Exploration Project 2: Blocks for ML

Learn with Buki: Spanish grammar for kids

Submitted by **Amanda Tran, Jyoti Poonia, and Veda Lakshminarayanan**

EDUC 5911 Artificial Intelligence for Children and Youth: Learning, Creating, and Understanding

---

Demo Video: *In this video, Veda demonstrates a correct and an incorrect question-answer interaction with the Buki device—1. Identifying a noun, and 2. First incorrectly classifying an adjective as an adverb, and then raising the correct sign for adjective.*

Download for PRG: Buki v1.3.5_Final.sb3

Original Inspiration: Bilingual Grammar Aid for Parts of Speech

1. Project Description

This project aims to exercise elementary aged learners' knowledge of parts of speech in Spanish. Inspired by a pre-existing grammar aid comprised of animated characters correlating to each part of speech, we incorporated the RAISE *sprite* as a guide for categorizing words by part of speech. We integrated a Teachable Machine (TM) image-based model, having been trained with data to classify visual signals as parts of speech, e.g. *noun, verb, adjective, adverb.* The RAISE model presents a sentence, *El perro marrón duerme tranquilamente.* (*The brown dog sleeps calmly.*) which contains a noun (perro), adjective (marrón), verb (duerme), and adverb (tranquilamente). We created tactile signs for each part of speech: red square for noun, yellow triangle for adjective, green circle for verb, and blue pentagon for adverb. In this order, the machine asks the learner to raise the correct signal for the highlighted word. If correct, the *sprite* will give affirmations like 'fantástico!' or 'buen trabajo'; if incorrect, the *sprite* will say 'qué? try again!', sometimes accompanied with hints.

2. Process

Our process unfolded in three stages—starting with an ambitious initial vision, to gaining key insights and learnings from exploration and subsequent refinements that shaped the final activity.

2.1 Initial Vision

Our initial vision for the project was to create a 'choose your own adventure' style story that required students to fill in blanks with correct Spanish vocabulary, powered by a voice classifier. Then, we thought to incorporate the grammar aid Veda had worked on previously (see *Figure 1*). With the new direction of teaching grammatical units, we decided to simplify the premise from storytelling to sentence-construction. Then, we drafted up three sentences to exercise all the different parts of speech. Ultimately, upon building the activity on RAISE Playground, we found that it may be best to present the student with *one* sentence, and have each question address a specific part of speech (see *Figure 2*). This was already an ambitious undertaking due to the sensitivity of the Teachable Machine model, which we will elaborate on in the upcoming section.

After scaling down on our initial idea, we began experimenting with ML blocks on RAISE. Following our plan, we were aiming to create different sprites matching different questions (linked to individual backdrops). With the sample sentence, each sprite, representing each part of speech, would ask "Where am I in this sentence?", then receive audio input of the learner saying the matching word from the sentence. The designated sprite would then respond adaptively to audio inputs from the learner, indicating whether student's audio input was correct.

Figure 1:

