

EXPANSION OF MINI BULEKU: A MANCHU DICTIONARY

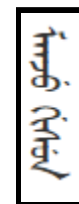
[Redacted]

Undergraduate Research Opportunities Program Proposal

Expansion of *Mini Buleku*: A Manchu Dictionary

I. Objective: The goal of this project is to expand *Mini Buleku* (henceforth *MB*), an online dictionary of the endangered Manchu language. The first and only Manchu dictionary of its kind, *MB* is unique for including publicly available, authentic recordings of native speakers pronouncing the word in each entry. By featuring recordings of this nearly-extinct language, *MB* has tremendous cultural benefits for the Manchu community and those learning the language—as well as research benefits for the linguistics community. *MB* was supported by 2021-2022 UROP and SURP funding, and due to the immensely positive reception of this resource among the Manchu community—as well as their increased demand for more content—I am applying for more funding to effectively double the size of this dictionary.¹

II. Background: Manchu, otherwise known as *manju gisun* (script: right image), is the language spoken by the Manchu people—an ethnic minority in China originating from the Northeast region known as Manchuria. Once the official language of the Qing Dynasty (1636-1912), the number of speakers of Manchu have significantly and drastically declined since the early 20th century (Gorelova 2002). Now, the number of proficient native speakers in China is around 500 individuals. Moreover, very few of these individuals can read and write in the traditional Manchu script, though there is a recent increase in the number of younger individuals learning the script (Zikmundová 2013; personal communication with members of the Manchu community). Due to both the sharp decline in native speakers and extremely low rates of proficiency in the written language, the Manchu language is categorized by the UNESCO Atlas of the World’s Languages in Danger as being critically endangered—with an increasing risk of becoming extinct within the next few generations.



This language is relatively unknown to people outside of the community—even unknown to the mass majority of Chinese people—but it is still nevertheless an important piece of cultural heritage. A language is a culture’s unique way of perceiving the world and interacting with it, so the loss of the Manchu language would equate to the loss of their unique perception of the world and thus a significant part of the Manchu culture. It would also be a loss to the world’s linguistic diversity in general, as it is one of the last living members of the Tungusic language family of Northeast Asia. Finally, it is important to mention that owing to its status as an official language of the Qing Dynasty, Manchu served as “the vehicle of an enormous translation literature”. In fact, about one-fifth (~2,000,000 items) of documents from the Qing Dynasty were written in Manchu (Elliott 2001). Documenting and revitalizing the language will uncover much more cultural knowledge about China during the Qing Dynasty. Not only that, documenting and revitalizing the language will contribute to the longevity of the Manchu culture and its traditions.

¹ Due to travel restrictions and safety concerns due to COVID-19, the entirety of this project has been—and will continue to be—conducted remotely. The native speakers with whom I work with are in mainland China, with a few in Hong Kong, Taiwan, and Japan.

III. Completed Work: While there are print dictionaries out there, there were no public online, recorded dictionaries featuring native speakers of Manchu—until the development of *MB*. I touched base with speakers of Manchu online in early 2021, and we began laying the groundwork for *MB* by developing a handful of digital resources. These resources can be found on my personal site (<https://bit.ly/39uYdAy>). Around the middle of 2021, I was accepted to the 2021-2022 UROP grant (09011s1) as well as the 2021 SURP program (57105s1)—both of which funded the creation of *MB*. The proposed dictionary was originally meant to be hosted on *Living Dictionaries*—a platform that hosts various other online, recorded dictionaries. We still deployed the dictionary to *Living Dictionaries* (<https://bit.ly/3F6tEgo>), but I decided it would be more beneficial to the Manchu community to develop a site that is solely dedicated to their language. Thus, at the end of the grant period, I coded and published *MB* (<https://minibuleku.github.io/>)².

