Houston Community College

#### 

# "Arrival" Through the Lens of NLP

#### Analysis

Trevon Woods

Deep Learning - ITAI - 2376

Anna Devarakonda

07/05/2025

“Arrival” is a movie filled with complex language theory and brain tingling communication concepts that will leave you in awe. It is a movie that follows the life of Dr. Louise Banks, a linguist who is recruited by the military to establish communication with aliens called Heptapods. These aliens land 12 ships at key locations all over the world. They communicate through a complex, non-linear circular written language akin to hieroglyphs. They also communicate through a series of deep, resonant sounds. Through this report I will identify the NLP challenges, analyze the communication methods, as well as give my own reflection on the communication concepts in the movie.

To begin, let's dive into the NLP challenges and their real-world parallels. I’d like to jump straight to an analysis of a communication method used in the movie. I will continue my analysis of communication methods later in the report. In the movie Dr. Banks tried to find a basis for communication by teaching the aliens basic kindergarten words. She did this because in order to answer the main question the military wants answered, "What is your purpose on Earth"? She needs to first get the aliens to understand what a question is, which is the nature of a request for information along with a response. Then they need to break down the meaning behind specific words and their idiomatic nature like the word your. The word your in this context is ambiguous, because it uses a specific "you" but has a collective connotation. She states that she wants the aliens to understand that she isn't asking what a specific one of them is doing on earth but the collective. Then she dives into “purpose” which requires an understanding of intent. She states that they have to understand if the aliens make conscious decisions or if their motivations are so instinctive that they don't understand a "why" question at all. Also beyond that they needed to have a big enough vocabulary between them so that they can correctly interpret their answer.

A real-world parallel to this could be in how we as an advanced civilization try to make connections with indigenous peoples when new lands are discovered. For example, the civilizations that still live in the Amazon Rainforest. These civilizations do not understand our language or intentions so we have to find some kind of common ground. This is usually done through a gift of supplies, medicine, or food which helps to convey a trustworthy and trusting atmosphere between us and these indigenous people. From there we can trade knowledge, asking questions and finding differences and commonalities in our languages. Understanding their language helps to understand the underlying thinking and perception of these indigenous people as stated in the movie. Now a real-world NLP parallel would be rule-based systems that relied on handcrafted rules, context free grammars, and lexicons (Chomsky, 1957). These systems were pretty cut and dry, meaning they operated on an “if this, then that” premise for finding patterns and relationships.

Another challenge would be how the various governments, especially the Chinese, interpreted the Heptapods' ambiguous use of the word “weapon”. This hieroglyph when fully understood actually means “tool” or “gift”. This in a way shows how the lack of context and understanding can lead to major miscommunications. The real-world NLP parallel that would most fit here would be Word Sense Disambiguation. WSD is a Natural Language Processing task that involves determining the correct meaning of a word when it has multiple meanings in a given context. In the movie Dr. Banks asked the Colonel to ask her competition what the Sanskrit meaning of war was. Her competition, professor Gavisti, said it meant an argument. Dr. Banks said it meant a desire for more cows which is a more accurate interpretation. This shows how the misinterpretation of words can negatively impact how we perceive them. If the military had hired Gavisti instead of Dr. Banks, they would have immediately attacked the aliens the moment they heard “weapon”. Dr. Banks’ desire to dig deeper into the meaning of what the aliens meant by “weapon” saved the world from its primitive nature. Without proper context NLP models will struggle to extrapolate the correct meaning from the data. This can lead to errors in translation, sentiment analysis, or information retrieval.

The next challenge in the movie pertains to idiomatic expressions. The Heptapods’ hieroglyphs function similarly to human idioms but aren’t exactly the same. Each of the symbols they create encompasses a full thought or concept. In order to understand these symbols you would have to understand these aliens’ underlying philosophy and how they perceive time. The real-world NLP challenge that is similar to this would be phrases like “Kill two birds with one stone” which simply means to get more things done. NLP models cannot on their own understand that these words don’t actually mean to kill two birds with one stone (Manning & Schütze, 1999). They will take the literal translation of the individual words and fail to acquire the actual meaning. This is why NLP models are trained on vast Corporas with these idiomatic expressions in them. It enables them to learn these double meanings through statistical co-occurrence or deep learning.

