Elizabeth Martinez
Elijah Ghaya
Yammine Yammine
Annette Bazan

A05 Analyzing "Arrival" Through the Lens of NLP

The movie *Arrival* (2016), directed by Denis Villeneuve, revolves around linguistic challenges in communicating with an alien species. It provides several parallels to Natural Language Processing (NLP) challenges, including ambiguity, idiomatic expressions, sarcasm, and cultural variations.

There is a scene in which the word “weapon" is misinterpreted as something that could mean the Heptapods mean to cause harm. In this scene, the Heptapods use a logogram that the humans initially interpret as a "weapon," causing international tensions. Dr. Louise Banks (Amy Adams) argues that "weapon" could also mean "tool" depending on context. As an everyday example, a kitchen knife to many is just a tool to use in the kitchen. But it can also be used as a weapon. Ambiguity in language processing, where words have multiple meanings depending on context (e.g., "bank" as a financial institution vs. a riverbank). NLP models struggle with correctly disambiguating words without sufficient context. The alien language lacks direct word-for-word translations. Which is just like how words in human languages can have multiple meanings.

There’s also the use of idiomatic expressions and metaphors. In the scene where Louise explains how language influences thought using the Sapir-Whorf hypothesis. The challenge arises when trying to translate idioms or metaphors that don't have direct equivalents in another language. Many NLP models struggle with idiomatic expressions (e.g., "kick the bucket" meaning "to die"). Literal translations

do not always capture intended meanings, requiring advanced models trained on extensive contextual data. In today’s world, even different cultures that speak the same language will have some form of idiomatic expression that will differ and needs to be explained to someone who is not familiar with it. Dr. Banks starts with a structured rule-based approach like you would train a NLP by teaching(training data) simple words and symbols.

Another NLP challenge is with sarcasm and implicit meaning. While not explicitly sarcastic, Louise's approach to humor and negotiation in the film demonstrates the difficulty of conveying non-literal meaning across languages. Sarcasm detection is a major issue in sentiment analysis. A phrase like "Oh great, another traffic jam!" is negative, even though "great" is usually positive. NLP models need advanced contextual awareness to detect sarcasm.

Global conflict arises due to differences in how various countries interpret the Heptapods' language. The Chinese team interprets the message as a threat, while Louise sees it as an invitation. Regional and cultural variations affect sentiment and meaning. Words and phrases can mean different things in different dialects (e.g., "pants" in the U.S. vs. the U.K.). NLP models need localization to adapt to different linguistic variations. This is similar to how NLP systems struggle with regional dialects and cultural context.

The Heptapods' language is non-linear, meaning they perceive time differently. Louise eventually learns to think in this way, which affects her perception of time. This relates to sequence modeling in NLP. Traditional language models process text sequentially, while advanced models like Transformers (e.g., GPT, BERT) process text in parallel and capture dependencies over long sequences, much like the Heptapods' language structure.

Louise and Ian (Jeremy Renner) start with almost no knowledge of the alien language and gradually build understanding through iterative learning and pattern recognition. Similar to low-resource languages where there is limited training data, NLP models often struggle to develop robust understanding in such cases. Transfer learning and few-shot learning methods help mitigate this issue.

The Heptapods communicate through circular logograms rather than linear text, requiring an entirely new way of interpreting meaning. NLP increasingly deals with multimodal learning, where models must process text alongside images, speech, or symbols (e.g., combining text with visual elements in AI models like CLIP). Dr. Banks touches upon statistical NLP when she starts to recognize patterns, analyzing frequencies and structures which is similar to probabilistic models in NLP.

The movie has some relation to rule-based NLP. Dr. Louise Banks starts by identifying patterns in the Heptapods' circular symbols (logograms) and mapping them to known linguistic structures. She breaks down their symbols into components (like phonemes or morphemes in human languages) to form a rule-based understanding. Early communication follows structured principles, much like learning a new human language.

Rule-based NLP relies on predefined linguistic rules and dictionaries to process language. Similar to Louise's approach, early NLP systems used rule-based translation (e.g., early machine translation models that followed structured syntax rules). Tools like Prolog-based parsers or regular expression-based text processing mirror this symbolic approach. Finite State Transducers (FSTs) – Used in early language processing tasks, especially for morphology. Grammars (e.g., CFGs,

Dependency Parsing) – Used in syntactic parsing and speech recognition. Lexicon-based Sentiment Analysis – Uses predefined word lists to assign meaning to text.

There's also a reference to pattern recognition and frequent analysis. As more Heptapod symbols are collected, Ian Donnelly (Jeremy Renner) helps analyze the statistical relationships between symbols and their context. They begin to notice recurring structures, frequency distributions, and dependencies within Heptapod speech. The logograms are interpreted probabilistically, rather than through strict one-to-one mappings. Statistical NLP relies on probability and frequency analysis to model language. Early statistical methods like n-grams or Hidden Markov Models (HMMs) analyze text based on observed patterns in large datasets. In *Arrival*, the team uses a data-driven approach similar to corpus-based methods in NLP.

References were made to deep NLP learning. As Louise immerses herself in the Heptapod language, she undergoes a cognitive shift, learning to think non-linearly. The Heptapods' language encodes meaning holistically rather than sequentially. This mirrors how deep learning models go beyond simple frequency analysis and instead learn abstract representations of meaning. Deep NLP leverages neural networks to learn abstract language representations. Much like Louise's brain rewiring to process Heptapod language, deep learning models undergo representation learning to extract meaning beyond surface-level patterns. Transformers (BERT, GPT) model entire sentences in parallel rather than sequentially, akin to Heptapod communication.

There was also visual and contextual learning. The Heptapods' communication is both visual (logograms) and spoken (sounds and gestures). The

team records, categorizes, and analyzes both aspects simultaneously. This parallels how modern NLP systems integrate text, images, and speech for richer understanding. Multimodal NLP integrates different data types, such as vision (images), text, and speech. Much like the movie's approach, modern AI systems use multimodal learning to extract meaning beyond text alone. It seems in the film there is a bit of dependency parsing and large language models that may reflect the methods used to understand the communication between the two species.

Overall *Arrival* displays a close correlation on how AI uses NLP to be able to understand human language. Without background knowledge and a model or dataset to base on, the human language is nearly impossible to understand at a glance. Just as Louise Banks and her team had to analyze patterns, context, and structure to decode the Heptapod language, NLP models rely on training data and algorithms to interpret meaning. The film highlights the challenges of ambiguity, culture differences, and non-linear communication issues that modern AI continues to be challenged by. *Arrival* showcases the complexity of language understanding and reflects the ongoing advancements in NLP, which help bridge gaps in communication.