Key Milestones in NLP

1948

The first NLP application can be considered a dictionary look-up system developed at Birkbeck College, London for automated translation. Post-World War II research focused on German-to-English translation, later shifting to Russian-to-English during the Cold War.

1957

Noam Chomsky's publication of "Syntactic Structures" revolutionized linguistics and significantly influenced NLP. His work informed the invention of Backus-Naur Form (BNF) notation in 1963 for representing programming language syntax and Regular Expressions in 1956 for specifying text search patterns.

1966

The ALPAC Report highlighted the limited success of machine translation, leading to a funding drought until 1980. Despite this, progress was made in areas like case grammar and semantic representations, although much of the work until the late 1960s focused on syntax.

1970

NLP shifted towards AI, emphasizing world knowledge and meaningful representations, increasing the importance of semantics. Systems like SHRDLU (1973) and LUNAR (1978) exemplified this trend, leading to the adoption of logic for knowledge representation and reasoning in the 1980s. The Prolog programming language, invented in 1970, also found applications in NLP.

1980

The growing adoption of Machine Learning marked the birth of statistical NLP. Annotated text corpora became the gold standard for evaluating ML models. ML approaches to NLP gained prominence throughout the 1990s, partly due to the success of Hidden Markov Models in speech recognition. The increased success of statistical methods over purely linguistic approaches is famously summarized by Fred Jelinek's quote: "Every time I fire a linguist, the performance of our speech recognition system goes up."

1982

Project Jabberwacky, aimed at simulating natural human conversations to pass the Turing Test, marked the beginning of chatbot development. It won third place in the Loebner Prize in October 2003.

1998

The FrameNet project was introduced, focusing on semantic role modeling—a form of shallow semantic parsing that remains an area of active research.

2001

Researchers proposed using feed-forward neural networks with vector inputs (word embeddings) for language modeling, replacing the classical N-gram model. This paved the way for the use of RNNs (2010) and LSTMs (2013) for language modeling.

2003

Latent Dirichlet Allocation (LDA) was invented and became a standard method for topic modeling in machine learning.

2013

Improvements in word embeddings and efficient implementations like Word2vec facilitated wider adoption of neural networks for NLP. RNNs and LSTMs became popular due to their ability to handle dynamic input sequences, while CNNs from computer vision were adapted for NLP due to their parallelizability. Recursive Neural Networks were also explored to leverage the hierarchical structure of language.

March 2016

Microsoft launched Tay, a chatbot on Twitter, which was shut down within 16 hours due to its adoption of racist and abusive language. Microsoft later launched the Zo chatbot.

September 2016

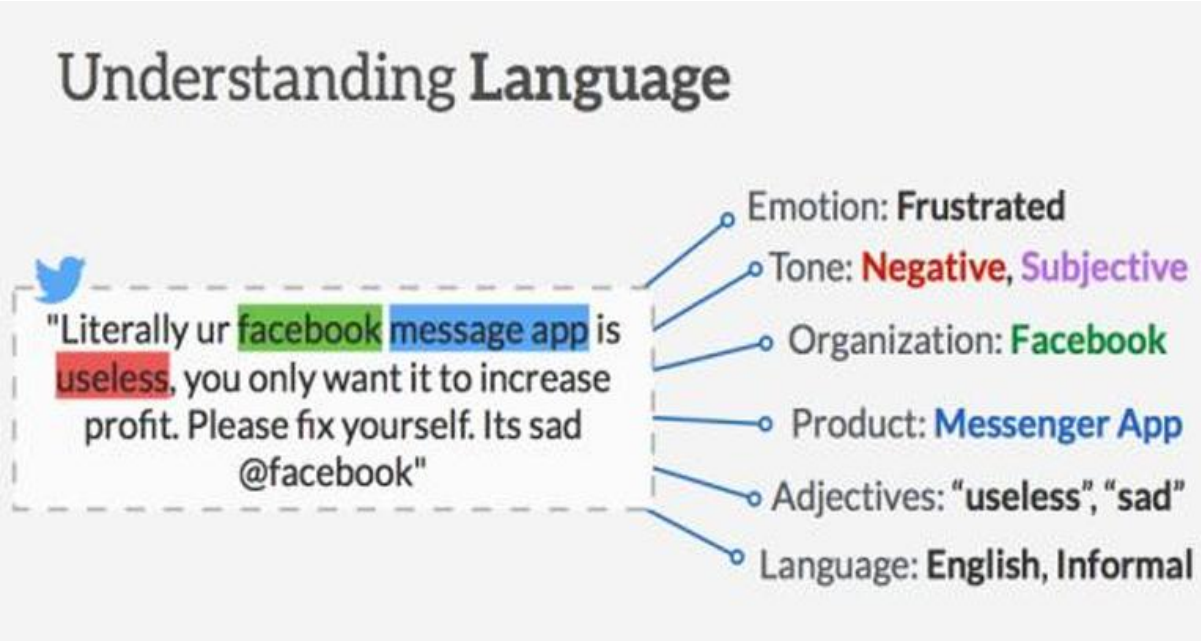
Google replaced its phrase-based translation system with Neural Machine Translation (NMT) using a deep LSTM network, reducing translation errors by 60%. This built upon sequence-to-sequence learning (proposed in 2014), which became a preferred technique for Natural Language Generation (NLG).

NLU vs NLG Comparison

NLU vs NLG

NLU	NLG
1. NLU is taking some spoken / typed sentence and working out what it means.	1. NLG is taking some formal representation of what you want to say & working out a way to express it in a natural language.
2. In NLU the system needs to disambiguate the input sentence to produce the machine representation language	2. In NLG the system needs to make decisions about how to put a concept into words.
3. Different levels of analysis required: morpological analysis, syntachic analysis, semantic analysis, discourse analysis.	3. Different levels of synthesis required: deep learning (what to say), syntachic generation.
4. NLU is most harder than NLG.	4. NLG is less harder than NLU.

Why do computers have difficulty with NLP?



1. Computers traditionally handle structured data (organized, indexed, and referenced, often in databases).
2. NLP deals with unstructured data like social media posts, news articles, emails, and reviews.

3. NLP must learn the structure and grammar of natural language (80% of enterprise data is unstructured).
4. Human language has complexities: ambiguous phrases, colloquialisms, metaphors, puns, sarcasm.
5. Words and text can have multiple meanings depending on context.
6. Language evolves, and imperfect communication (spelling, grammar, punctuation errors) is common yet understandable to humans.
7. Ambiguities are lexical, syntactic, or referential.
8. Speech adds challenges: accent, tone, loudness, noise, pronunciation, emotion, pauses.

Examples of English Complexities:

1. "One morning I shot an elephant in my pajamas"— ambiguity about who wore the pajamas.
2. "Listening to loud music slowly gives me a headache"— ambiguity about what happened slowly.
3. "The complex houses married and single soldiers and their families"— "complex" can be a noun or adjective, requiring part-of-speech tagging.
4. "John had a card for Helga, but couldn't deliver it because he was in her way"— coreference resolution needed to understand "he".
5. "The Kiwis won the match"— requires contextual knowledge of the nickname "Kiwis" for New Zealanders.