# Picard understanding Darmok: A Dataset and Model for Metaphor-Rich Translation in a Constructed Language

**Peter A. Jansen**
University of Arizona
`pajansen@arizona.edu`

**Jordan Boyd-Graber**
University of Maryland
`jbg@umiacs.umd.edu`

## Abstract

Tamarian, a fictional language introduced in the *Star Trek* episode *Darmok*, communicates meaning through utterances of metaphorical references, such as *"Darmok and Jalad at Tanagra"* instead of *"We should work together."* This work assembles a Tamarian-English dictionary of utterances from the original episode and several follow-on novels, and uses this to construct a parallel corpus of 456 English-Tamarian utterances. A machine translation system based on a large language model (T5) is trained using this parallel corpus, and is shown to produce an accuracy of 76% when translating from English to Tamarian on known utterances.[1]

## 1 Introduction

Science fiction and fantasy literature has long created constructed languages for their characters, from Elvish in *Lord of the Rings* and Klingon in *Star Trek* to Heptapod in *Arrival* (Cheyne, 2008). These languages often have many of the same syntactic or semantic features as human languages, and some (such as Klingon) have been developed to a level where full dictionaries (Okrand, 1992) and online translators are available.[2]

An unconventional language was proposed in an episode of *Star Trek: The Next Generation* called *"Darmok"*, where a race of aliens called the Tamarians speak a language that is communicated exclusively through metaphors. Instead of direct reference (e.g. *"I want to give this to you"*), Tamarians speak in metaphorical references grounded in stories (e.g. *"Temba, his arms wide"*) that (like symbols) have learned associations with their true meaning meaning. In the *Darmok* story, the unusual nature of the language poses a challenge for both the automated translation systems and the
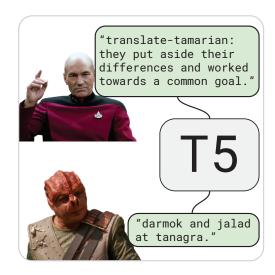


Figure 1: An example of translating English to the metaphor-grounded Tamarian language using T5.

characters in the story to learn. The creator of the language, Joe Mendowsky was inspired by the difficulty of translating across cultures (Block and Erdmann, 2012), and Tamarian has since been the subject of repeated informal study (Bogost, 2014) in the 30 years since the episode aired.

This work investigates the feasibility of translating this artificial metaphor-rich language via our new parallel corpus of English-Tamarian phrases (Figure 1). Our machine translation system based on a large language model (Raffel et al., 2020, T5) has 76% accuracy in translating English phrases to Tamarian metaphorical utterances. This suggests automatically translating metaphor-grounded languages may be feasible, though we discuss several pragmatic challenges in representing complex expressions and generating a parallel corpus preventing scaling the approach.

## 2 English-Tamarian Parallel Corpus

Comparatively few Tamarian utterances have been authored, effectively limiting the size and scope of the effort. To maximize the number of available utterances, all utterances from the original broadcast

---

[1] Data and code available at: `https://github.com/cognitiveailab/darmok`

[2] https://www.translate.com/klingon-english

| | Tamarian Utterance | Inferred Meaning | English Example |
|---|---|---|---|
| 1 | Darmok and Jalad at Tanagra | Working together | Knowing they would both be needed, they went together. |
| 2 | Temba, his arms wide. | Giving | The child offered his toy to his friend. |
| 3 | Kira at Bashi. | Story-telling | They described what had happened to those who listened. |
| 4 | Chenza at court, the court of silence. | Incontestability | The results were beyond reproach. |
| 5 | Zima at Anzo, Zima and Bakor. | Persistence | They continued their task, undeterred from past failures. |
| 6 | Fendit, refusing the flame. | Refusing help | She preferred to work alone, without assistance. |
| 7 | Chatha and Teribium, the fire warm. | Hospitality | Their household was offered for rest and comfort. |
| 8 | Jeral, her arms weary. | Being tired | She was spent at the end of the day. |
| 9 | Pirakee, with clouds parted. | Visibility | She turned on a flashlight, making it easier to see. |
| 10 | Hammat dancing. | Liking something | It filled them with delight. |

Table 1: Example Tamarian utterances, their inferred meaning, and an English example from the parallel corpus.

episode, as well as those in three licensed novels featuring a Tamarian main character were used (Beyer, 2012, 2014, 2015). Approximately twenty utterances are provided in the *Darmok* episode, while an additional forty-eight are used in the novels, for a total of sixty-eight utterances.

