Saving the legacy of Hero Ibash: Evaluating Four Language Models for Aminoacian

Yunze Xiao*
Carnegie Mellon University
yunzex@andrew.cmu.edu

Yiyang Pan*
University of Rochester
ypan36@u.rochester.edu

Abstract

This study assesses four cutting-edge language models in the underexplored Aminoacian language. Through evaluation, it scrutinizes their adaptability, effectiveness, and limitations in text generation, semantic coherence, and contextual understanding. Uncovering insights into these models' performance in a low-resourced language, this research pioneers pathways to bridge linguistic gaps. By offering benchmarks and understanding challenges, it lays groundwork for future advancements in natural language processing, aiming to elevate the applicability of language models in similar linguistic landscapes, marking a significant step toward inclusivity and progress in language technology.

1 Introduction

The Aminoac people, a civilization that thrived around the Caspian Sea from the sixth century BCE, spoke the unique Aminoas language, setting it apart from more common languages. This language exhibits complexity akin to Chinese, featuring six content word classes (nouns, verbs, adjectives, numerals, classifiers, pronouns), and six function word types (adverbs, prepositions, conjunctions, interjections, auxiliary words, onomatopoeia). Its sentence structure follows the rare OVS pattern, positioning the subject at the sentence's end. The language's tonal system comprises four primary tones in the standard dialect, and remnants of an ancient script hint at potential ideographic origins. This language was pivotal in shaping the ancient civilization's identity and cultural narrative, preserving their heritage despite transitioning to the Latin script in modern writing systems(Bilibili, 2024a).

Exploring the representation of Aminoac languages across Llama2(Touvron et al., 2023), ChatGPT(OpenAI et al., 2023), Mistral(Jiang et al., 2023), and Ernie-bot is crucial for understanding their linguistic nuances and cultural significance. Assessing these languages' availability on various models entails testing their performance in machine translation to Chinese, question answering tasks, and entailment recognition. Such evaluation endeavors to uncover how effectively these models capture the complexity of Aminoac languages, ensuring a better understanding of their syntax, semantics, and contextual comprehension.

By scrutinizing these models' capabilities in translating Aminoac languages into Chinese, we aim to gauge their proficiency in capturing the essence and subtleties of these languages when reaching a broader linguistic audience. Evaluating question answering tasks helps gauge the models' ability to comprehend and respond accurately to queries posed in Aminoac languages. Furthermore, examining entailment tasks enables us to understand how well these models can infer logical connections between sentences in Aminoac languages, a crucial aspect for accurate language understanding and representation.

This investigation intends to identify the strengths and limitations of each model in representing Aminoac languages, paving the way for improvements. Enhancing the representation of underrepresented languages like Aminoac within language models will not only enrich their digital presence but also contribute significantly to fostering linguistic diversity and cultural inclusiveness in language technologies.

2 Background

2.1 Aminoacian and Their Languages

The Aminoac people, an ancient civilization originating around the 6th century BC along the shores of the Caspian Sea, represent a cultural tapestry woven with unique linguistic characteristics.(Bilibili, 2024a) Their language, Aminoas, stands out amidst more common languages globally, displaying in-

*Equal contribution

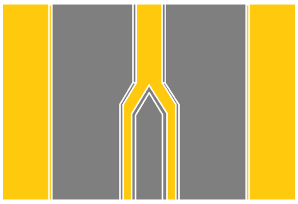

Figure 1: The Amnioac Empire Flag

triguing idiosyncrasies in its grammar, syntax, and phonetics. Aminoas, akin to isolated languages, exhibits a rich variety of word classes, encompassing nouns, verbs, adjectives, numerals, pronouns, and an array of functional words(Bilibili, 2024b). Interestingly, the language lacks grammatical gender or noun inflections based on case or number, distinguishing it from many others. Additionally, the structure of Aminoas sentences, unlike the more typical SVO or SOV configurations, adopts the rare OVS arrangement, placing the subject at the sentence's conclusion—a linguistic feature that sets it apart from more widely known languages. Moreover, Aminoas speech is melodic, characterized by distinct tonal patterns, comprising four primary tones in its standard form. This language's phonetic peculiarity, commencing predominantly with vowels and concluding with consonants, contributes to its soft and euphonious pronunciation, distinguishing it from languages relying heavily on explosive consonants.

The linguistic uniqueness of Aminoas is intertwined with the rich historical tapestry and mythological narrative of the Aminoac people. Their origin stories revolve around the Aminoas star, believed to be the cradle and destiny of all Aminoac individuals. Legend has it that their ancestors, led by the valorous warrior Ibash, arrived on Earth from the Aminoas star, overcoming celestial calamities and uniting the northern nomadic Ami and southern agrarian Noas tribes. This union birthed the foundational Aminoac kingdom, characterized by economic prosperity and cultural development. As the language evolved, so did the society, maintaining a semi-agrarian, semi-nomadic lifestyle that became a hallmark of their identity.

