BSCS5002: **Introduction to Natural Language Processing**

Lecture 2 : History of NLP

Parameswari Krishnamurthy

Language Technologies Research Centre
IIIT-Hyderabad

param.krishna@iiit.ac.in

Introduction to NLP

- NLP is a field of artificial intelligence that focuses on the interaction between computers and humans using natural language.
- The goal is to enable computers to understand, interpret, and generate human language in a way that is both valuable and meaningful.
- NLP combines computational linguistics with machine learning, deep learning, and other AI techniques.

Early Beginnings:1950-1960

- Warren Weaver and his *Translation Memorandum (1949)*
- Foundation: His ideas were rooted in information theory, successes in code breaking during WWII
- **1950s:** Machine translation emerged as one of the first major applications, igniting significant interest in NLP.
- The initial systems were simplistic, primarily using dictionary lookups and basic word order rules.
- Georgetown–IBM Experiment (1954): First public demonstration of a machine translation system at IBM's New York headquarters.
- Translated 49 carefully selected Russian sentences into English, mainly in the field of chemistry
- In 1957, Noam Chomsky introduced the concept of generative grammar in his renowned book *Syntactic Structures.*

Rule-based MT

English-Spanish:

source: https://medium.com/nlplanet/a-brief-timeline-of-nlp-bc45b640f07d

1960-1970

- **1960s** - **1970s:** Continued development of rule-based systems like ELIZA, a program that mimicked a Rogerian psychotherapist.

```
Welcome to
                    EEEEEE LL      IIII  ZZZZZZ  AAAAA
                    EE     LL       II       ZZ  AA   AA
                    EEEEE  LL       II      ZZZ  AAAAAAA
                    EE     LL       II     ZZ    AA   AA
                    EEEEEE LLLLLL  IIII  ZZZZZZ  AA   AA

   Eliza is a mock Rogerian psychotherapist.
   The original program was described by Joseph Weizenbaum in 1966.
   This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```
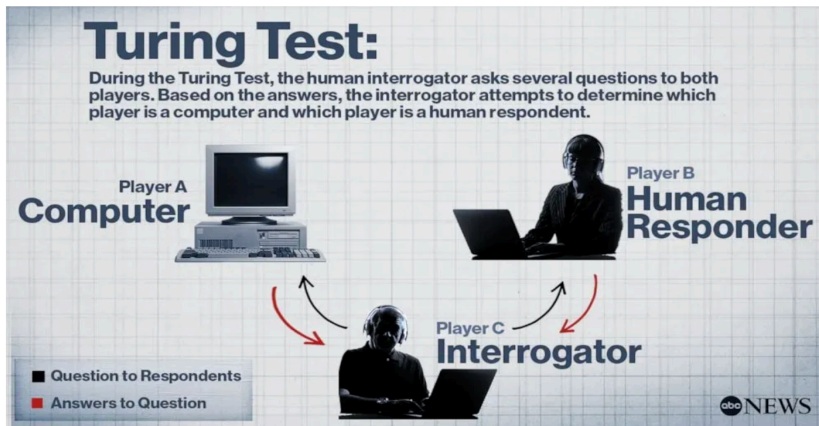
source: https://en.wikipedia.org/wiki/ELIZA

1960-1970

- Turing Test: A Benchmark for Machine Intelligence

Source: ABC News

1960-1970

- ALPAC Report (1966): Shifting Focus in Machine Translation
- US govt formed it to evaluate progress in computational linguistics.
- However, the limitations of these systems eventually led to the first **AI winter**, a period of reduced funding and interest.

Late 1970s

- Despite a slowdown in NLP research during the 1970s, significant advancements were made in computationally tractable theories of grammar.
- **Case Grammars:**
  - Focused on the role of noun phrases and their relationships to verbs.
  - Provided a framework for understanding sentence structure based on semantic roles.
- **Semantic Networks:**
  - Graph-based representations of knowledge.
  - Used to model relationships between concepts and entities, enhancing natural language understanding.
- **Conceptual Dependency Theory:**
  - Aimed at representing the meaning of sentences in a structured format.
  - Emphasized the importance of semantic relationships over syntactic structures.

NLP in the 1980s: Expert Systems

- Symbolic Approaches also known as **Expert Systems**, these approaches dominated NLP in the 1980s.
- Utilized **hard-coded rules** and **ontologies** (knowledge bases containing facts, concepts, and relationships within a domain).
- **Ontologies:**
    - Ontologies stored facts and relationships, essential for reasoning in expert systems.
    - For example, if the system knows that "All humans are mortal" and "Socrates is a human," it can infer that "Socrates is mortal."

Transition to Statistical Models: Late 1980s - Early 1990s

- **Shift from Symbolic to Statistical Models:**
  - Statistical models began to replace expert systems.
