**IBALOI NLP LEXICON SYSTEM**

**USER MANUAL**

**A Digital Platform for Ibaloi Language**

**Preservation Through NLP**

**Prepared By:**

*NLP Team Mors3*

**TABLE OF CONTENTS**

**[Introduction 3](#_25keqridcsv7)**

**[Purpose of the System 3](#_qu6btfqsnrxy)**

**[Target Users 4](#_kj6yh5rrt165)**

**[System Requirements 4](#_a2ocr4qn4zae)**

**[How to Use This Manual 4](#_7cx29d2b8lgr)**

**[Figure Guide 6](#_jhxhpl3cujbm)**

[Figure 1. Ibaloi NLP – Home Page Interface 6](#_reffy37u8k0a)

[Figure 2. Ibaloi NLP – Home Page Search Interface 7](#_mu4n4q5fssm5)

[Figure 3. Ibaloi NLP – Research and System Features Section 8](#_2jvijnh8jk6m)

[Figure 4. Ibaloi NLP – Interactive Tools Interface 9](#_mr3hhf2x16vc)

[Figure 5. Ibaloi NLP – Lexicon Tab Interface 10](#_55qstwa079ux)

[Figure 6. Ibaloi NLP – Lexicon Search Result (Example: “adivay”) 12](#_78ryovofq5ll)

[Figure 7. Ibaloi NLP – Research Paper View 13](#_c8byjr2lpatw)

[Figure 8. Sentence Builder – Main Word Selection Interface 14](#_waw7vzxa7mvk)

[Figure 9. Ibaloi NLP – LLM Translator Tab Interface 15](#_xkm2c3z1wa2u)

[Figure 10. Ibaloi NLP – Lexicon Matching and Context Injection View 16](#_wlm4eibkpvm8)

[Figure 11. Ibaloi NLP – Reverse Translation (Ibaloi to English) Interface 17](#_whrl7jcjq5b1)

# Introduction

The **Ibaloi NLP Lexicon System** is a digital platform designed to support the preservation, study, and modernization of the Ibaloi language through Natural Language Processing (NLP). Developed as part of an academic initiative, the system integrates lexicon browsing, sentence construction, and machine-assisted translation to enhance accessibility for learners, researchers, and community members.

This User Manual provides step-by-step guidance for navigating the system’s features including the Digital Lexicon, Sentence Builder, Research Paper viewer, and the LLM-based Translation Module. Each figure in this manual is accompanied by an explanation to help users understand interface components, system functions, and expected outputs.

Whether you are a language learner, a researcher conducting linguistic analysis, or a contributor helping enrich the Ibaloi dataset, this manual will assist you in maximizing the use of the Ibaloi NLP platform.

# Purpose of the System

The Ibaloi NLP System was created to:

* Provide a structured and searchable **Ibaloi Digital Lexicon**
* Enable community participation through a **Sentence Builder** for example sentences
* Demonstrate NLP applications using a **LLM-based translation tool**
* Present the full research study behind the project in an accessible, web-friendly format
* Support language documentation, preservation, and modernization efforts

# Target Users

This manual is intended for:

* Students and researchers studying Ibaloi or related Philippine languages
* Educators teaching regional languages or indigenous studies
* Linguists conducting morphological or lexical analysis
* Community members who want to explore and contribute to the digital lexicon
* Developers and NLP practitioners evaluating low-resource language models

# System Requirements

To access the Ibaloi NLP System, users need:

* A modern web browser (Chrome, Firefox, Safari, or Edge)
* A stable internet connection
* (Optional) A GitHub account for viewing the project repository

No installation is required. The entire system runs through a web-based interface.

# How to Use This Manual

This manual includes:

* **Figures** illustrating each part of the system
* **Descriptions** explaining interface elements and functions
* **Step-by-step instructions** where applicable
* **Examples** for clarity

Each section corresponds to major features of the system, allowing users to follow along visually.