**Tamarian-to-English dictionary:** To create a parallel English-Tamarian corpus, first a Tamarian-to-English dictionary that captures the inferred meaning of each Tamarian utterance was required. The meanings of the twenty broadcast utterances was ascertained from a Reddit thread with extensive discussion of the topic.[3] The meanings of the remaining forty-eight utterances was inferred as best as possible from the surrounding context of where they appeared in their respective novels.

**Tamarian-English Parallel Corpus:** Training a machine translation system requires a parallel corpus, where utterances of one language are paired with utterances of a second language, where the utterances in both languages have the same meaning. Tamarian utterances abstractly refer to specific types of situations that could be applicable to many circumstances. Thus, for each Tamarian utterance a set of $k$ English examples were manually authored, with ten examples authored for thirty-nine utterances, and five examples authored for eleven utterances. Eighteen Tamarian utterances were not included in the parallel corpus as they have relatively narrow meanings, and generating a large number of parallel examples for them in English proved challenging. The final parallel corpus contains fifty Tamarian utterances, paired with 456 parallel English utterances (Table 1).

## 3 Translation Model

**Approach:** Here, English-to-Tamarian is modeled as a sequence-to-sequence (seq2seq) learning task, using English utterances as the source sentence, and a single Tamarian translation of that English utterance as the target sentence.

**Models:** Modeling used T5 (Raffel et al., 2020), a large pre-trained multi-task language model. T5 includes pre-training for a variety of tasks, including question answering, summarization, and translation. Several model sizes were explored, including T5-small (66M parameters), T5-base (220M parameters) and T5-large (220M parameters). The model prompt took the form of:

```
translate English to Tamarian: {src}
```

where *{src}* is the English source sentence to translate (e.g. *"She offered it to them"*). The model then generated a corresponding target sequence corresponding to the Tamarian translation of the source sentence (e.g. *"Temba. His arms wide."*). The model was implemented using the Huggingface Transformers library (Wolf et al., 2020).

**Dataset splits:** Due to small dataset, we use 5-fold crossvalidation: with 60% of data used for training, 20% for development, and 20% for test. For utterances with ten examples, this corresponds to six train, two development, and two test samples per run, while for utterances with five examples, this corresponds to three train, one development, and one test sample per run.

**Evaluation Metrics:** Translation performance was evaluated using SACREBLEU (Post, 2018), a metric that measures translation performance using $n$-grams, while taking partial matches into account. Here, because only fifty Tamarian utterances are

| | | Translation Performance | | | |
|---|---|---|---|---|---|
| | | Dev. | | Test | |
| | Model | BLEU | Acc. | BLEU | Acc. |
| T5 | T5-Small | 38 | 34.4% | 41 | 38.0% |
| | T5-Base | 71 | 72.8% | 70 | 72.4% |
| | T5-Large | **80** | **82.4%** | 74 | **76.4%** |

Table 2: Average English-to-Tamarian translation performance on both development and test sets. BLEU measures per-token accuracy, while *Acc.* refers to the average binary classification accuracy of choosing the correct Tamarian utterance for a given English input sentence.

available, and their surface presentation is generally constant, we also consider evaluating translation as an *N-class* classification task where a given English input sentence can be classified as one of fifty Tamarian utterances.

# 4 Results

Models were trained until performance (BLEU) asymptoted on the development set, at thirty epochs. The best performing model achieves a translation accuracy of 76% on the unseen test set, which corresponds to translating approximately three out of four English utterances from the corpus correctly into Tamarian (Table 2).

# 5 Discussion

As a constructed language for a fictional universe, Tamarian is a low resource language with fewer than one hundred known utterances. What might it take to grow Tamarian (or a metaphorically-grounded Tamarian-like language) into a more complete artificial language similar to Klingon? This section attempts to address the challenges of scaling beyond this work in the context of two central difficulties: growing the parallel corpus of metaphors, and challenges associated with the semantics of translating complex ideas in Tamarian.

