



Antonios Anastasopoulos
Department of Computer Science
George Mason University
4400 University Drive, Fairfax, VA, 22030
(703)-993-6821
antonis@gmu.edu
<https://cs.gmu.edu/~antonis>

April 29, 2023

Reference for Shruti Rijhwani's Job Application

To Whom It May Concern:

It is my utmost pleasure to write this letter in support of Shruti Rijhwani's job application at your company. I came to know Shruti as a CS PhD student at the Language Technologies Institute at Carnegie Mellon University when I was a post-doc there starting in January 2019, and I have been working with her for the past three years (including after I transitioned into an assistant professor role at George Mason University).¹

I will start by summarizing by impression of Shruti: she is **a talented researcher, an excellent communicator and organizer, and quite literally an ideal collaborator on research projects**. Shruti is easily the best graduate student I have had the chance to work with and I consider myself very lucky to have had the opportunity to work with her. I will expand on the above aspects in the following sections.

Shruti and I collaborated mainly on a 3-year-long project on developing optical character recognition (OCR) models for endangered language. Even though OCR is considered a solved problem by many, the particularities of indigenous and other endangered languages make it a very challenging setting. Put plainly, current approaches simply fail in such settings. The impact of solving this problem is potentially enormous: it would render currently inaccessible collections machine-readable and thus searchable, it would allow communities to reconnect with their literary materials, and it could facilitate the downstream development of language technologies for thousands of languages that are currently under-served by modern systems. Shruti explored a several approaches that tackle the problem from a variety of angles, showcasing the breadth of the tool-box that she can wield: unsupervised and semi-supervised approaches, incorporating structural priors into modeling, and multilingual and multimodal learning.

Beyond publishing her research, **Shruti takes pride in developing systems that are actually usable by the intended audience**, and that's exactly what she did for our OCR project. She packaged her research code in very easy-to-use code bundles and created meticulous documentation and well as tutorial run-throughs. As a result, her work is currently being used by indigenous communities and by researchers all over the world: to my knowledge, as of today, it has been used by researchers in Darwin University in Australia to digitize Aboriginal languages texts, by researchers in University of Georgia digitizing Quechua books, by the Kwak'wala community in British Columbia in Canada, and by researchers working in the Tibetan language. I could go on and on about the impact of this work, but perhaps it would suffice to say that it is a rare occasion that an academic project finds such widespread adoption so quickly after its launch – and this is all due to Shruti's work.

Above I only focused on the biggest project in which I have collaborated with Shruti – we have also worked on metrics for the morphosyntactic evaluation of automatically generated text, in speech-related projects focusing on phoneme recognition, and on multimodal data collection for low-resource languages, and on modeling NLP performance. Beyond these, Shruti has also done very impactful work in a variety of tasks, touching on practical problems including code switching, zero-shot transfer for cross-lingual learning, and

¹A bit about myself: I am an assistant professor at the Computer Science Department at George Mason University since August 2020. Previously I did a post-doc with Graham Neubig at the Language Technologies Institute at Carnegie Mellon University after concluding my CS PhD degree at University of Notre Dame in 2019 (advised by David Chiang). My research evolves around natural language processing, computational linguistics, and machine learning, with a primary focus on multilinguality and low-resource settings. I have published over 60 papers in top conferences, journals, and workshops in the area. I have supervised or co-supervised 15 masters and doctoral students.

social media analysis.

I believe that Shruti has explained the details of her research work much more ably than I would be able to; instead, I would like to emphasize three aspects of her work that make me confident that she will continue to excel in industry research.

First, her work has always been **intensely collaborative**, and working with Shruti is nothing but an absolute pleasure. Second, her work makes use of **a wide variety of tools and traditions**: it includes work involving linguistic theory, linguistic annotation, unsupervised and semi-supervised learning, and deep learning, in a time when most people are just focused on the latest deep learning methods. To me, this shows a maturity and depth that will make her a great researcher and will help her maintain a steady research agenda that sets trends instead of following them. Third, **her research breaks new ground** in areas basically untouched by others, but Shruti has already made significant strides in solving some very hard problems.

These three factors put together have made Rijhwani a rising star in the areas of NLP and low-resource NLP in particular. This subfield has grown a lot in the last few years, due in no small part to her own contributions. It's no wonder that Forbes recently included her in their 30-under-30 list for Science!

All in all, I could not recommend Shruti enough for any position relating to not only research in NLP, computational linguistics, ML, and the intersection thereof, but also in the development and deployment of related language systems.

If I can be of any further assistance or provide you with any further information, please do not hesitate to contact me.

Sincerely,
Antonios Anastasopoulos,
Assistant Professor