# Figure Guide



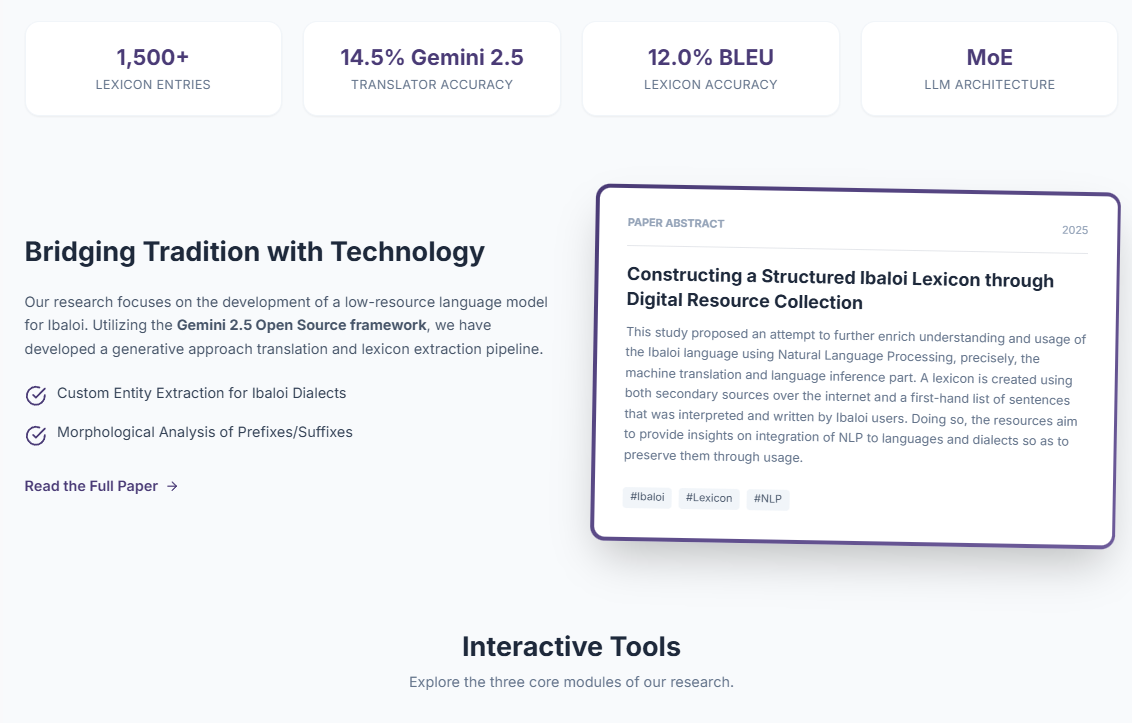
## *Figure 1. Ibaloi NLP – Home Page Interface*

This figure displays the main home page of the Ibaloi NLP System. It includes the system header with navigation links for **Lexicon**, **Research Paper**, **Sentence Builder**, **LLM Translator**, and **About Us**. The center section presents the project title *“Preserving Ibaloi through Neural Processing”* along with a search bar that allows users to look up entries within the Ibaloi Lexicon. Summary statistics—**1,500+ Lexicon Entries**, and **MoE LLM Architecture**—are shown beneath the search bar to provide users with an overview of system coverage and performance.



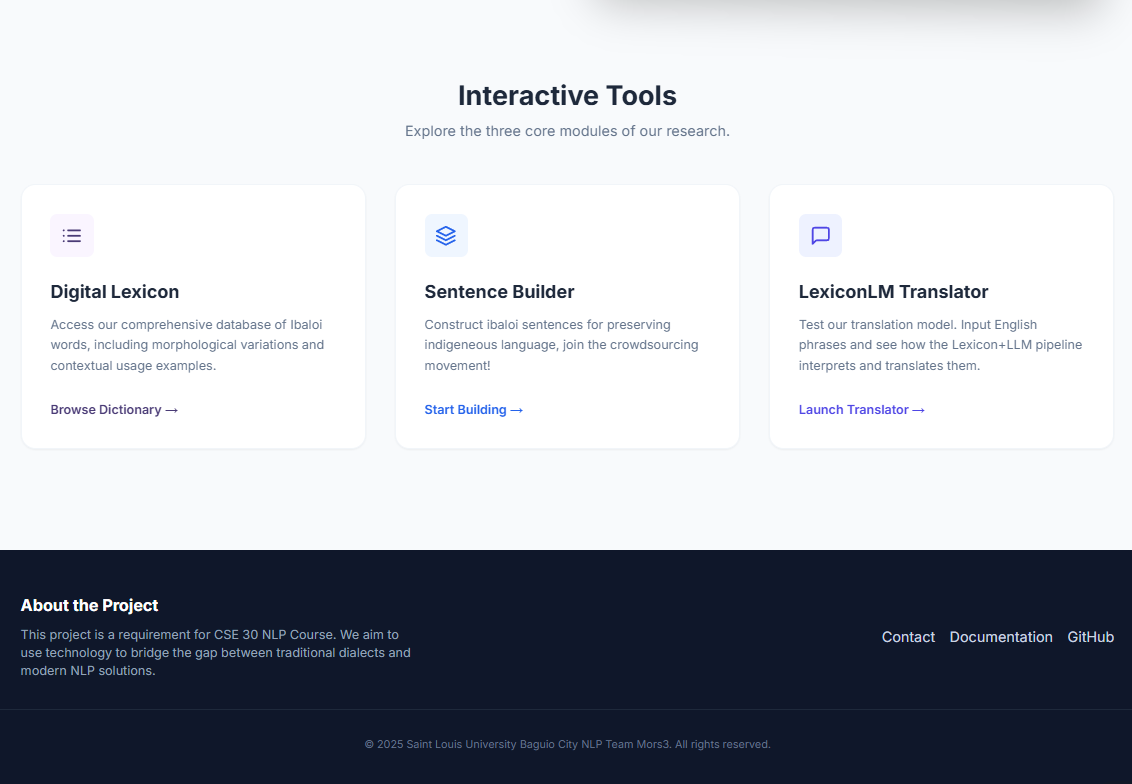
## *Figure 2. Ibaloi NLP – Home Page Search Interface*

A prominent search bar is positioned beneath the project title, allowing users to enter any Ibaloi word. When a query such as “adivay” is typed into the field, users can click the Search button to retrieve lexical results. The system is designed to automatically redirect the user to the Lexicon tab which will be shown in *Figure 6* to display the matched entry or entries.



