Parnal Sinha, Amisha Singleton
ITAI 2376 – Deep Learning in AI
A05: Exploring “Arrival” and NLP
Group 3

Exploring "Arrival" and Natural Language Processing: Decoding Alien and Human Communication

1. Introduction

"Arrival," a 2016 science fiction film directed by Denis Villeneuve, examines communication between humans and extraterrestrials, focusing on linguistic challenges. When twelve alien spacecraft, known as “shells,” appear globally, linguist Dr. Louise Banks is recruited to decipher their circular, symbolic language to understand their intent, amid rising international tensions. This narrative connects directly to Natural Language Processing (NLP), a field that enables computers to interpret human language, highlighting challenges such as ambiguity, sarcasm, and cultural variation. This report analyzes key scenes from "Arrival," linking them to NLP difficulties based on class materials, presenting findings in a clear, organized manner. The work reflects collaboration between Amisha and I, where we discussed and verified our findings on ambiguity, communication methods, and the movie’s broader implications for NLP.

2. Plot Overview: Linguistic Focus in "Arrival"

The film depicts the sudden arrival of twelve alien spacecraft across Earth, prompting global concern. Dr. Louise Banks, a linguist, collaborates with physicist Ian Donnelly to interpret the Heptapods’ circular, semasiographic symbols within a Montana shell, separated by a glass barrier. Nations misinterpret the message “Offer weapon” as a threat, but Banks decodes the Heptapod language, revealing its nonlinear temporal structure, enabling simultaneous perception of past, present, and future. This linguistic breakthrough prevents international conflict and alters Banks’ understanding of time. The narrative emphasizes language’s role in cognition and communication, illustrating the Sapir-Whorf hypothesis’s relevance to human and alien interaction.

3. NLP Challenges in "Arrival"

"Arrival" reflects NLP challenges evident in human language processing, identified through specific scenes:

a. **Ambiguity:** The Heptapod message “Offer weapon” is interpreted variably as a weapon, tool, or gift, paralleling NLP’s difficulty with ambiguous texts, such as “That’s cool” indicating either positivity or sarcasm based on context. Additionally, Louise’s initial misinterpretation of Heptapod symbols as threats, similar to national misreadings, mirrors NLP errors in social media sentiment analysis, where tweets like “Love this update… not” are misclassified due to sarcasm.

   Technical Insight: This requires probabilistic disambiguation, using models like Bayesian inference or neural networks to assess contextual likelihoods and resolve multiple meanings.

b. **Idiomatic Expressions:** Heptapod semasiography, with its nonlinear, circular symbols, contrasts with human linear syntax, challenging Louise akin to NLP's struggles with idioms like "kick the bucket," which lack literal meaning.

   Technical Insight: This necessitates context-sensitive parsing, addressed by statistical models (e.g., n-grams) or deep learning (e.g., transformers) to capture syntactic dependencies and non-literal meanings.

c. **Sarcasm/Intent:** Nations misinterpret Heptapod intent, such as China viewing "Offer weapon" as hostile, analogous to NLP's challenges detecting sarcasm in tweets, e.g., "Great, another meeting" implying negativity despite positive phrasing.

   Technical Insight: Intent detection leverages sentiment analysis with recurrent neural networks (RNNs) or BERT, modeling contextual cues like tone or negation to predict underlying intent accurately.

d. **Cultural Variations:** Heptapod nonlinear time perception differs from human linearity, resembling NLP's difficulties with regional dialects, such as "bodega" (New York) versus "corner shop" (London).

   Technical Insight: This is addressed by multilingual embeddings (e.g., multilingual BERT) trained on diverse corpora, enabling models to handle cultural and syntactic variations effectively.

These parallels align with class materials on NLP textual data challenges, highlighting areas for technological improvement.

4. Communication Methods with Extraterrestrials and NLP Parallels

Louise Banks' approach to decoding Heptapod language corresponds to NLP methodologies:

a. **Rule-Based NLP:** Banks initially teaches basic vocabulary (e.g., "human," "Louise") using grammatical rules, similar to rule-based NLP systems relying on deterministic grammars for syntactic parsing. For example, her writing of "Offer weapon, question mark" on the glass in Session 36 tests basic syntax, akin to early NLP parsing simple sentences but limited by Heptapod complexity.