Bilingual Grammar Aid for Parts of Speech

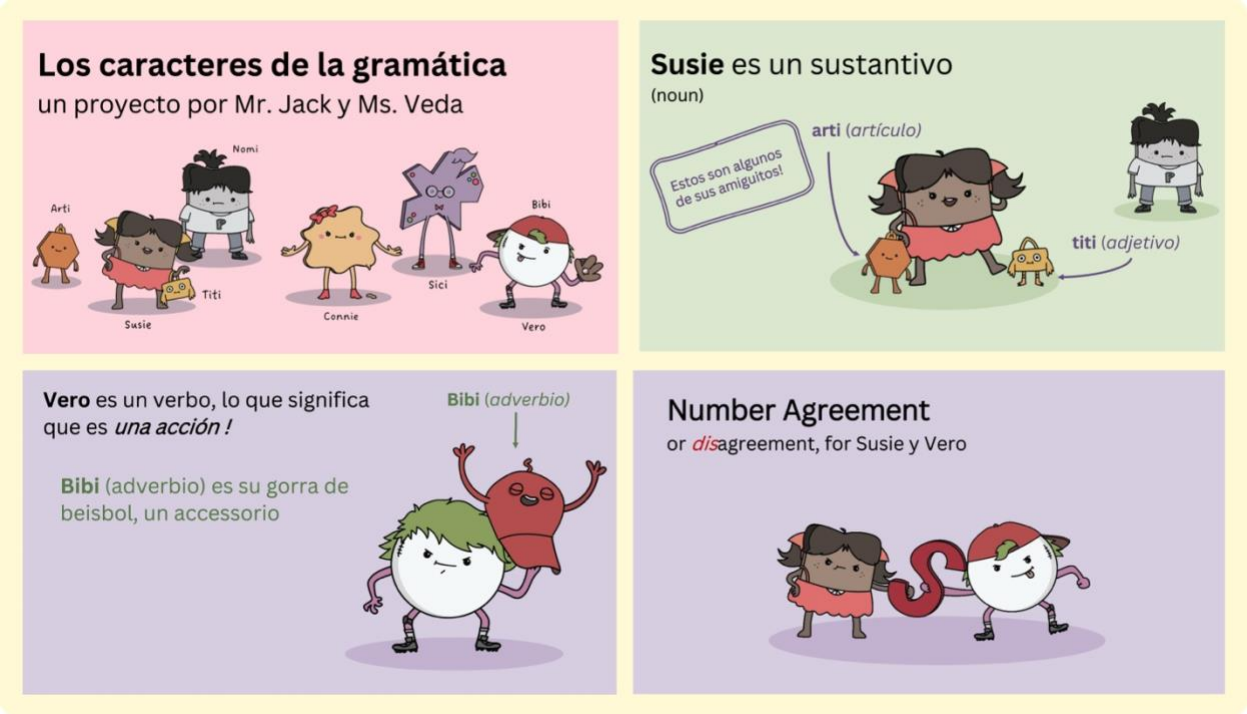
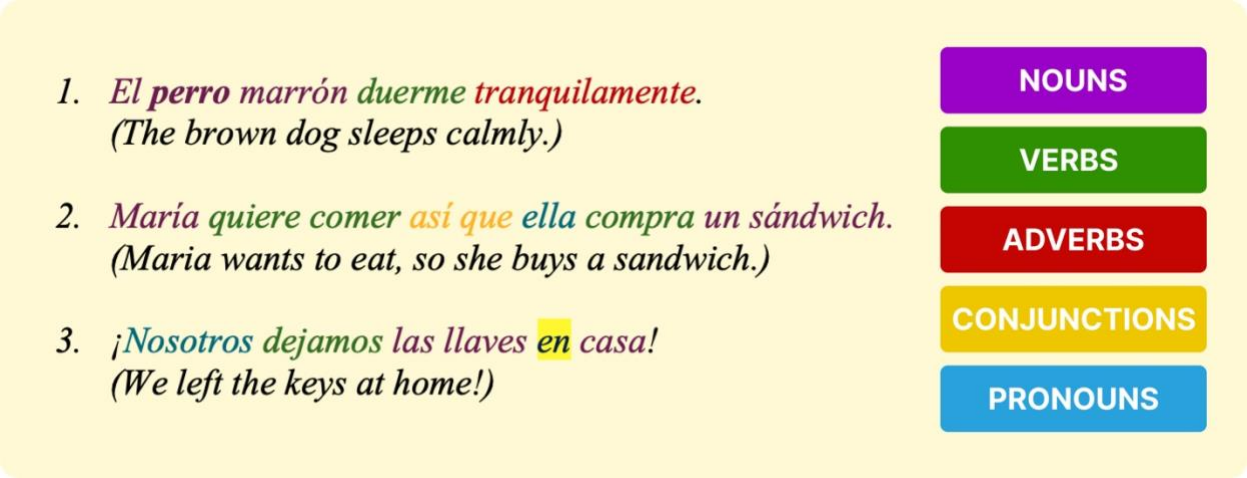