## *Figure 3. Ibaloi NLP – Research and System Features Section*

This figure shows the research overview section of the Ibaloi NLP System. It highlights the key research focus areas, such as **custom entity extraction** and **morphological analysis** for Ibaloi dialects. A preview of the research abstract titled *“Constructing a Structured Ibaloi Lexicon through Digital Resource Collection”* is shown on the right side. The section also includes a link to read the full paper.

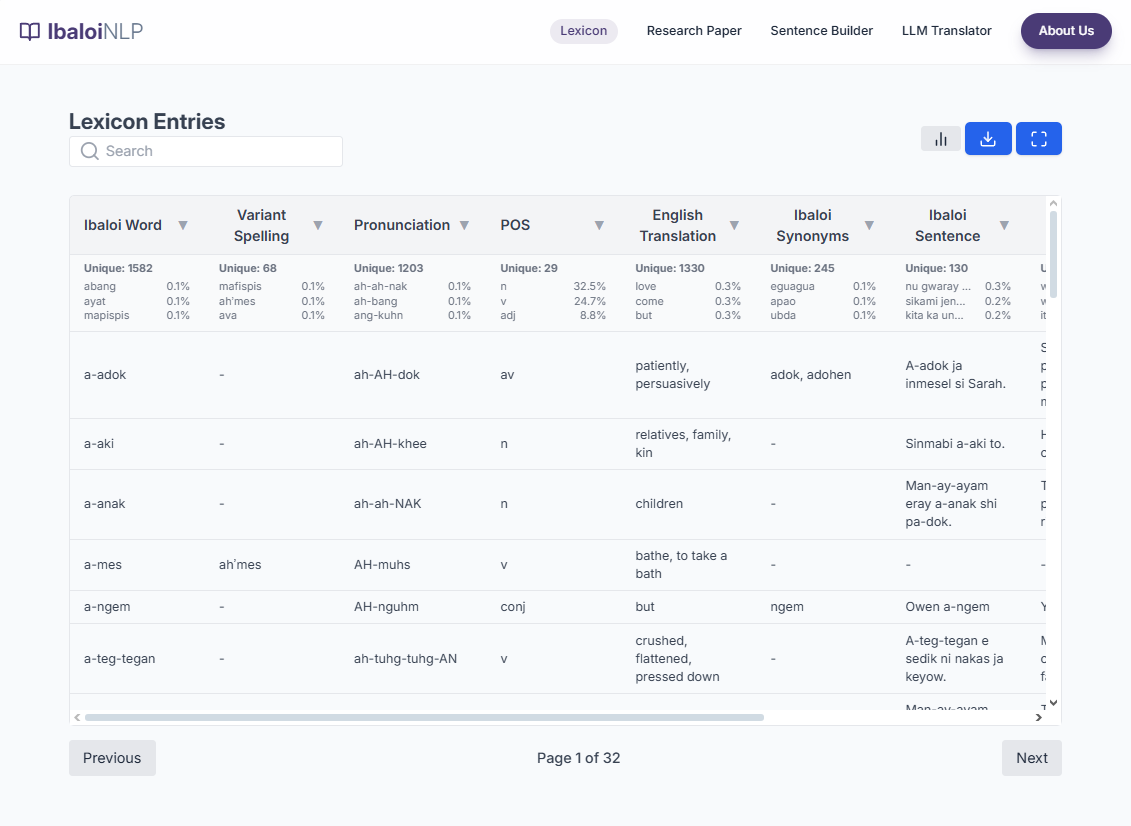


## *Figure 4. Ibaloi NLP – Interactive Tools Interface*

This figure displays the primary **Interactive Tools** section of the Ibaloi NLP system, which allows users to explore the three core parts developed during the research. Each part is presented as a distinct card with a descriptive title and a call-to-action link:

* **Digital Lexicon:**
  + **Description:** Provides access to the comprehensive database of Ibaloi words, including morphological variations and contextual usage examples.
  + **Action:** Users can click **"Browse Dictionary"** to begin exploring the lexicon.
* **Sentence Builder:**
  + **Description:** Engages the user in a crowdsourcing movement to contribute to the project by inputting and building Ibaloi sentences.
  + **Action:** Users can click **"Start Building"** to contribute data to the system.
* **LLM Translator:**
  + **Description:** Allows users to test the project's Machine Learning (ML) model by inputting English phrases and observing how the LLM pipeline interprets and translates them into Ibaloi.
  + **Action:** Users can click **"Launch Translator"** to begin testing the translation model.

Below the interactive tools, the interface includes an **"About the Project"** footer section which states the project's purpose: a requirement for the CSE 30 NLP Course, aiming to bridge the gap between traditional dialects and modern NLP solutions. This footer also provides contact information and links to the **Documentation** and **GitHub** repository.