## 5.1 Growing the Parallel Corpus

Growing the vocabulary of metaphors in Tamarian presents a unique challenge for constructed languages. Where human languages typically expresses base-level semantics at the level of the morpheme or word, Tamarian's most atomic construction is a single metaphor, making approaches that start with translating a dictionary challenging to adapt. One approach to growing Tamarian would be to continue the current manual approach, identifying a set of atomic events that convey common situations (such as *eating, giving, taking, or helping*),

| Tamarian Utterance | Inferred Meaning |
|---|---|
| *Gesture/Context Hypothesis* | |
| Temba, his arms wide. *Also: Pointing at item* | Hand me the blue screwdriver I am pointing at |
| *Specificity Hypothesis* | |
| Jeral, her gift. | Give me a blue screwdriver on the left |
| *Modifier Hypothesis* | |
| Temba, his arms wide. | Giving |
| Paris, in the garage. | Screwdriver |
| Tolanis painting, in winter. | Blue |
| Bakor, examining. | Look to the left |

Table 3: Examples of the three hypotheses for how fine-grained semantics could be inferred or composed in Tamarian.

and authoring utterances grounded in an expanded Tamarian mythology—for example, *"Timba, his stomach rumbling"* to convey the notion of hunger. The prerequisite for having an exhaustive list of possible event schemas to translate would likely make this approach challenging to scale.

**Automatic Generation:** An alternate approach was suggested by Picard in *Darmok* – to use the existing body of human literature (such as the *Epic of Gilgamesh*) to build a Tamarian-like language grounded in metaphors inferred from classic literature. Picard suggests that *"Gilgamesh and Enkidu at Uruk"* might be an utterance to represent a central component of the story – two people who were first in conflict coming together in friendship. Such an automatic approach to building a Tamarian-like language is in principle feasible, potentially making use of recent successes in automatic summarization to extract key elements of a story in templated form (e.g. {PERSON X} AND {PERSON Y} AT {LOCATION}) to generate novel utterances. One of the challenges with this approach is that narratives often contain many events, specified both at a low-level (e.g. Enkidu entering the city of Uruk) and high-level (e.g. Gilgamesh and Enkidu eventually forming a friendship in spite of their differences), and identifying only a single idea to be represented by the utterance would be difficult.

## 5.2 The Challenge of Translating Fine-grained Semantics

It has been hypothesized that Tamarian may not be well suited to expressing fine-grained semantics, and would present challenges for translating utterances such as *"Hand me the blue screw driver on*

*the left"* (Bogost, 2014). While the few observed multi-utterance exchanges of Tamarian have (so far) typically conveyed steps in a story, we present three hypotheses for how fine-grained semantics might be achieved, with examples shown in Table 3:

1. *Gesture/Context hypothesis:* The spoken Tamarian language may ground ambiguity through gestures or other situated contextual cues, as the Tamarian captain does when he utters *"Temba, his arms wide" (take)* and gestures to a weapon.
2. *Specificity hypothesis:* Though impractical, the Tamarian language may have many utterances to refer to very specific situations.
3. *Modifier hypothesis:* Unobserved classes of utterances may serve as modifiers, providing additional clarification to an utterance.

There is partial observation of both the *gesture/context* and *modifier* hypotheses provided in the original *Darmok* episode, and we believe the modifier hypothesis likely provides a mechanism for composing larger units of meaning akin to a generative grammar.

The more fundamental challenge of extending Tamarian is that every sentence must be connected to an underlying mythology: if you want to translate a sentence you must first create a universe (Sagan et al., 1983). While we can invent Tamarian sounding proper nouns, a more fundamental challenge is to build a world where there are characters who would have or invent a screwdriver, a character who could successfully use it, a character who would use it incorrectly, and perhaps someone else who could address when you've accidentally stripped the head of the screwdriver.

Thus, the challenge is not just creating enough examples but also building the cultural cannon to support those examples. While this is a unique linguistic challenge for Tamarian, it follows the course of other constructed languages: Quenya was developed alongside the backstory of Middle Earth (Lewis, 1995) and the creator of the Klingon language also ensured that the Klingon mythology was recorded in the Klingon language (Schönfeld et al., 2011). Tamarian foregrounds this challenge of obtaining enough cultural context to translate (Keesing, 1985; Maitland, 2017).

# 6 Related Work: (Computational) Linguistics for Constructed Languages

The elephant in the room is whether it is worthwhile to study constructed languages at all. This section seeks to answer that question with a resounding yes by discussing the other insights that have come from scholarly investigations of constructed languages.

Tamarian is from the Star Trek Universe, so it is instructive to spend a little time first with the oldest Star Trek language, Klingon. Klingon is often used in NLP education because it has features that are rare in natural languages but it is incredibly regular: a morphological analyzer can get 100% accuracy but still have fascinating properties like affixes for honorifics, completion, and tense (Wicentowski, 2004). Likewise, because Klingon is by construction meant to feel literally alien, its OVS structure can also upend students' part of speech tagging expectations (Boyd-Graber, 2014).