Figure 2:
Initially planned sentences



2.2 Key insights from exploration

After (attempting to) execute our initial idea on RAISE playground, we have come to see two biggest takeaways on the technical end that drove us to adapt our approach to this project.
First, we ran into an issue with isolating operations of multiple *sprites*. As illustrated in *Figure 3*, each *sprite* was programmed to simplistically handle learner's input through a conditional script. When we engage multiple *sprites* into the process, there was overlapping of *sprites*' response to learner's input, most evident in the chaos of multiple simultaneous audio outputs. In other words, the initial logic enabled all *sprites* to respond to inputs throughout the user "journey", rather than segmented to their designated question. Our exploration from tutorial videos or forums on the Internet gave some hints to hiding and muting *sprites*, but these solutions did not address the issue (see *Figure 4*). From this, we decided to centralize the responses to one *sprite* and utilize the costumes functionality instead.

Figure 3:
Sample of initial sprite logic

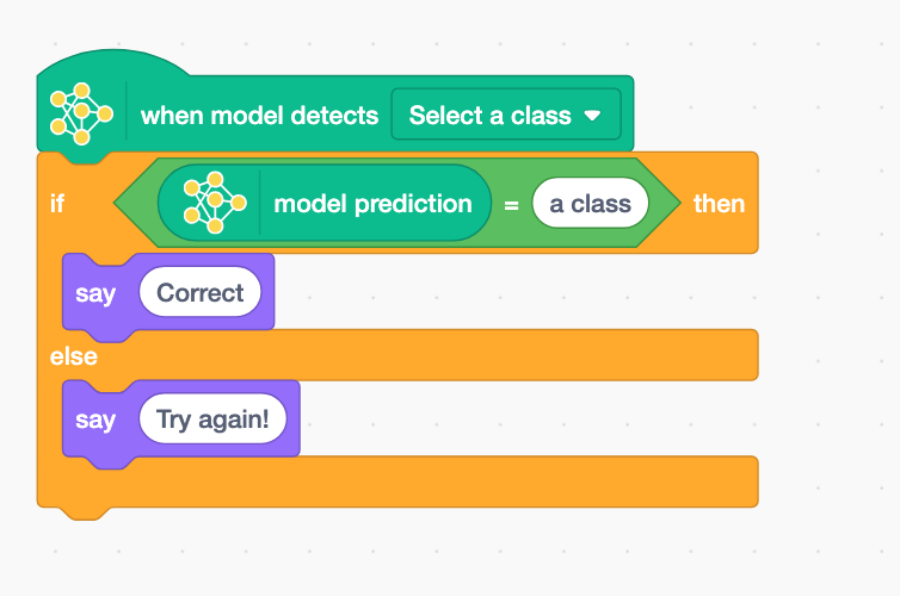
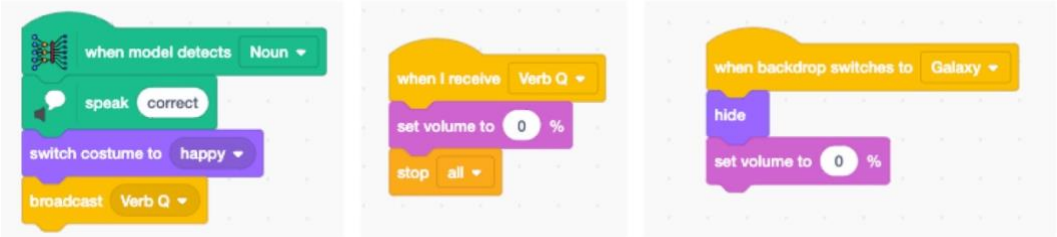


Figure 4:
Samples of trying to reduce volume overlap when one sprite changes to another



We also ran into obstacles with using audio input for our model. As we experimented with a TM-hosted audio classifier, there were issues with model sensitivity that we think would be exacerbated in a practical setting with children talking simultaneously and background noise. It was also difficult to control for model accuracy: as we have learned from our experience programming with TM, the confidence level could fluctuate with live testing, and for transferring the model to RAISE we were unclear what confidence threshold has to be reach for a prediction to be "validated" (was it 75% and above, or a different threshold?). As a result, the model could incorrectly classify the learner's input and confuse the learner even more. With this consideration in mind, we decided to seek alternative modes of inputs like visual artifacts, which might make building the model more manageable, and consequentially make the application more scalable and learner friendly.