## *Figure 5. Ibaloi NLP – Lexicon Tab Interface*

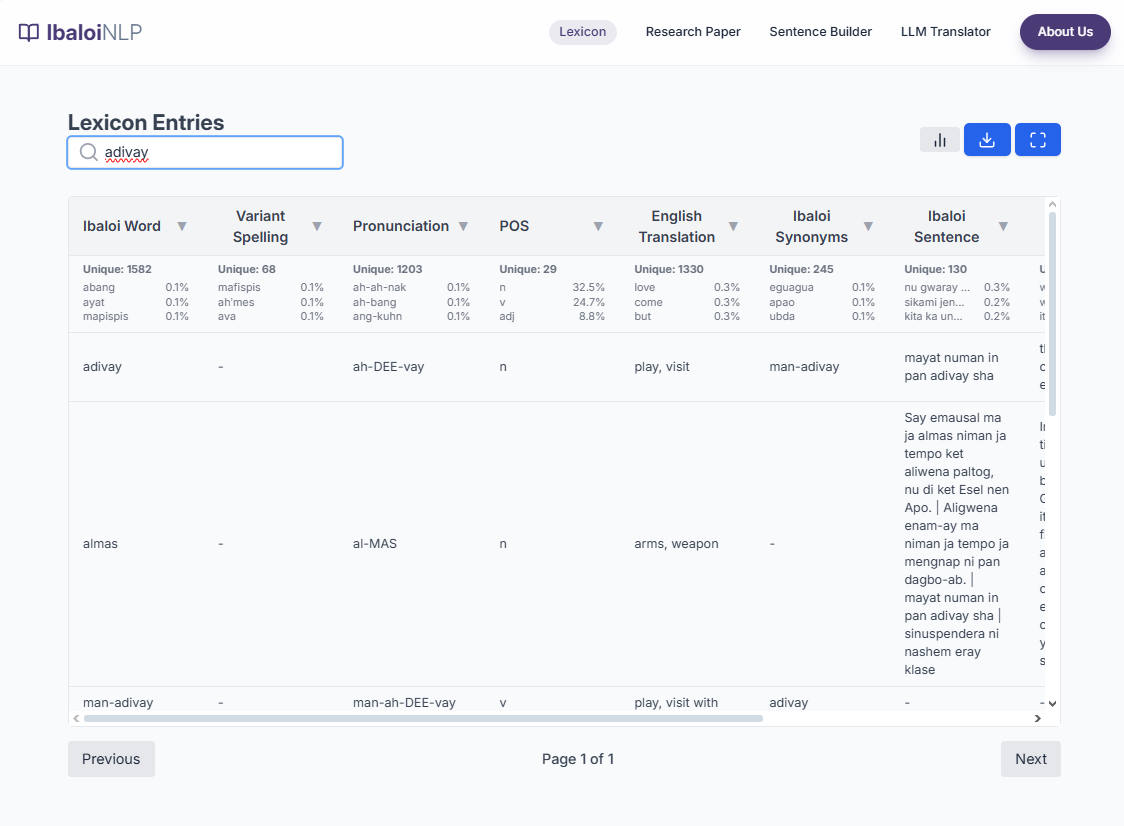
This figure presents the **Lexicon** tab of the Ibaloi NLP System, showing the full interface used for browsing and searching entries within the structured Ibaloi dictionary. The central portion of the interface features a **search bar** that allows users to look up specific Ibaloi words, spellings, or partial text queries. Once a term is entered, the lexicon table automatically filters results to display only matching entries.

Below the search bar is the **Lexicon Entries Table**, which is divided into several columns that provide detailed linguistic information. These include:

* **Ibaloi Word**
* **Variant Spelling**
* **Pronunciation**
* **Part of Speech (POS)**
* **English Translation**
* **Ibaloi Synonyms**
* **Ibaloi Sentence**
* **English Sentence**
* **Notes**
* **isNew**