  - These models could **learn from data** rather than relying on manually coded rules.
- **Machine Learning:**
  - Statistical models utilized **machine learning** to learn patterns and rules automatically.
  - This marked a significant shift in NLP research and application.
- **Advances in Computational Resources:**
  - Increased computational power in the late 1980s and 1990s enabled the training of more complex models.
  - Facilitated the training of the first **Recurrent Neural Networks (RNNs)**.

# Neural Networks in the 2000s

- **Increased Use of Neural Networks:**
  - Initially applied for **language modeling**.
  - Focused on predicting the next word in a sequence based on previous words.
- **Introduction of Word Embeddings:**
  - Represented words as **dense vectors of numbers**.
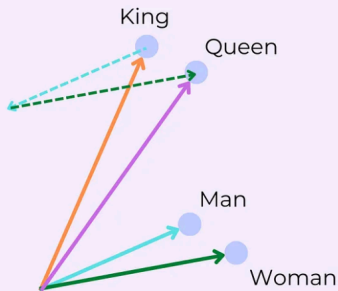 - Words with similar meanings are mapped to similar vector representations.

Graphical Representation of Vectors

Mathematics

Statistics
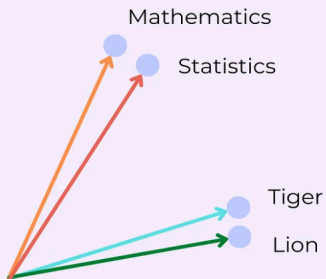
Tiger

Lion

- "Mathematics" is similar to "Statistics"
- "Tiger" is similar to "Lion"

Graphical Representation of Vectors

$$\text{vector(king)} - \text{vector(man)} + \text{vector(woman)} \approx \text{vector(queen)}$$

Challenges and Innovations

- Struggled to efficiently learn word representations.
- Often trained on small text corpora, leading to **suboptimal embeddings**.
- Google Translate: Released in 2006 as the first commercially successful NLP system.
- Utilized statistical models for automatic document translation.

Impact of Word Embeddings

- Pre-trained Embeddings:
  - Using pre-trained embeddings as features improved performance across various NLP tasks.
  - Enabled better encapsulation of text meaning.
- Common Neural Networks:
  - Dominant architectures included **LSTM RNNs** and **Convolutional Neural Networks (CNNs)**.
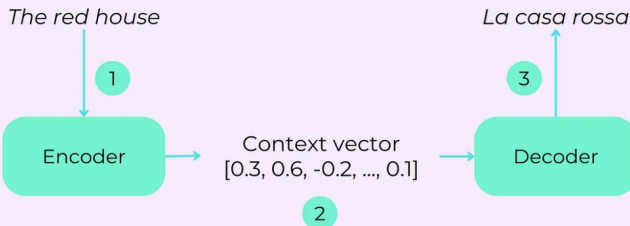
Encoder-Decoder Model: 2014

- General formalization of **sequence-to-sequence** problems, crucial for tasks like machine translation.
- **Encoder:** Encodes input into a context vector.
- **Decoder:** Decodes the context vector into an output sequence.

Encoder-Decoder based MT

English-Spanish:

source: https://medium.com/nlplanet/a-brief-timeline-of-nlp-bc45b640f07d

Success of Transformers and Attention Mechanisms

- Attention Mechanisms: Presented in the landmark paper "Attention Is All You Need."
- Revolutionized NLP by eliminating recurrent connections and relying solely on attention mechanisms.
- Capable of capturing long-range dependencies and context efficiently.
- Facilitated training on large datasets, leading to better performance across NLP tasks.
- Enhance information flow between encoder and decoder.
- Improved performance of sequence-to-sequence models.

Advancements in Pre-trained Language Models

- Training on Large Datasets:
  - Transformers trained on vast amounts of internet text in a self-supervised manner.
  - Led to the development of powerful pre-trained models.
- Fine-tuning:
  - Pre-trained models can be adapted to various tasks with minimal additional training (fine-tuning).
  - Enables quick application to new tasks and domains.

History of NLP

Conclusion

- NLP has evolved significantly from its early beginnings and continues to be a rapidly growing field.
- The 2000s to 2020s saw a significant transformation in NLP through the introduction of neural networks, word embeddings, and transformer models.
- These advancements enhanced the ability to understand and generate human language effectively.
- Ongoing trends include the development of increasingly larger language models that excel in a wide range of NLP tasks.
- The future may involve more advanced pre-trained models, better understanding of context, and more robust multilingual models.
- Ethical considerations, such as bias in NLP models, are becoming increasingly important.