2.3 Revisions post exploration: the final activity

2.3a Content

We have decided on framing the lesson to still focus on teaching Spanish parts of speech, but instead of involving multiple characters to refer to these parts of speech, we grounded our approach to using one main character, Buki (see *Figure 5*) and maintained English terminology for the grammar. This is because the learning objective was not necessarily to immerse the learner in Spanish, but to recognize syntax. Eventually, this activity as demonstrated in *Figure 6*, would be effective for both L1 and L2 learners of Spanish of the target age.

Figure 5
Buki and his costumes

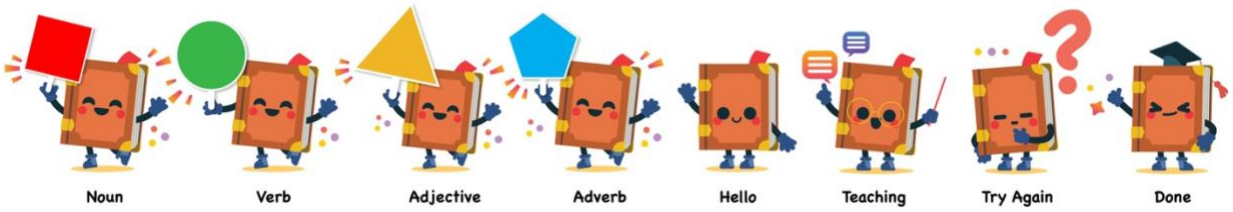
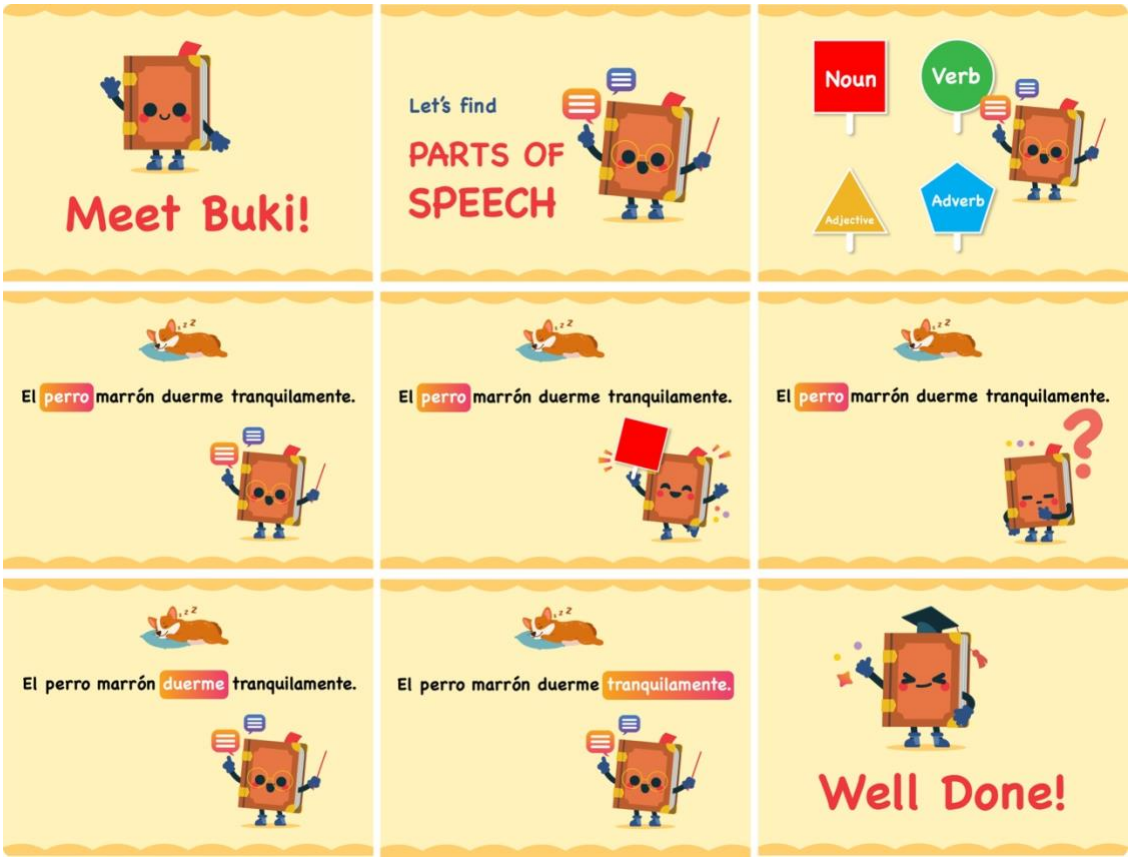