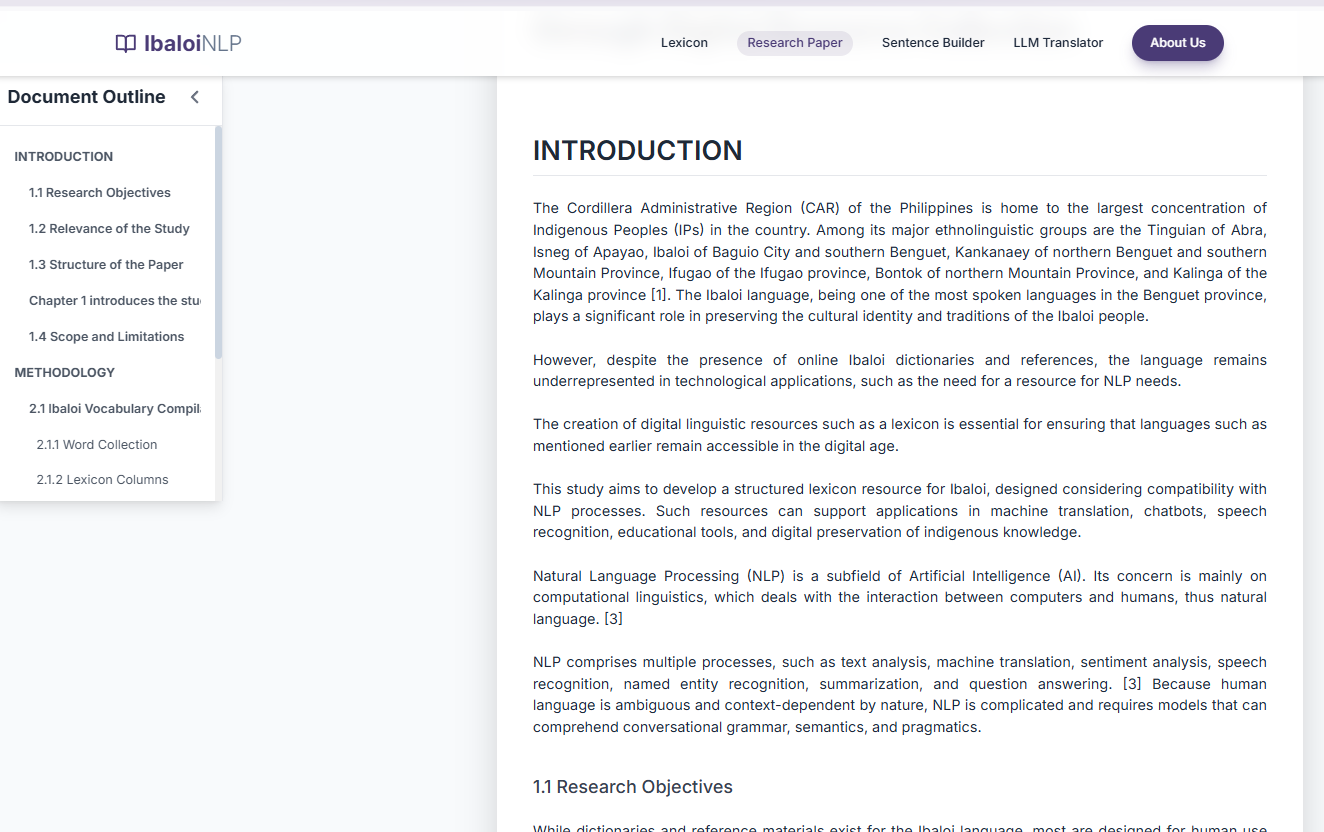
Each row in the table represents a single lexicon entry, enabling users to review pronunciation guides, grammatical categories, translations, and contextual usage. Scroll controls are available at the bottom, allowing navigation through multiple pages of lexicon data.

On the upper-right corner of the table, the interface includes three functional buttons:  
 (1) a **statistics icon** for viewing lexical frequency data,  
 (2) a **download icon** for exporting lexicon files, and  
 (3) a **fullscreen icon** for expanding the lexicon table view.



## *Figure 6. Ibaloi NLP – Lexicon Search Result (Example: “adivay”)*

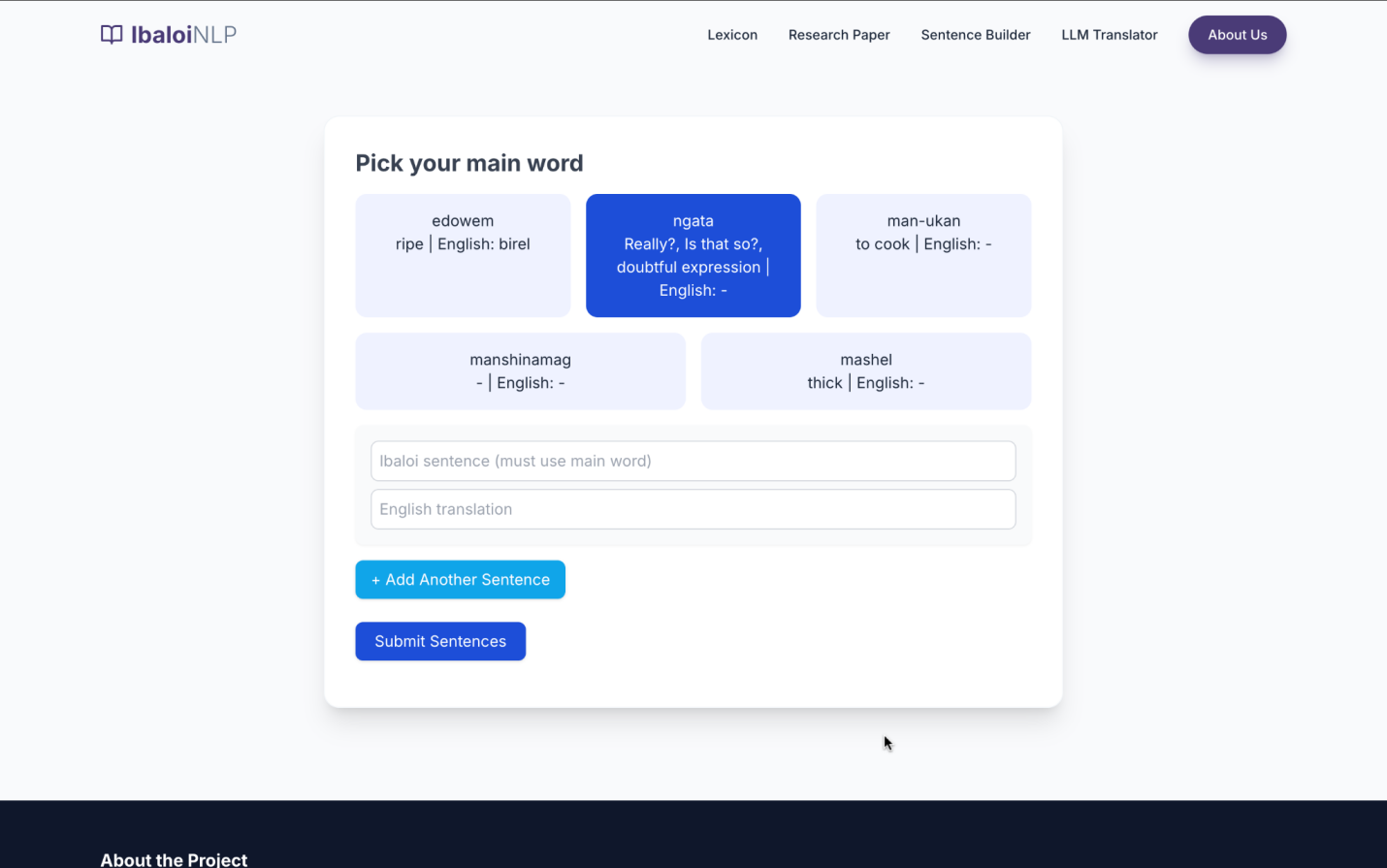
This figure illustrates the Lexicon interface after performing a search for the word **“adivay.”** Once the term is entered in the search bar, the system filters the table to show a single matched entry. The displayed row includes the standard fields, such as pronunciation (“ah-DEE-vay”), part of speech (noun), English translation (“play, visit”), and the synonym “man-adivay.” This filtered view helps users quickly locate detailed lexical information for a specific term without navigating multiple pages.