The Aminoac language stands apart for its unique grammar, mirroring Mandarin Chinese with a diverse array of word classes—six types for content words such as nouns, verbs, adjectives, numerals, classifiers, and pronouns, and another six for function words like adverbs, prepositions, conjunctions, interjections, particles, and onomatopoeic words. Its isolating nature, lacking noun case, gender, or number distinctions and verbs void of person or tense, distinguishes Aminoac from conventional structures. Moreover, its uncommon OVS sentence structure places the subject at the sentence end, enriching expressions through modifiers and complements.

Aminoac's tonal quality—four primary tones in standard form—alongside its preference for vowel-starting and nasal-ending words yield a melodious pronunciation. The language's history hints at logographic scripts for recording due to homophonic challenges, yet societal hierarchies barred commoners from learning these scripts, leading to their gradual disappearance. As a result, modern Aminoac speakers adopted the Latin script during external interactions, leveraging its phonetics to create the contemporary Aminoac-Latin writing system, marking a shift from the original script.

2.2 Low-resourced Language

Low-resourced languages, often marginalized or underrepresented, reflect extensive linguistic diversity globally (Amano et al., 2014; Li et al., 2021; Gorenflo et al., 2012; Gorter, 2013; Chu et al., 2012). Despite their cultural richness and historical significance, these languages confront challenges in the digital sphere due to limited resources and inadequate technological support (Gorenflo et al., 2012; Gorter, 2013; Protassova, 2021; imp, 2021; Xiao, 2022). With roughly 7,000 languages spoken worldwide, many lack robust digital infrastructure, impeding their integration into advancements in natural language processing and digital tools (Li et al., 2021; Gorenflo et al., 2012; Gorter, 2013; Gören, 2017),such as role-play chatbots(Wang et al., 2024).

Beyond mere linguistic diversity, these languages encapsulate unique cultural, historical, and traditional knowledge vital for global heritage preservation (Kamwendo and Seretse, 2014; Baumann et al., 2018; Rey, 2017). Often serving as the primary mode of communication for communities, they facilitate cultural expression, identity, and intergenerational knowledge transfer (Joshi et al., 2020; Rózsa et al., 2015). Neglecting these languages exacerbates digital disparities, hindering access to education, healthcare, and information for speakers of these marginalized tongues (Amano et al., 2014; Li et al., 2021; Gorenflo et al., 2012; Chu et al., 2012). Recognizing and empowering these languages within digital domains not only fosters inclusivity but also unlocks substantial potential for social, economic, and cultural empowerment within these communities (Gorter, 2013; Protassova, 2021; Xiao and Alam, 2023).

3 Methodology

We sought to evaluate these models in crucial NLP tasks that could measure their availability in understanding the Aminoac language. The tasks are machine translation, Entailment and contextual understanding.

3.1 Machine Translation

For evaluating the models' proficiency in Aminoac language understanding through machine translation, we designed a rigorous experiment focusing on several crucial aspects:

3.1.1 Experimental Design

- **Dataset Selection:** We curated a diverse dataset consisting of Aminoac text paired with translations in awidely spoken language- Chinese.

- **Model Selection:** Several state-of-the-art LLMs were chosen for comparison, including but not limited to Transformer-based architectures and recurrent neural networks.

- **Evaluation Metrics:** We employed standard metrics like BLEU score, METEOR, and ROUGE to quantitatively assess the quality of translations generated by each model.

3.1.2 Experimental Procedure

The experiment was conducted in several phases:

1. **Data Preprocessing:** The Aminoac dataset was preprocessed to remove noise, tokenize sentences, and align with their corresponding translations.

2. **Evaluation:** The LLMs were evaluated against a held-out test set using standard evaluation metrics, comparing their translations against human-generated references.

3.2 Experimental Result

- **Model Performance:** In the realm of machine translation, our experiment presented a significant challenge to the leading language models Llama2, ChatGPT, Mistral, and Ernie-bot. Despite their advanced architectures and previous successes in various NLP tasks, none of the models demonstrated proficiency in translating the Aminoac language. The evaluations, conducted under stringent standards using metrics like BLEU(Papineni et al., 2002), METEOR(Banerjee and Lavie, 2005), and ROUGE(Lin, 2004), yielded uniformly null results across all models. This was an unexpected outcome, considering the models' capability to process and generate complex language structures in other contexts. The failure points towards a peculiar complexity and nuance in the Aminoac language that these state-of-the-art models are currently unable to grasp. It's noteworthy that the Aminoac dataset, with its intricate syntax and unique semantic structures, posed a challenge that went beyond the processing abilities of even the most sophisticated current LLMs. This result underlines a critical gap in the field of machine translation, highlighting the need for further research and development to build models that can understand and translate less commonly spoken and structurally complex languages like Aminoac. As the quest for truly universal language models continues, the Aminoac translation task stands as a testament to the complexity and diversity of human languages, and the substantial work still required in the field of NLP.