But Klingon is not just a fun exercise for programmers and linguists; the creation of parallel data (as discussed above for Tamarian) also explores the interplay between culture and translation. For the translation of *Hamlet* into Klingon, cultural adaptation (Peskov et al., 2021) is also needed: for example, Fortinbras becomes "the most insuborinate head of the House of Duras" (Kazimierczak, 2010). The art of translation often relies on metaphor (Veale, 2016) and cultural knowledge (Vinay and Darbelnet, 1995), and just as exploring Klingon can reveal limitations of our understanding of affix morphology and OVS word order, Tamarian can help illuminate the limitations of metaphor in communication.

All extant constructed languages are low resource languages, which typically pose challenges for machine translation (Haddow et al., 2021). Like how Klingon can emphasize particular aspects of a language (word order, morphology), Tamarian helps focus attention on the role of mythology, inter-personal relationships, and multiword expressions for translation.

# 7 Conclusion

This paper is an initial English–Tamarian translation model. This task is difficult because it not only maps words to words but also maps metaphor to typical translation phrases. While Tamarian is a constructed language, it shows large language models' ability and limitations for metaphor.

# References

Kirsten Beyer. 2012. *Star Trek Voyager: Eternal Tide*. Pocket Books.

Kirsten Beyer. 2014. *Star Trek Voyager: Acts of Contrition*. Pocket Books.

Kirsten Beyer. 2015. *Star Trek Voyager: Atonement*. Pocket Books.

P.M. Block and T.J. Erdmann. 2012. *Star Trek: The Next Generation 365*. ABRAMS, Incorporated (Ignition).

Ian Bogost. 2014. Shaka, when the walls fell. *The Atlantic*.

Jordan Boyd-Graber. 2014. Homework 5: Qu' bopbe' paqvam.

Ria Cheyne. 2008. Created languages in science fiction. *Science Fiction Studies*, 35(3):386–403.

Barry Haddow, Rachel Bawden, Antonio Valerio Miceli Barone, Jindřich Helcl, and Alexandra Birch. 2021. Survey of low-resource machine translation.

Karolina Kazimierczak. 2010. Adapting shakespeare for "star trek" and "star trek" for shakespeare: "the klingon hamlet" and the spaces of translation. *Studies in Popular Culture*, 32(2):35–55.

Roger M. Keesing. 1985. Conventional metaphors and anthropological metaphysics: The problematic of cultural translation. *Journal of Anthropological Research*, 41(2):201–217.

Alex Lewis. 1995. Historical bias in the making of "the silmarillion". *Mallorn: The Journal of the Tolkien Society*, (33):158–166.

S. Maitland. 2017. *What Is Cultural Translation?* Bloomsbury Advances in Translation. Bloomsbury Academic.

Marc Okrand. 1992. *The Klingon Dictionary: The Official Guide to Klingon Words and Phrases*. Simon and Schuster.

Denis Peskov, Viktor Hangya, Jordan Boyd-Graber, and Alexander Fraser. 2021. Adapting entities across languages and cultures. In *Findings of Empirical Methods in Natural Language Processing*.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Carl Sagan, Ann Druyan, and Steven Soter. 1983. The lives of the stars. *Cosmos*, 1(9).

Floris Schönfeld, Mar Okrand, Kees Ligtelijn, and Vincent W.J Van Gerven Oei, editors. 2011. *paq'balth: The Klingon Epic*. Punctum Books, Brooklyn, NY.

Tony Veale. 2016. Round up the usual suspects: Knowledge-based metaphor generation. In *Proceedings of the Fourth Workshop on Metaphor in NLP*.

Jean-Paul Vinay and Jean Darbelnet. 1995. *Comparative stylistics of French and English: A methodology for translation*, volume 11. John Benjamins Publishing.

Richard Wicentowski. 2004. Multilingual noise-robust supervised morphological analysis using the Word-Frame model. In *Proceedings of the 7th Meeting of the ACL Special Interest Group in Computational Phonology: Current Themes in Computational Phonology and Morphology*, pages 70–77, Barcelona, Spain. Association for Computational Linguistics.

Thomas Wolf, Julien Chaumond, Lysandre Debut, Victor Sanh, Clement Delangue, Anthony Moi, Pierric Cistac, Morgan Funtowicz, Joe Davison, Sam Shleifer, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45.