**(250) Clearly articulate the specific research question and the goals of the project:**

About one spoken language dies out every two weeks. Almost four thousand -- about half of all those spoken today -- will disappear entirely in the next century. It’s impossible for anyone to revive a dying language, but it is our duty to see that they are at the very least preserved. Terraling is a linguistics web application -- the development of which is the focus of this research project -- dedicated to the collection, preservation, and examination of the various linguistic properties that define a language. It is not a dictionary, and it does not store word-value pairs or direct translations. It is intended to be broader, to be a way for linguists to store and compare the defining semantic and syntactic properties of a language. These, of course, depend on the linguists who use it.

The goal of this research project is not to create Terraling from scratch. It already exists, albeit with limited functionality. The goal of this research project is to develop Terraling further, and to introduce new features: some already planned, some that will arise from the needs of the linguists who use it. The system is designed so that linguists who are native or expert in a particular language L (and understand English) can enter data about L and then the system does inter-language comparisons. dataset is also open-source, which means we don’t need in-house linguists to take advantage of this project -- we can draw on the knowledge of linguists everywhere to aid in the preservation of our languages.

**(600) Provide sufficient background to contextualize your question or problem. An educated, non-expert reader should be able to fully understand your topic. Be sure to describe the significance of this research as well: e.g. How is it unique? Why is it important? What will it contribute to the field? Include references:**

“The circumstances that have led to the present language mortality known to us range from outright genocide, social or economic or habitat destruction, displacement, demographic submersion, language suppression in forced assimilation or assimilatory education, to electronic media bombardment…” (Krauss 6). Languages are dying out, and rapidly. Estimates of how many languages will decay over the next century range from a conservative 43% (Moseley) to a more alarming 90% (Krauss).

We need to find an efficient way to conserve these languages, and there has been significant research on this topic. The linguist John Staczek compares language conservation to the conservation of art, involving “maintenance, preservation (even protection from neglect, damage, or deterioration...), and restoration” (285). Maintenance is not a luxury afforded to languages in the process of dying, so to conserve we must preserve, with the ultimate goal of restoring.

Early linguists worked only with languages that they knew, and only studied the linguistic properties that they were aware of. Although there is no concrete research into the limits of human language learning, we could posit some upper bound that most people will never reach; that is, most people will only ever be fluent in at most 20 or so languages. Thus, we must create some paradigm through which to offload, annotate, and store the linguistic properties they have observed in order to analyze even more languages.

A database satisfies precisely these needs: a database -- any conceptual store of data -- specializes in preserving data with built-in hierarchies and strong organizational structures. A database is in practice able to store linguistic knowledge orders of magnitudes greater than even the most learned polyglot. If the system allows people to compare languages based on their properties, the advantages that Terraling brings to linguistic research are evident. Some of the functional enhancements I plan to bring to the project include ….

This is, of course, not the first project to recognize the potentials of technology to assist with linguistic research. Resources such as The Endangered Language Project already exist to tackle the problem of language death, but these operate on a meta-level, classifying endangered languages and raising public awareness, but not taking concrete steps towards their conservation through maintenance, preservation, or restoration. Other linguistics databases also exist, but consist usually only a collection of scholarly articles (a search on any journal can confirm this), and do not fulfill the needs of modern linguists when it comes to preservation. The closest analog to Terraling that exists today is likely TIMIT (Garfolo), but even that serves only English-studying phoneticians and phonologists primarily. By contrast, Terraling takes full advantage of the “information revolution” (Dimitriadis 1) to allow for better preservation, conservation, and restoration of the world’s dying languages.

Dimitriadis, Alexis et al. *The Use Of Databases In Cross-Linguistic Studies*. Mouton De Gruyter, 2009.

Garofolo, John S., et al. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium, 1993.

Krauss, Michael. "The world’s languages in crisis." *Language* 68.1 (1992): 4-10.

