#### **What is Natural Language Processing?**

In the 1950s, Actually, Natural Language Processing (NLP) began as the incision of linguistics and artificial intelligence. Originally NLP was distinct from information retrieval (IR) of text, which engages highly with scalable statistics which is based on techniques to identify and index huge amounts of text expertly: Manning et. al [1], introduced a recent definition of Information Retrieval (IR). However, Natural Language Processing and Information Retrieval have worked with the same concept, with time. Recently, NLP covers not only many several and diverse fields but also NLP researchers and developers think broadly to gain knowledge-base efficiently.

Early simplistic approaches, for example, word-for-word Russian-to-English machine translation,2 were defeated by *homographs*—identically spelled words with multiple meanings—and metaphor, leading to the apocryphal story of the Biblical, ‘the spirit is willing, but the flesh is weak’ being translated to ‘the vodka is agreeable, but the meat is spoiled.’

Natural language's vastly large size, unrestrictive nature, and ambiguity led to two problems when using standard parsing approaches that relied purely on symbolic, hand-crafted rules:

* NLP must ultimately extract meaning (‘semantics’) from the text: formal grammars that specify the relationship between text units—parts of speech such as nouns, verbs, and adjectives—address syntax primarily. One can extend grammars to address natural-language semantics by greatly expanding sub-categorization, with additional rules/constraints (eg, ‘eat’ applies only to ingestible-item nouns). Unfortunately, the rules may now become unmanageably numerous, often interacting unpredictably, with more frequent *ambiguous parses* (multiple interpretations of a word sequence are possible). (Puns—ambiguous parses used for humorous effect—antedate NLP.)
* Handwritten rules handle ‘ungrammatical’ spoken prose and (in medical contexts) the highly telegraphic prose of in-hospital progress notes very poorly, although such prose is human-comprehensible.

#### **Why a Survey on Natural Language Processing?**

#### **Goals of this Survey?**

#### **NLP Areas:**

We enumerate common sub-problems in NLP: Jurafksy and Martin's text20 provides additional details. The solutions to some sub-problems have become workable and affordable, if imperfect—for example, speech synthesis (desktop operating systems' accessibility features) and connected-speech recognition (several commercial systems). Others, such as question answering, remain difficult.

In the account below, we mention clinical-context issues that complicate certain sub-problems, citing recent biomedical NLP work against each where appropriate.

* Anaphora
* Chatbot
* Character Annotation
* Digit Recognition
* Grapheme
* Language Machine Translation
* Lemmatizing
* Morphological Analysis
* Named Entity Recognition
* Optical Character Recognition
* Parser
* Pattern Recognition
* Parts of Speech Tagging
* Question Answering System
* Semantic Analysis
* Sentiment Analysis
* Sign Language Recognition
* Speech to Text
* Stemmer
* Syntax Analysis
* Text Classification
* Text Detection
* Text Generation
* Text Summarization
* Text to Speech
* Word Embedding