The added words are split into three categories. First are named entities (mostly locations) taken from a board game set after the fall of the Durrani empire (1823), which spanned present day Afghanistan, parts of Pakistan, Iran, Turkmenistan, and India. Second are scientific/latin bird names. Third are named entities from a novel. These names seem Polynesian-inspired, but since the novel is secondary world (not a fictionalised version of the real world), the baseline model could only have knowledge of the entities by coincidence. The table on the next page contains the words, their respective categories, and the model output for one test. The model's output varies somewhat between tests.

There is generally good performance on the Afghan words. Presumably, many or most of these already existed in the baseline model's vocabulary, but it still had a hard time with many of them pre-training. Barring some small differences, the only big mistake post-training was for *Qanat*, which could not overcome the presumably very high probability the underlying language model has for *cannot*.

Performance on the latin birds is not great even post-training. None are entirely correct, and *Butorides Virescens* gets a full (nonsensical) English phrase instead. While the words in the training data for our modified model is put on the radar by the training, they clearly do not get a (much) privileged position in our implementation.

The fictional-entity category is also not great. The baseline language model probabilities seem to outweigh the new words quite often (*oh no ma* for *uunuma*, *paolo* for *paiolo*, *end*, for *enderal* for example). On a relisten, test audiofile clearly has the extra syllibles for, for example, *enderal*, but *end* clearly has too high weights in the model to be overcome.

It should be noted that the model training was performed by uploading a text file with the desired words, as described in a lab session. In cases where the target learning words overlap even partially with common English words, it might be necessary to upload audio data, phonetic transcriptions, or some other more robust training data.

Word	**Category**	Model output
Herat	Afg	Herat
Transcaspia	Afg	Transcaspia
Wakhan	Afg	Wakhan
Durrani	Afg	Durrani
Qanat	Afg	Cannot
Ghilzai	Afg	Gilzai
Baloch	Afg	Baloch
Nasrullah	Afg	Nasrallah
Khyber	Afg	Khyber
Farah	Afg	Farah
Butorides Virescens	Bird	But the reader’s free rescinds
Zenaida Macroura	Bird	Zenaida macora
Podilymbus Podiceps	Bird	Portalimbus Podiceps
Mycteria Americana	Bird	Magteria americana
Enderal	Fict	End
Uunuma	Fict	Oh no ma
Paiolo	Fict	Paolo
Maitepo	Fict	Mai temple
Maitemi	Fict	My tammy
Uunili	Fict	U unile
Lehomai	Fict	Le ho mai
Kilay	Fict	Kilay
Nehrimese	Fict	Nareemies
Qyranian	Fict	Kiranian
Vyn	Fict	Vin
Al-rashim	Fict	Al rashim
Naka	Fict	Naka
Makehu	Fict	Make