## *Figure 7. Ibaloi NLP – Research Paper View*

This figure displays the interface when a user navigates to the **Research Paper** tab (or the Research Paper link in the header). This view is dedicated to presenting the full academic paper detailing the project's methodology and findings.

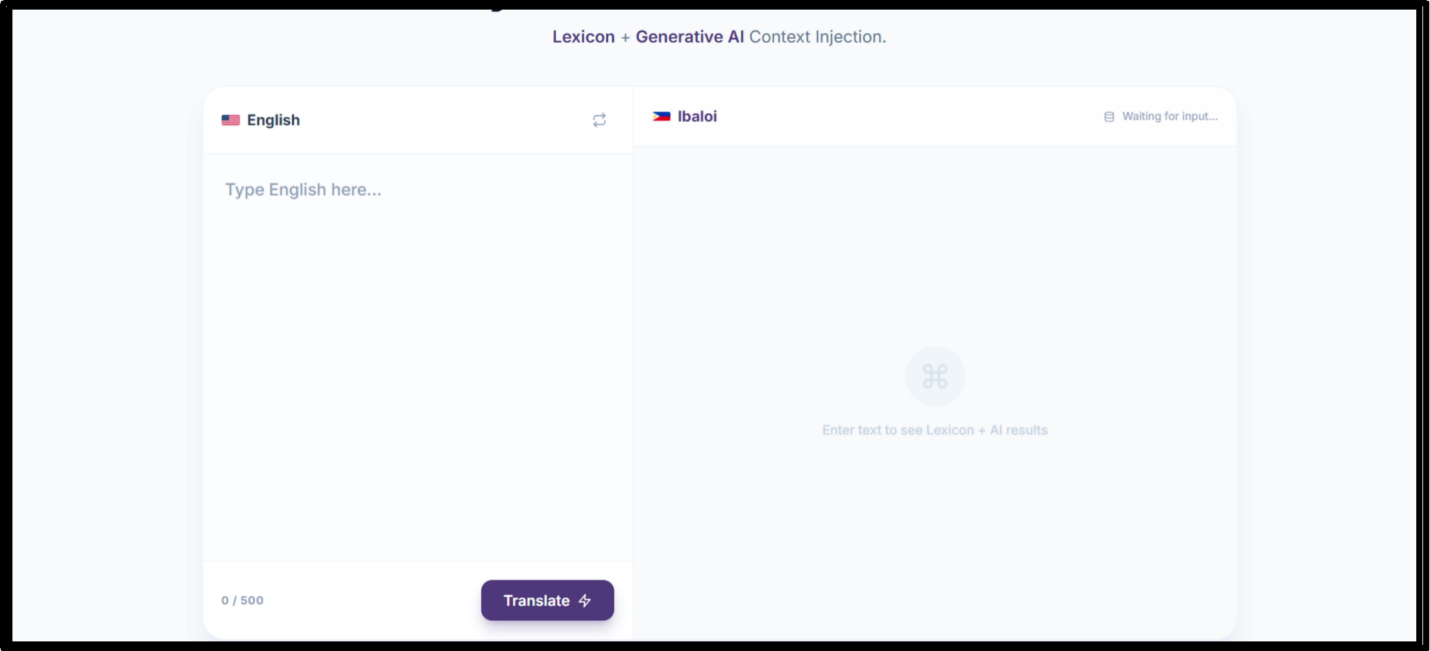
* **Header:** The standard navigation bar is maintained, allowing users to easily switch back to the other sections (Lexicon, Sentence Builder, LLM Translator).
* **Document Title:** The primary title of the research paper is prominently displayed: **“Constructing a Structured Ibaloi Lexicon through Digital Resource Collection.”**
* **Document Outline (Sidebar):** On the left, a vertical **Document Outline** is visible. This interactive table of contents allows users to navigate quickly through the paper by clicking on section headings (e.g., INTRODUCTION, 1.1 Definition and Domain, 2.1 Data Gathering).
* **Main Content:** The right and center portion of the screen is dedicated to the body of the paper. This view shows the introductory text, which explains the context of the Ibaloi language in the Cordillera Administrative Region (CAR), the problem of underrepresentation in technological applications, and the study's goal to develop a structured lexicon resource compatible with Natural Language Processing (NLP) processes.