From the start of the UROP/SURP grant period to January 2022, a total of 2660 entries (and audio recordings) were created for the dictionary. Besides including audio recordings, *MB* features both Mandarin and English translations, which makes it a unique resource given that a majority of the dictionaries for Manchu are written only in Mandarin. Also unlike other previous dictionaries, *MB* displays the actual words as they are written in the traditional Manchu script; other dictionaries only display the words in English letters. Finally, this resource is unique in that it features transcriptions of each word in the International Phonetic Alphabet (IPA). The IPA is an alphabet with letters representing all the sounds in human language, and it is primarily meant for linguists and other people working with world languages. Individuals not familiar with Manchu, but familiar with the IPA, can read and pronounce the words of the dictionary by referring to these transcriptions. I discuss the benefits of these features to the Manchu community and linguists in **VI. Cultural and Linguistic Benefits**. Given the benefits of this dictionary—as well as the increased demand among the community for the native speakers and I to add more entries—my hopes are to utilize UROP funding to expand this dictionary by 50%, equating to roughly 1,500 entries and audio files.

² It is important to note that the speakers I collaborated with are native speakers of a dialect of Manchu known as Sibe. *MB*, being a resource created for use among the community of speakers and learners, is referred to as a “Sibe” dictionary. The terms “Manchu” and “Sibe” for the most part are used interchangeably when referring to the language with which I am working.

Figure 1: Sample entry in MB

The screenshot shows the 'Mini Buleku' dictionary interface. At the top left is a logo with the Sibe word 'buluNkw' written vertically. The main title is 'Mini Buleku (米尼詞典)' with the subtitle 'A RECORDED SIBE DICTIONARY'. A search bar on the right contains the text 'mirror' and a close button. Below the search bar is a table of search results for the term 'mirror'.

Audio	Spoken	IPA	Written	Script (Text)	Script (Image)	Mandarin	English
	buluNkw	buluŋkw	buleku	𑖀𑖔𑖔𑖔		镜子	mirror

IV. Methodology: For all data collected and inputted to the online dictionary, I have been maintaining a spreadsheet, which I refer to as the “Master Spreadsheet”, that tracks the information for each entry on Google Drive. Each entry is given a numerical code, and an audio file named after that code is uploaded to Google Drive in a separate folder that is dedicated to the audio files. For the continuation of this project, I plan to continue using the Master Spreadsheet, shown in **Figure 2**. Creating new dictionary entries follows a 5-stage method, which I summarize in the following paragraphs as well as the timeline in **Table 1** (on pg. 5).

Figure 2: Sample of Master Spreadsheet

The screenshot shows a Google Sheet titled 'Master Spreadsheet'. The table contains the following data:

	A	B	C	D	E	I	J
	Entry	Spoken	IPA	Written	Script	Mandarin	English
1							
2	2	aba	aba	aba	𑖀𑖔𑖔	打围	hunt, roundup hunt
3	3	acem	aɕɕum	acimbi	𑖀𑖔𑖔𑖔	驮带	to pack, to load
4	4	acem	aɕɕum	acambi	𑖀𑖔𑖔𑖔	合	to meet, to join
5	5	acenem	aɕɕunum	acanambi	𑖀𑖔𑖔𑖔𑖔	对	to match; to fit; to suit
6	6	acevem	aɕɕuvum	acabumbi	𑖀𑖔𑖔𑖔𑖔	使合	causative form of acam (= acambi)
7	7	achen	aɕɕun	aciha	𑖀𑖔𑖔	驮物	items carried on an animal
8	8	adaky	adaki	adaki	𑖀𑖔𑖔	邻	adjacent, next to, neighboring
9							

Stage 1—Data preparation: For this stage³, I develop the list of words that will eventually be translated into Manchu and displayed on the website. To do this, I will consult a resource known as the [Master Elicitation List](#), which was developed by the same organization that developed *Living Dictionaries*. This list features around 4,000 commonly-used words found across languages, and all of these words are in English. Hoping to add around 1,500 entries to the dictionary, I will choose a selection of 1,500 words with input from the Manchu speakers, translate these words into Mandarin Chinese, and input them into a new, separate spreadsheet. I will then send this spreadsheet to a few native speakers, and ask them to translate these words into Manchu, record the audio files, and provide the additional data for each entry (cf. **Figure 1** for all of the relevant data points).