For regional/cultural variational challenges, in this movie each nation interpreted messages of the aliens differently based on their own cultural biases, political agendas, and their limited data. This can be seen particularly in how the Chinese government reacted to the alien’s message. They became extremely concerned and aggressive purely based on how they perceived the danger through their cultural lense and military dogma. This shows how different national ideologies steer understanding. The real-world NLP equivalent is in how models trained on one domain are unable to produce correct outcomes when used in other domains. This is called domain adaptation (Blitzer et al., 2006). For example, an AI agent that is trained on cooking data to be a cooking assistant would do poorly in giving legal advice. Language use, vocabulary, jargon, and common expressions vary significantly across different fields, industries, or even regional dialects.

Now I'd like to continue my analysis of the communication methods and their equivalent NLP approaches. Next up would be Statistical NLP which is iterative learning and pattern recognition. In the movie, Dr. Banks gathers all of the symbols and tries to extract how they’re used in different ways. She starts to notice that certain parts of these circular hieroglyphs have smaller meanings that when combined create a larger meaning. They also use mathematical analysis to break down the symbols. In Statistical NLP approaches techniques like N-grams, Hidden Markov Models, and Latent Semantic Analysis analyze word frequencies, sequences, and statistical relationships to infer meaning and structure. This can be seen in how Machine Translation systems use statistical analysis to translate words to a different language.

Moving on to Deep NLP, which is a holistic understanding and uses contextual embeddings. This is shown in how Dr. Banks begins to grasp the whole of the alien’s language along with how it connects to their non-linear perception of time. She basically internalizes the language and the way the aliens experience reality. This internalization is what made her able to fully interpret the alien language and even see the future. Their language is basically an embedding for the alien’s entire world view. This is very similar to how models like Word2Vec, GloVe, and transformer models like BERT or GPT understand text (Goldberg, 2017; Vaswani et al., 2017). They learn dense embeddings for words in sentences. They move beyond rules or simple statistics to learn hierarchical representations of language. This enables a more "human-like" understanding of context and meaning.

There are many analogous tools and technologies, beginning with the frequency analysis tools that may have been used in the movie. There was a scene when Ian was looking at his computer and I believe he had a seismograph up. I’m sure this was used to try and find patterns in the deep, resonant sounds that the Heptapods make when communicating. They probably also used tools to analyze the frequency of the symbols and their combinations. This is similar to how we use tokenizers and frequency counters to denote importance and to find patterns between different words and context. They probably also used pattern recognition software to find patterns in the symbols of Heptapods language. Especially when the Heptapods spewed out a crazy number of symbols towards the end of the movie. They probably also tried Machine Translation tools in an attempt to link the alien’s language to some language from earth. Dr. Banks in the movie creates a sort of mental knowledge graph connecting the dots of Heptapod concepts. Knowledge graphs like WordNet or Wikidata are used to gain structured representations that can help models understand relationships between entities and concepts.

This movie was so good the first time and even better the second. My reflection is that it housed a plethora of challenges and examples of NLP techniques and systems. It was the most interesting when Dr. Banks started talking about the Sapir-Whorf hypothesis. It basically states that language doesn’t just describe reality but shapes one's perception and cognition. Dr. Banks' ability to perceive time non-linearly after learning the alien’s language suggests that mastering a new language can altogether alter a person's cognitive framework. Also the concept that misinterpretations can have catastrophic consequences is a prevailing lesson that I take from this move. NLP models still struggle with the understanding of some words and phrases because these words and phrases may have double meanings. Now I understand why Chatbots like GPT’s training data and process was scrutinized by OpenAI so much. They have to ensure that the intent behind their user’s queries is clearly understood or there could be catastrophic and potentially costly consequences for Open AI.

In conclusion, “Arrival” serves as a profound thought experiment for the field of NLP. It pushes us to consider the philosophical underpinnings of language. As well as the cognitive impact of linguistic structures, and the immense challenges of true cross-cultural understanding. It suggests that the future of NLP might involve not just building more accurate translation or understanding systems, but also developing AI that can adapt its own cognitive framework to truly bridge linguistic and conceptual divides. I also left out sarcasm in my analysis because I don’t think the Heptapods had any sarcasm in their language. I didn’t see any instances of this in the movie.

**References:**

Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Prentice Hall.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.

Blitzer, J., McDonald, R., & Pereira, F. (2006, July). *Domain adaptation with structural correspondence learning*. Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia.

Chomsky, N. (1957). *Syntactic structures*. Mouton & Co.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.

Goldberg, Y. (2017). *Neural network methods for natural language processing*. Morgan & Claypool Publishers.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. Advances in neural information processing systems, *30*.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). MIT Press.