## Figure 8. Sentence Builder – Main Word Selection Interface

This figure shows the **Sentence Builder module** of the Ibaloi NLP System. At the top of the interface, the user is prompted to **“Pick your main word”**, which is required for constructing valid Ibaloi example sentences. A selection of 5 main word cards is displayed randomly, including entries such as *edowem*, *ngata*, *man-ukan*, *manshinamag*, and *mashel*. Each card provides the word and, where available, its English translation or meaning.

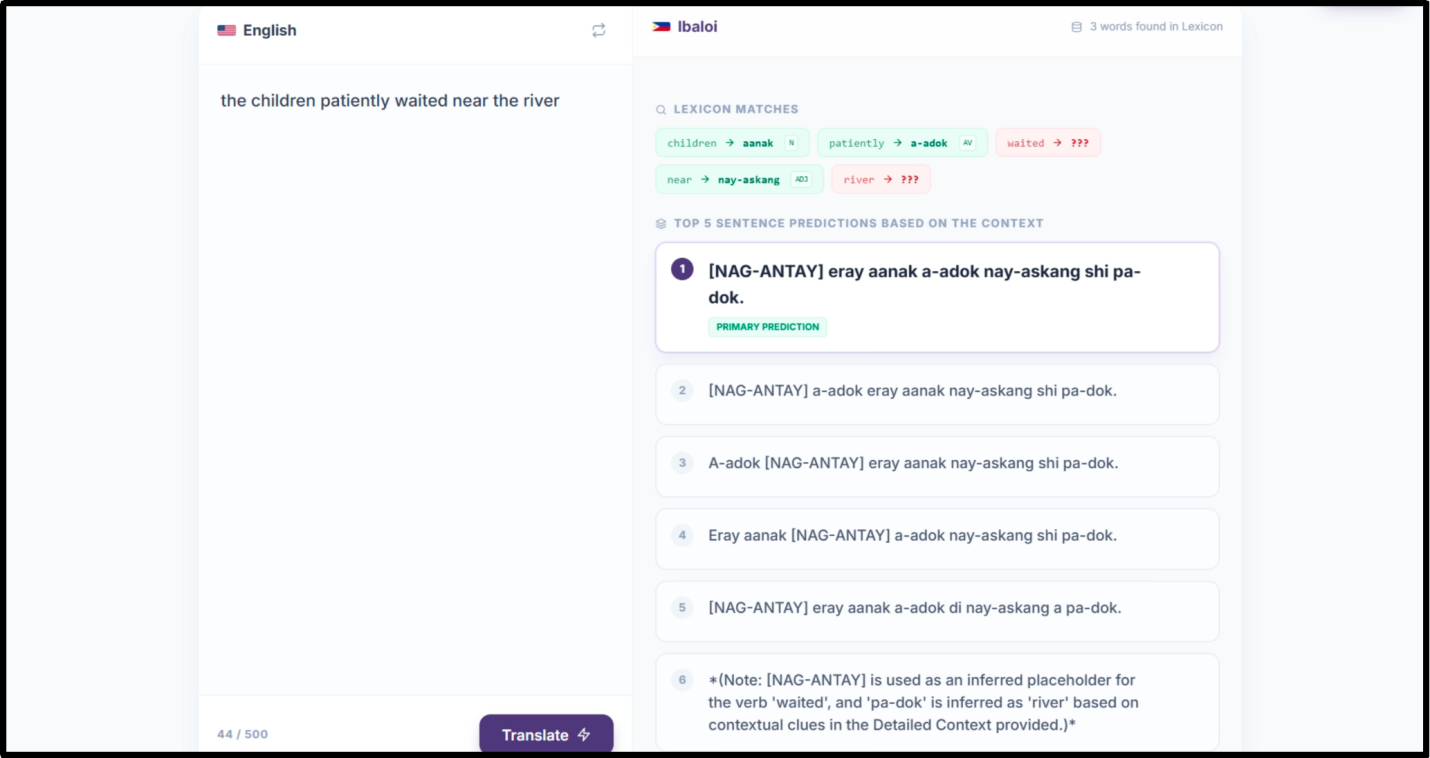
When a main word is selected, text fields appear below for entering an **Ibaloi sentence** that uses the chosen word, and a corresponding **English translation**. The interface also includes two action buttons: **“Add Another Sentence”** for submitting multiple examples, and **“Submit Sentences”** to finalize the input. This tool is used to gather structured Ibaloi example sentences for lexicon enrichment and model training.



## *Figure 9. Ibaloi NLP –* LLM *Translator Tab Interface*

This figure presents the **LLM Translator** tab of the Ibaloi NLP System. The interface is divided into two main panels: the **source language input panel** (left) and the **target language output panel** (right). Users can select the language direction (English to Ibaloi or Ibaloi to English) at the top of each panel. The left panel allows users to type text, with a character counter displayed at the bottom. A **Translate** button is located below the input box to initiate the translation process.