   Technical Insight: This employs context-free grammars for syntax, analogous to regular expressions in text processing, but requires statistical or deep learning for nonlinear semantics.

b. **Statistical NLP:** Banks identifies patterns in symbol frequency and structure, paralleling statistical NLP's use of n-gram models or hidden Markov models (HMMs) to estimate probabilities, as in early machine translation.

Technical Insight: This uses maximum likelihood or Bayesian inference to model co-occurrences, capturing Heptapod patterns but necessitating deep learning for semantic depth, modeled via probabilistic graphical models.

c. **Deep NLP:** Banks' breakthrough in understanding Heptapod nonlinear time via their symbols reflects deep NLP's use of neural networks, such as transformers or convolutional layers, to extract hierarchical features.

Technical Insight: This leverages transformer architectures (e.g., BERT) or convolutional layers for high-level pattern recognition, modeling Heptapod temporal correlations with attention mechanisms, akin to advanced text processing for complex contexts.

d. **Tools/Analogies:** Banks' whiteboards and recordings equate to NLP tokenization and lexical analysis, while the glass barrier represents the interpretability challenges in NLP, similar to black-box models in machine learning.

Technical Insight: These correspond to text processing pipelines, where whiteboards tokenize language, recordings score patterns, and the barrier reflects interpretability issues (e.g., SHAP) in neural models.

5. Reflections on Arrival's Language Portrayal and NLP Implications

"Arrival" highlights language's role in shaping cognition, suggesting NLP must evolve to address non-standard, nonlinear data. The Heptapod semasiography, embodying the Sapir-Whorf hypothesis, indicates potential for NLP to incorporate context-driven, intent-focused systems, possibly using advanced neural architectures for temporal or cultural variability. This challenges current NLP to enhance models with broader data sets, potentially integrating multimodal or non-sequential approaches to improve accuracy and adaptability, inspired by the film's emphasis on communication's complexity. Lucy's (n.d.) analysis of linguistic relativity supports this, suggesting language structures influence perception, prompting NLP to develop flexible models for diverse linguistic structures like Heptapod's.

6. What is the Sapir-Whorf Hypothesis?

The Sapir-Whorf hypothesis, named after linguists Edward Sapir and Benjamin Lee Whorf, proposes that language influences how people think and perceive the world. It has two versions:

Strong Version (Linguistic Determinism): This suggests language completely determines thought and perception. For instance, if a language lacks terms for certain colors, speakers might not distinguish those colors, though this is highly controversial and not widely accepted today.

Weak Version (Linguistic Relativity): This posits that language shapes, but does not fully control, thought and perception. It argues that linguistic structure and vocabulary subtly affect

understanding, such as languages with numerous snow-related terms enhancing speakers’ awareness of snow variations.

In "Arrival," the Heptapods’ nonlinear, circular language allows them to perceive time simultaneously—past, present, and future—because their symbols are not linear like human language. Louise’s mastery of this language alters her time perception, illustrating the hypothesis. For NLP, this suggests that system design could influence language interpretation; focusing solely on linear, human-centric models might limit understanding of non-human languages like the Heptapods’, prompting development of more adaptable NLP approaches. Lucy (n.d.) reinforces this, emphasizing language’s perceptual impact.

7. Conclusion

"Arrival" illustrates NLP challenges and opportunities, from ambiguous messages to cultural differences, through an extraterrestrial perspective. Louise’s linguistic efforts demonstrate communication’s challenges and potential, encouraging NLP advancements in contextual understanding, cultural sensitivity, and adaptability. This project, developed collaboratively by our team, provided valuable insights into NLP’s practical and theoretical boundaries. I found this assignment to be interesting and particularly compelling as it forced me to reconcile and coqalesce current NLP developments with potential future applications in NLP developments, possibly extending its capabilities to diverse data types and enhancing system performance.

References

1. Villeneuve, D. (Director). (2016). *Arrival* [Film]. Paramount Pictures.
2. ITAI 2376 Course Materials. (n.d.). Challenges of textual data and domains of NLP [Lecture notes].
3. Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
4. Lucy, J. A. (n.d.). *Linguistic relativity* [PDF]. Center for the Study of Languages and Cultures, University of Notre Dame. https://cslc.nd.edu/assets/142525/lucy_linguistic_relativity.pdf