Figure 6
Story board for the activity



2.3b Model

Since we have switched our approach to using visual artifacts (i.e., signs), our final Teachable Machine model is an image classifier of 4 classes (see *Figure 7*): noun (represented by the red square), adjective (represented by the yellow triangle), verb (represented by the green circle), and adverb (represented by the blue pentagon). For each class, we entered roughly 230 samples, positioning each sign in various orientations, with our hands covering different parts. We made sure to hold down the 'Record' button while moving the sign toward and away from the camera. I also alternated between covering my face and having my face appear larger than the signs in these data recordings. We were particularly careful with the blue pentagon and yellow triangle, as we suspected the machine may confuse the two shapes. Additionally, yellow and red are not as distinct from human skin tone as blue and green, so we wanted to make sure we input enough imagery with my face near the *noun* and *adjective* signs so that it could distinguish those two.

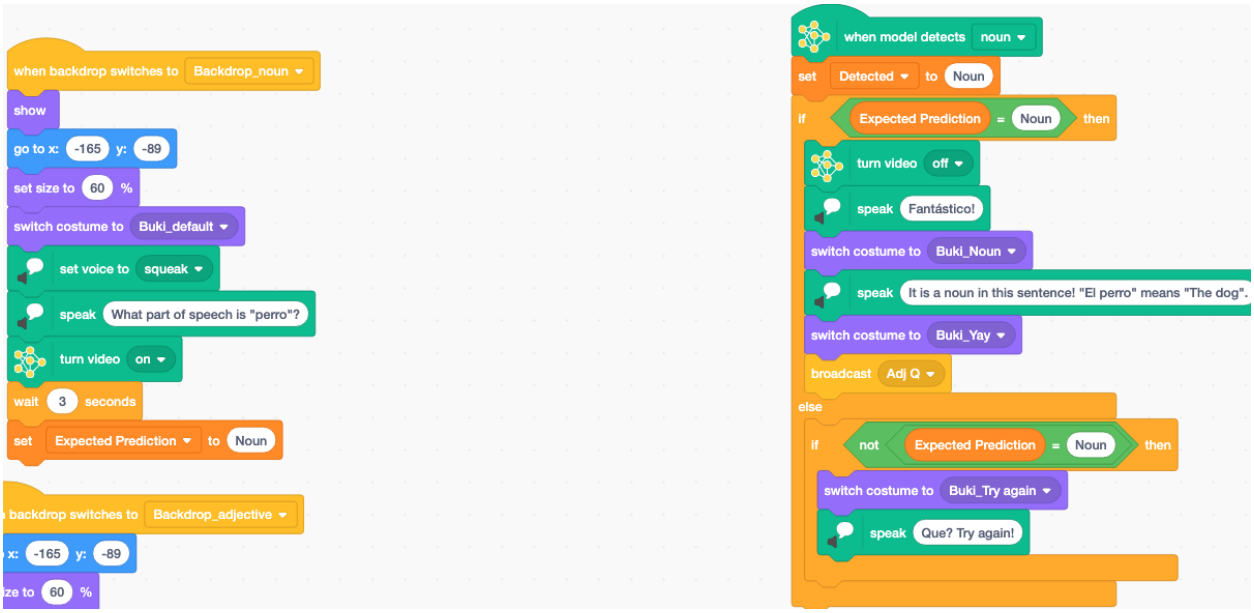
Figure 7
Training the teachable machine model

2.3c Project/application

For the application, we created 2 variables: 1 global variable “Detected” for the TM-predicted class of the visual input and a 1 variable local to our *sprite* Buki, titled “Expected”, to track the expected response. “Expected” is reset every time the learner moves to a new question, indicated by a change of backdrop.