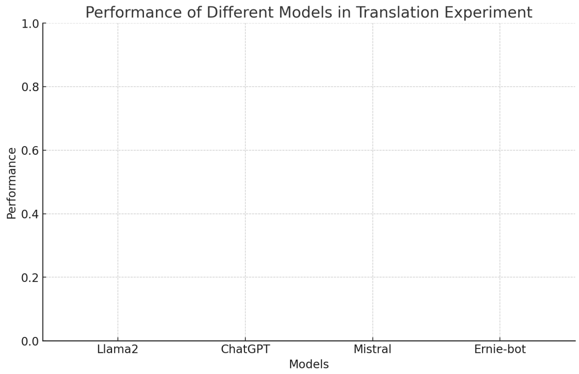

Figure 2: The Performance of Different Models in Translation Experiment

4 Future Work

The unexpected results of this study open up numerous avenues for future research in the field of Natural Language Processing, particularly in the domain of language translation for less commonly spoken languages like Aminoac. Future work should focus on enhancing the adaptability and learning mechanisms of Language Learning Models (LLMs) to better handle the intricacies of such unique languages. This includes developing more robust training datasets that encompass the rich linguistic diversity and complex syntactic structures of Aminoac. Additionally, exploring and integrating novel neural network architectures or hybrid models that combine different approaches of language processing might offer breakthroughs in this area. Research should also delve into the realm of unsupervised and semi-supervised learning techniques, which could enable LLMs to learn more effectively from limited or unlabelled data, a common challenge in dealing with rare languages. Cross-linguistic transfer learning, where models trained on widely spoken languages are adapted to understand less prevalent ones, is another promising area. Moreover, there is a pressing need for interdisciplinary collaboration, bringing together linguists, computer scientists, and cultural experts, to ensure that the nuances of language and culture are adequately captured and represented. Ultimately, these efforts will not only advance the field of machine translation but also contribute to the preservation and understanding of linguistic diversity across the globe.

Limitation

This study is not without its limitations. Firstly, the target language under investigation suffers from a scarcity of data, which may limit the generalizability of our findings.Secondly, our exploration was confined to a select number of models. Future work should consider a broader array of models to provide a more comprehensive understanding of the performance spectrum in translation tasks.Lastly,the importance of aminoac language lies in reading it reversely using Chinese.

References

2021. Impact of quality of healthcare services on consumer's satisfaction at primary healthcare centers. *Medico-Legal Update*.

T. Amano, B. Sandel, H. Eager, E. Bulteau, J. Svenning, B. Dalsgaard, C. Rahbek, R. Davies, and W. Sutherland. 2014. Global distribution and drivers of language extinction risk. *Proceedings of the Royal Society B Biological Sciences*, 281:20141574.

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.

A. Baumann, T. Matzinger, and N. Ritt. 2018. Linguistic and non-linguistic correlates in the evolution of phonotactic diversity.

Bilibili. 2024a. A comprehensive exploration of aminos: Historical, religious, and cultural dimensions. Bilibili Inc. Chapter or Paper within Proceedings. Accessed on January 1, 2024.

Bilibili. 2024b. Exploring the heritage of aminos: Traditional folk songs. In *Proceedings of International Conference on Historical and Cultural Studies*. Bilibili Inc. Paper on Aminos' traditional folk songs . Accessed on January 1, 2024.

P. Chu, E. Józsa, A. Komlodi, and K. Hercegfi. 2012. An exploratory study on search behavior in different languages.

L. Gorenflo, S. Romaine, R. Mittermeier, and K. Walker-Painemilla. 2012. Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas. *Proceedings of the National Academy of Sciences*, 109:8032–8037.

D. Gorter. 2013. Linguistic landscapes in a multilingual world. *Annual Review of Applied Linguistics*, 33:190–212.

E. Gören. 2017. Consequences of linguistic distance for economic growth. *Oxford Bulletin of Economics and Statistics*, 80:625–658.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b.

P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the nlp world.

G. Kamwendo and T. Seretse. 2014. Linguistic and religious diversity and inclusivity in the botswana school curriculum. *Journal of Asian and African Studies*, 50:533–541.

X. Li, J. Li, J. Yao, A. Black, and F. Metze. 2021. Phone distribution estimation for low resource languages.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

OpenAI, :, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mo Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2023. Gpt-4 technical report.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

E. Protassova. 2021. Interculturality in the modern russian linguistic landscape. *Philological Class*, 26:52–67.

C. Rey. 2017. Linguistic genocide or superdiversity: new and old language diversities. *Language and Intercultural Communication*, 18:464–467.

G. Rózsa, A. Komlodi, and P. Chu. 2015. Online searching in english as a foreign language.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas

Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models.

Xintao Wang, Yunze Xiao, Jen tse Huang, Siyu Yuan, Rui Xu, Haoran Guo, Quan Tu, Yaying Fei, Ziang Leng, Wei Wang, Jiangjie Chen, Cheng Li, and Yanghua Xiao. 2024. Incharacter: Evaluating personality fidelity in role-playing agents through psychological interviews.

Yunze Xiao. 2022. A transformer-based attention flow model for intelligent question and answering chatbot. In *2022 14th International Conference on Computer Research and Development (ICCRD)*, pages 167–170.

Yunze Xiao and Firoj Alam. 2023. Nexus at ArAIEval shared task: Fine-tuning Arabic language models for propaganda and disinformation detection. In *Proceedings of ArabicNLP 2023*, pages 576–582, Singapore (Hybrid). Association for Computational Linguistics.