Stage 2—Data collection: This stage is primarily dedicated to giving the Manchu speakers time to input the relevant data on the spreadsheet and record the audio files. During my last time of collecting data for *MB* as part of UROP/SURP, there was a lot of back-and-forth between the native speakers and I, where we discussed how to effectively translate each word—among other issues. I expect this process to be the same this time around, so that is why I am dedicating much of the research period to this stage (also see note below **Table 1**). The speakers, when done with the spreadsheet and audio files, will send them back to me.

Stage 3—Data processing: Following my receiving of the new data spreadsheet and audio files from the native speakers, I will input the data from the new spreadsheet into the Master Spreadsheet, and upload the new audio files to the Google Drive folder reserved for these files. For each entry, I will also obtain the IPA transcriptions and traditional Manchu script forms using online software. For the IPA transcriptions, I will use a tool I developed known as [Manchu to IPA](#). I will use [this](#) externally-developed online tool to transcribe the entries provided by the native speakers into the traditional Manchu script, and these will be inputted into the Master Spreadsheet.

Stage 4—Data cross-checking: Once I am done processing the data in the updated Master Spreadsheet, I will send a copy of the Master Spreadsheet—abridged to only include the new data collected during this grant period—to the native speakers once again. For this stage, they will correct any errors observed that were made during the translation/data collection process. After I obtain their approval, I proceed to the final stage.

Stage 5—Date deployment: *MB* is designed in a way to allow immediate transfer of the contents from the Master Spreadsheet to the website. For this stage, I simply have to download the Master

³ Note that the native speakers of Manchu that I work with do not know English, so Mandarin Chinese is the language we communicate with and translate entries from.

Spreadsheet and upload to the website. All of the new additions—including the 1,500 entries and audio files—will be publicly available for use.

Table 1: Timeline of Project

Period	Progress
Summer 2022	<ul style="list-style-type: none"> ● Stage 1: Data preparation <ul style="list-style-type: none"> ○ Choose selection of 1,500 words from Living Tongues' Master Elicitation list ○ Input selection of words (and Mandarin translations) into new spreadsheet ○ Send spreadsheet of words to native speakers for translation
Fall 2022	<ul style="list-style-type: none"> ● Stage 2: Data collection* <ul style="list-style-type: none"> ○ Answer any questions native speakers have about translations ○ At the end of this stage, receive spreadsheet with Manchu translations, along with audio files ● Stage 3: Data processing <ul style="list-style-type: none"> ○ Upload audio files to Google Drive ○ Input data from translation spreadsheet into personal Master Spreadsheet ○ Retrieve script entries, IPA transcriptions
Winter 2023	<ul style="list-style-type: none"> ● Stage 4: Data cross-checking <ul style="list-style-type: none"> ○ Send Master Spreadsheet back to native speakers to review any errors ● Stage 5: Data deployment <ul style="list-style-type: none"> ○ Upload Master Spreadsheet to <i>MB</i>

* For reasons of brevity, I include this process under Fall 2022, when in reality it will span both Summer and Fall. I want to give enough time for the native speakers to input the data and for me to provide any support they may need.

V. Funding: I am requesting an amount of \$1,440, all of which will be going back to the Manchu community, and specifically the three native speakers I am working with to develop the dictionary. I am planning to pay \$16USD for each hour of their time. It is important to note that this is over the typical amount, but this is necessary to fairly pay the speakers for the time I anticipate them taking—all in the goal of effectively documenting this language and contributing to its preservation. An outline of the expenses is shown in **Table 2**.