The application is triggered by clicking the green flag. To begin, Buki is programmed to walk the learner through the instructions for the activity. Then, when a background is triggered for a question, Buki is programmed to announce the question. As the “Expected” variable is set, the camera is turned on (programmed this way to avoid sensitivity from TM live testing). The model then detects user input – accordingly, “Detected” is set and incorporated into the conditionals that determined Buki’s adaptive output. We made sure to include hints for when the learner answered incorrectly; in the future, this can even be tweaked so that students can involve teacher’s assistance.

Figure 8
Snippet of Buki's final program logic



3 Reflection on the learning

3.1 Data collection phase

In the data collection phase of our project, we considered various factors while training Teachable Machine for both audio and video inputs. Initially, we explored audio data but later transitioned to video data due to challenges in model learnability and sensitivity. While we faced learnability issues while training our model on audio data, the model trained on video data was very sensitive to live video inputs as the accuracy fluctuates rapidly with any movement. We hypothesize that this might be because we used test set with very close resemblance to training set (Tseng et al., 2024, p. 3), without a separate validation set, which may have influenced model performance. We tried to tackle this with the code phase with RAISE. This shift allowed for more consistent and reliable results. This is a particularly salient consideration when designing such interventions for children as we cannot expect them to use these activities exactly as intended. We recommend that such designs should be as immune as possible to discrepancies and bias that can be inherent to working with data. This limitation could be addressed by incorporating a validation dataset to better assess generalizability.

3.2 Working with Teachable Machines

While working with Teachable Machine, we observed significant fluctuations in the model's confidence levels due to the live video input. As shapes were moved in front of the screen, the model's accuracy varied dynamically, leading to inconsistent classifications. In some instances, Buki detected multiple possibilities simultaneously—one at 100% accuracy and another at a

lower confidence level, influenced by these fluctuations. This variability highlights the challenges of using real-time video input and underscores the need for stability in live classification scenarios.

3.3 Working with RAISE

A point of curiosity we had while experimenting with RAISE Playground, just as with other block-based programming platforms, was the intentionality of keeping the workspace so open and unstructured. We found that for our group members with less familiarity with block-based programming, it was more difficult to navigate the workspace without structure or, at least, functionality for code annotations.

In contrast with the Dialogue Flow Design of the Interactive Character, Elinor (Xu, et, al., 2024), our Buki follows a more scripted approach, emphasizing the role of a guiding figure, such as a parent or teacher, in the learning process. We felt it was important to have an involved mentor to support and scaffold children's understanding and incorporated this in designing the interactions with Buki. In real-world scenarios, for instance, if a child provides an incorrect answer, we envisioned Buki prompting a teacher for guidance. This design choice reflects our belief that AI-driven learning tools should complement, rather than replace, human interaction, fostering a more supportive and structured learning experience.

References

Tseng, T., Davidson, M. J., Morales-Navarro, L., Chen, J. K., Delaney, V., Leibowitz, M., Beason, J. & Shapiro, R. B. (2023). Co-ML: Collaborative Machine Learning Model Building for Developing Dataset Design Practices. https://doi.org/10.48550/arXiv.2311.0908

Xu, Y., He, K., Levine, J., Ritchie, D., Pan, Z., Bustamante, A., & Warschauer, M. (2024). Artificial intelligence enhances children's science learning from television shows.*Journal of Educational Psychology, 116*(7), 1071–1092. https://doi.org/10.1037/edu0000889