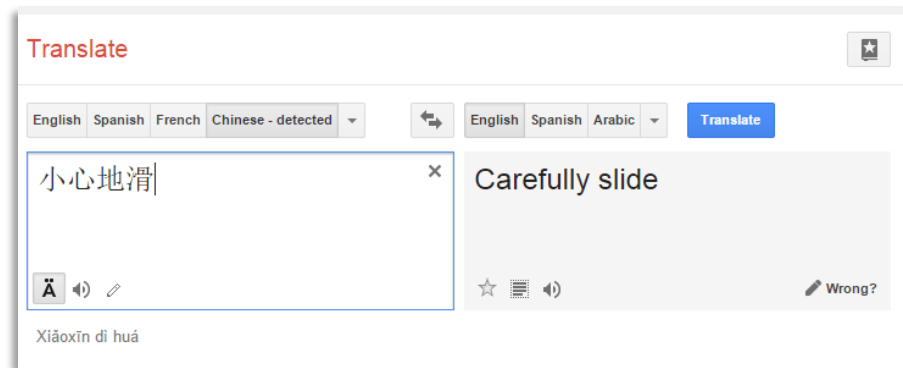


小心:

Carefully
Careful
Take
Care
Caution

地滑:

Slide
Landslip
Wet Floor
Smooth



Challenges – scale

- Examples:
 - Bible (King James version): ~700K
 - Penn Tree bank ~1M from Wall street journal
 - Newswire collection: 500M+
 - Wikipedia: 2.9 billion word (English)
 - Web: several billions of words

Challenges

1. Contextual words and phrases and homonyms

- Homonyms
 - Write and right
 - Their and there
- Words and phrases have different meaning in different context
 - e.g. 'run/ran'
 - I ran to the store as we ran out of milk
 - Can this program run on this machine
 - The house is looking really run down today

Challenges

2. Synonyms

- Words that convey same meaning
- Difficult for contextual understanding as we use many words to convey the same meaning
- Additionally, some of these words convey exactly the same meaning, while some may be levels of complexity and different people use synonyms to denote slightly different meanings
- E.g., Small, little, tiny, minute

Challenges

3. Irony and sarcasm

- words that strictly by definition are either positive or negative but truly mean the opposite
 - Yeah right, whatever!
 - “It’s okay if you don’t like me. Not everyone has good taste.”
 - “Unless your name is Google stop acting like you know everything.”
 - “If you think nobody cares if you’re alive, try missing a couple of loan installments.”

Challenges

4. Ambiguity and Uncertainty in Language

- Natural language is very ambiguous.
- Ambiguity, generally used in natural language processing, can be referred as the ability of being understood in more than one way.
- NLP has the following types of ambiguities –

Lexical Ambiguity

- The ambiguity of a single word is called lexical ambiguity.
- For example, treating the word **silver** as a noun, an adjective, or a verb.

Syntactic Ambiguity

- This kind of ambiguity occurs when a sentence is parsed in different ways. For example, the sentence “*I saw the girl on the beach with my binoculars*”.
- It is ambiguous whether the man saw the girl through the binoculars or did he see her carrying his binoculars.

Ambiguity and Uncertainty in Language

Semantic Ambiguity

- This kind of ambiguity occurs when the meaning of the words themselves can be misinterpreted.
- In other words, semantic ambiguity happens when a sentence contains an ambiguous word or phrase.
- For example, the sentence

“The car hit the pole while it was moving”

- is having semantic ambiguity because the interpretations can be

“The car, while moving, hit the pole”

and

“The car hit the pole while the pole was moving”.

Ambiguity and Uncertainty in Language

Anaphoric Ambiguity

- This kind of ambiguity arises due to the use of anaphora entities in discourse.
- Anaphora – use of expression whose interpretation depends on another expression in context (antecedent or postcedent)
- For example,

The horse ran up the hill. It was very steep. It soon got tired.

Here, the anaphoric reference of “it” in two situations cause ambiguity.

Pragmatic ambiguity

- Such kind of ambiguity refers to the situation where the context of a phrase gives it multiple interpretations.
- In simple words, we can say that pragmatic ambiguity arises when the statement is not specific.
- For example,

“I like you too”

can have multiple interpretations

I like you (just like you like me),

and

I like you (just like someone else does).

Challenges

- 5. Misspelled words
- 6. Idioms and slangs
- 7. Domain specific language
- 8. Low-resource languages