Fast Jet Learning Rockets (FJLR)

The Report

NLP and the film “Arrival”

**Table of Contents**

1. Introduction
2. NLP Challenges in “Arrival”

2.1. Ambiguity

2.2. Idiomatic Expressions and Metaphors

2.3. Sarcasm and Pragmatics

2.4. Cultural and Regional Variations

2.5. Data Sparsity and Few-Shot Learning

3. Analysis of Communication Methods and NLP Approaches

3.1. Analogous NLP Tools & Technologies

4. Reflection on Language and NLP in “Arrival”

4.1. Expanding NLP Understanding

5. Conclusion

1. Introduction

Language is one of the most complex aspects of human intelligence, and its study has significant implications for both linguistics and artificial intelligence. The 2016 film “Arrival”, directed by Denis Villeneuve, explores the challenges of understanding an entirely new language, providing a compelling parallel to the challenges faced in Natural Language Processing (NLP). This technical report examines how specific scenes in the movie reflect NLP challenges, analyzes the communication methods used by the protagonists, and relates them to real-world NLP approaches.

2. NLP Challenges in “Arrival”

The process of deciphering the Heptapod language in “Arrival” closely mirrors key challenges in NLP:

2.1. Ambiguity

In the film, the Heptapod symbols are highly abstract, and their meanings are initially unclear (timestamp 41:10). This reflects word sense disambiguation in NLP, where the same word can have multiple meanings depending on context. For example, the word “bank” can mean a financial institution or the side of a river, the word “light” can mean illumination or not heavy, etc. In the “Arival” the phrase “Use your weapon” was initially interpreted by humans as a threat, implying aggression or warfare. However, as Louise deciphers more of their language, she realizes that “weapon” in Heptapod language means “tool” or “ability” (timestamp 1:12), referring to the nonlinear perception of time that they are trying to share with humanity. Deep Learning NLP models, such as BERT, GPT, T5 and Word2Vec, use context to resolve such ambiguities.

Изображение выглядит как облако, туман, небо, на открытом воздухе

Контент, сгенерированный ИИ, может содержать ошибки.

2.2. Idiomatic Expressions and Metaphors

The Heptapod language does not operate on a one-to-one word translation basis but rather conveys meaning holistically. This is similar to NLP struggles with idiomatic expressions (e.g., “spill the beans” meaning “reveal a secret”, “break a leg” meaning “good luck and not literal injury, “the ball is in your court” meaning “it is your turn to decide, not a reference to a physical bal), which require models to understand cultural and contextual nuances. It is the major challenge for NLP systems when translating across languages.

A circle of life in the water

AI-generated content may be incorrect.

2.3. Sarcasm and Pragmatics

In the film, the Heptapods’ intent is initially misinterpreted, similar to how NLP models struggle with sarcasm and pragmatics. Sentiment analysis models often misinterpret sarcastic statements because they rely heavily on explicit sentiment words rather than deeper contextual meaning.

This misinterpretation of pragmatics shows once the Heptapods write down “Offer Weapon” (timestamp 1:06:30). This leads to a scene where they discuss and argue, one side saying it could be seen as a threat, demand, or form of aggression. While Dr. Banks explains that without context weapon can be seen as tool, they could be making a request or offering them technology. She says they need to go back and clear it up, such as how when an NLP model may be inaccurate or wrong, you need to retrain and “clear up” the meaning of these words within context.

A person wearing headphones

AI-generated content may be incorrect.

2.4. Cultural and Regional Variations

The film explores the Sapir-Whorf Hypothesis, suggesting that language shapes perception. Similarly, NLP must handle languages with different syntactic and semantic structures (e.g., English vs. Chinese). This challenge is particularly evident in machine translation systems, which must adapt to linguistic variations.

One example of how this challenge presents itself in the movie is when attempting to spy on General Shang, they use a satellite to listen in on his conversation (timestamp 1:03:30). They’re unable to fully understand what he’s saying through literal translation at first, as he mentions Suits, Honor, and Flowers and there’s little context within the sentence to go off of. This is until Dr. Banks realizes they’re tile sets in the game Mahjong, which could only be known if you’re aware and familiar with the culture.

A group of people in a dark room

AI-generated content may be incorrect.

2.5. Data Sparsity and Few-Shot Learning

At the start, Louise has no prior knowledge of the Heptapod language and must learn it with minimal data (timestamp 43:21). This is analogous to low-resource languages in NLP, where AI models struggle due to a lack of training data. Recent advances, such as zero-shot and few-shot learning, attempt to address this challenge.

When having difficulty translating the Heptapod’s language, Dr. Banks has to explain why they have to work slowly to prevent any misunderstandings (timestamp 39:50). She comes up with a story about the Aboriginal people when sailors arrived, when they pointed to the jumping animals with babies in their pouch the Aboriginal people said “Kangaroo”. With no knowledge of the language where kangaroo means “I don’t understand.” the sailors named that animal “Kangaroo”. NLP models can also face this challenge when they have very little data to work on, and have to make predictions based off minimal labels.



3. Analysis of Communication Methods and NLP Approaches

Louise, Ian, and their team use various techniques to decode the Heptapod language, many of which resemble real-world NLP approaches:

|  |  |  |
| --- | --- | --- |
| **Method in Arrival** | **NLP Approach** | **Description** |
| Visual analysis of symbols | Rule-Based NLP | Early stages involve manually mapping symbols to meanings, similar to rule-based NLP |
| Pattern recognition and frequency analysis | Statistical NLP | Identifying recurring symbols mirrors n-gram models and probabilistic language models |
| Iterative learning and prediction | Deep NLP | Louise refines her understanding iteratively, like how transformers (GPT, BERT) predict meaning based on previous data |
| Bidirectional interpretation | Contextual NLP | The Heptapods perceive time non-linearly, which resembles transformer-based models that process text with full context |

3.1. Analogous NLP Tools & Technologies

Several modern NLP technologies resemble the techniques used in “Arrival”:

• Word Embeddings (Word2Vec, GloVe) – Mapping meanings based on relationships between symbols.

• Neural Machine Translation (Google Translate, T5) – Predicting entire sentence structures instead of word-by-word translation.

• Explainable AI in NLP – Like Louise needing to justify her translations, modern NLP models require interpretability.

4. Reflection on Language and NLP in “Arrival”

The film challenges conventional views of language by presenting a system that is deeply tied to cognition and time perception. It raises thought-provoking questions about how language influences thought and how AI might one day overcome the limitations of human language processing.

4.1. Expanding NLP Understanding

• The Heptapod language suggests that meaning is not always linear, similar to how AI models now prioritize contextual embeddings rather than simple word mappings.

• The film highlights theimportance of multi-model NLP, as Louise relies on both visual and linguistic elements to decode meaning.

• Future NLP advancements may integrate cognitive models that adapt to different ways of thinking, similar to how Louise’s brain adapts to Heptapod communication.

5. Conclusion

The film “Arrival” provides a compelling case study for real-world NLP challenges, from ambiguity and idioms to machine learning approaches for language understanding. The film illustrates how the process of deciphering an alien language mirrors the complexities of NLP, from rule-based approaches to deep learning. As AI continues to evolve, the challenges highlighted in the film remain relevant, pushing the boundaries of language understanding in both human and machine communication.