Table 2: Requested Funding

Expense Name	Amount	Hours	Sub-Total
Correspondent Compensation #1	\$16USD/hr	30	\$480USD
Correspondent Compensation #2	\$16USD/hr	30	\$480USD
Correspondent Compensation #3	\$16USD/hr	30	\$480USD
Total Amount Requested:			\$1440USD

VI. Cultural and Linguistic Benefits: Following the creation of *MB*, the native speaker with whom I primarily collaborated stated, “Our dictionary has had great impacts on the Manchu world”. Similarly, the student of the previously quoted speaker stated that “It would be amazing if this dictionary can be expanded; there are no [other] recorded dictionaries for Manchu”. Expanding upon the community’s feedback of the dictionary, there are three audiences to whom the dictionary has—and will continue to—provide benefits and impact in a positive way. These three audiences and types of benefits are outlined in (1)-(3),

- | | | |
|---|---|------------------------------|
| (1) Manchu community members who speak their language | } | Cultural benefits |
| (2) Manchu community members do not speak their language | | |
| (3) Linguists and other people interested in the language | } | Linguistic/research benefits |

In terms of (1), it was already mentioned that very few members of the Manchu community can write in their language. The dictionary, by featuring entries in the traditional Manchu script, provides a reference tool for members of the Manchu community to learn the script. For (2), the audio recordings and translations can be referenced by Manchu people who cannot speak their language—as a way to learn the language and contribute to its preservation. The audio recordings are also especially beneficial for the population described in (3). There has been very little research done on the sounds of the Manchu language, and that is why these authentic recordings provide valuable acoustic information for analyses of the language as it is spoken. In fact, I plan to use the recordings in *MB* to perform such an analysis of the language, with the

hopes of producing a descriptive paper of the sound system of Manchu. After all, the general goal when doing linguistics work is to understand how human languages work, and this is done not by looking at a small subset of languages (such as English and other European languages), but by drawing from a full range of languages from all over the world. Manchu, as one of the last surviving languages of the Tungusic language family, has the potential to reveal more about how human languages work—especially within the context of sounds.

With 2660 entries, *MB* has already contributed to ongoing efforts in the Manchu world to document and revitalize the language. As numerous as the entries are, there is still a growing demand and pressing need to expand this resource. With the remaining (mostly elderly) speakers aging and an increasing shift to Mandarin Chinese, it is our hope—both mine and that of the native speakers with whom I am collaborating—to expand this valuable resource and contribute to the longevity of the Manchu language.

VII. Selected References:

- Anderson, G., & Daigneault A. (2021). Living Dictionaries: An Electronic Lexicography Tool for Community Activists. In *Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference*. 5–7 July 2021, virtual. Brno: Lexical Computing CZ.
- Elliott, M. (2001). The Manchu-language archives of the Qing Dynasty and the origins of the palace memorial system. *Late Imperial China* 22(1), 1-70.
- Gorelova, L. (2002). *Manchu Grammar*. Handbook of Oriental Studies. Section Eight, Central Asia, vol. 7. Leiden ; Boston: Brill.
- Kodner, J. A, & Meng, R. (2021). *Mini Buleku: A Recorded Sibe Dictionary*. Retrieved from <https://escholarship.org/uc/item/3kh1q0g4>
- Moseley, C. (2010). *Atlas of the World's Languages in Danger*. 3rd ed. Paris: UNESCO Publishing.
<http://www.unesco.org/culture/en/endangeredlanguages/atlas>.
- Norman, J. (2013). *A Comprehensive Manchu-English Dictionary*. Harvard University Asia Center.
- Zikmundova, V. (2013). *Spoken Sibe: Morphology of the Inflected Parts of Speech* (Vol. 55060). Karolinum Press.

VIII. Links:

- Personal website
<https://jkodner.weebly.com/my-projects.html>
- Manchu page on *Living Dictionaries*:
<https://livingdictionaries.app/sibe/entries/list>
- *Mini Buleku*:
<https://minibuleku.github.io/>
- Master Elicitation List:
https://livingtongues.org/wp-content/uploads/2020/05/TOOLKIT_Master_Elicitation_Spreadsheet_V1.pdf
- *Manchu to IPA*:
<https://jkodner.github.io/ManchuIPA.github.io/>
- Script tool:
<http://www.anakv.com/eng/>