Moseley, Christopher (ed.). 2010. Atlas of the World’s Languages in Danger, 3rd edn. Paris, UNESCO Publishing. Online version: [http://www.unesco.org/culture/en/endangeredlanguages/atlas](http://www.unesco.org/culture/languages-atlas/en/atlasmap.html)

Staczek, John J. "PARTICULAR TREASURES: ON BEING AND BECOMING A SPEAKER OF ENGLISH" *Proceedings of the Georgetown University Round Table on Language and Linguistics* (1987): 1-15. Online version: <https://repository.library.georgetown.edu/bitstream/handle/10822/555480/GURT_1987.pdf>

**(600) Describe the methodological approach you will employ to carry out your proposed research:**

Since the project’s inception, much of the setup work has already been completed. What is left to do is to improve existing features, and to program more useful ones: cataloguing, database depth (the ability to store an infinite level of “sub”-languages derived from the parent), cross searches, comparison searches, and other similar features.

Since an important part of the project is the involvement of external linguists, it will not be useful to postulate any more features without these experts’ input.

On a more technical level, most of the work revolves around the Design Cycle, which in short follows these broad steps:

1. Identify a specific feature that needs to be added, tweaked, or otherwise edited
2. Specify the details needed for such a feature, the more specific the better
3. Create the feature, following the design specifications as a roadmap
4. Test the feature, using an appropriate suite of tests to assess its robustness

Much of this process involves repeated consultation with the linguistic experts and my mentor. Although the third step, creation, is generally what people expect of a coding project, the other three are equally as important.

Identification happens primarily on the part of the linguist, who identifies a need in their research that requires a separate feature be added to Terraling. For example, a linguist might need to repeat a search many times while making minor changes to the search parameters, and requests a feature for editing search parameters after a query.

Specification requires collaboration between the linguist and the programmer, and is the step where the details of the feature are hammered out. If the repeat-search feature is implemented, would it conserve or discard the previous search? Should we instead create a history of searches for more flexibility instead of storing only one? Should the search history persist through sessions or be discarded? How could someone ‘break’ this feature?

The creation step involves the actual creation of the feature, the steps for which vary from feature to feature. For the search-repeat feature, the programmer would need to find a way to store the previous search parameters, store the parameters in the browser cache, access the cache again should the feature be used, gracefully discard the current search, and generally make the process occur as seamlessly as possible.

Finally, the testing stage simply requires stress-testing the feature to ensure it conforms to the outlined specifications, and that it does not display unintended behavior. Pressure points might include ensuring nobody could use the repeat search feature to inject unwanted commands into the database, compromising user security, or to perform so many database requests as to overload the server.

**(150) Describe your timeline for completing this research, including the project's start and end dates and estimated number of hours per week dedicated to the work:**

The project requires constant maintenance, in addition to the development of new features. It is currently hosted on AWS (Amazon Web Services), and is running on fairly outdated dependencies. This means that catching bugs and ensuring smooth operation may take longer than on a contemporary web application running on contemporary technologies. My involvement on the project has already started; I began working on October 1st, and so far I have dedicated on average 10 hours per week on the project.

The project is not expected to have any conclusive end date. To be a useful application, development needs to be ongoing, with new features added as its users -- the linguists -- find the need for them.

**(150) How is your project relevant to your academic interests and goals?**

As a Computer Science major and French minor, I am uniquely interested in the intersection of technology and linguistics. This passion for language was sparked by my growing up in a multilingual household, where three or more different languages wove themselves into conversations with ease. Different languages rely on different structures and constructions, and so lend themselves to different modes of thinking. By letting languages die out, then, are we losing an entirely unique paradigm of thought?

On a more practical level, this love for language also led me to computer programming at a young age; a match that resulted in my pursuing a Computer Science degree. Terraling will allow me the unique practical experience of working on a live web app with experience developers and with clients who have specific and detailed needs.

**(200) Describe your relationship with your project mentor, addressing the following points: How did you identify your project mentor? In what capacity did you work with your mentor in developing your research project? How will you work with your faculty mentor on this project? How often will you meet?**

Professor Shasha was recommended to me by several classmates as a professor with an extensive background in research, a fondness for introducing students to said research, and a uniquely warm personality. Although Professor Shasha is the lead developer on the Terraling project, I will work mainly with another developer who will then report to the professor.

Because web development is far less physical than lab work, most of the work on Terraling will be done remotely through code. However, I will meet with Professor Shasha every Monday, both to report further developments on the Terraling project and to discuss any other changes we need to make to the application.