The right panel displays the system’s translation results and real-time processing status. When input is provided, the system performs **lexicon-based matching** and highlights recognized words using color-coded tags. Unknown or out-of-vocabulary terms are flagged for user awareness. This design allows users to visually distinguish between dictionary-supported translations and AI-inferred outputs.

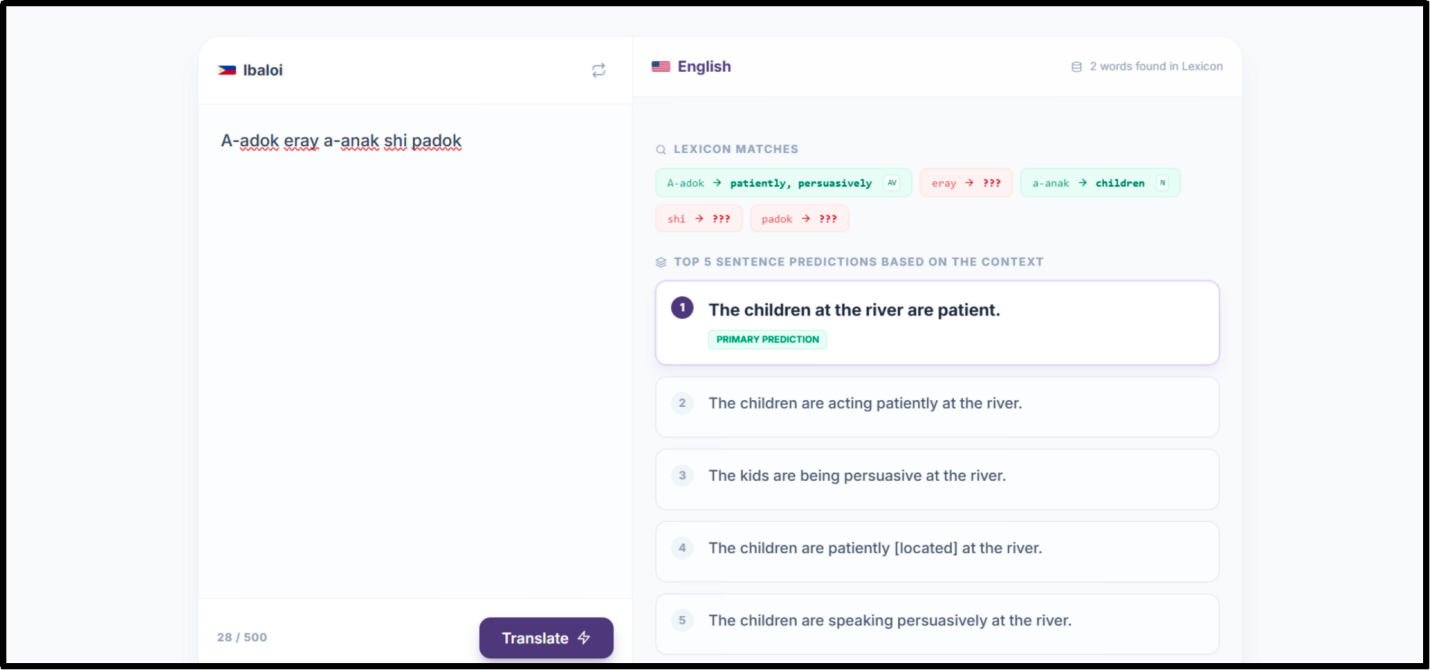


## *Figure 10. Ibaloi NLP – Lexicon Matching and Context Injection View*

This figure illustrates the internal processing view of the NLP Translator. After a user submits a sentence, the system displays a **Lexicon Matches** section, which shows word-level alignments between the source and target languages. Each token is tagged by its part of speech (e.g., noun, adjective, verb) and its corresponding translation if found in the structured lexicon.

Below this, the **Top Sentence Predictions Based on the Context** panel presents multiple candidate translations generated by the hybrid architecture. The primary prediction is visually emphasized to guide the user. This feature demonstrates the system’s **Lexicon + Generative AI context injection** approach, where dictionary rules are combined with neural language modeling to produce grammatically coherent translations even when some words are missing from the lexicon. **[Show top 5 possi; Lexicon lookup + GenAI; lexi lookup task is to extract each entries of user input, serve as context for GenAI; GenAI receive prompts, in which is prompt engineered to form sentence patterns based on ibaloi syntax along with sentence particles.]**

**[this model ain’t using one-to-one translation, but rather we are utilizing context clues based on the lexicon lookup and engineered prompts]**



## *Figure 11. Ibaloi NLP – Reverse Translation (Ibaloi to English) Interface*

This figure shows the LLM Translator operating in the reverse translation mode. In this configuration, the left panel accepts Ibaloi text, while the right panel displays the English output. The system again highlights recognized tokens from the lexicon and marks unknown words.

The prediction panel provides multiple English sentence alternatives, with the most accurate translation labeled as the **Primary Prediction**. This functionality supports both language learners and researchers by allowing users to analyze how Ibaloi sentence structures are interpreted and rendered into English through the LLM-based natural language understanding pipeline.