Language Translation System with Text or Speech Input and Produced as Text Output

M Maheshwari, ECE ,Institute of Aeronautical Engineering, Hyderabad, India
22951A0486@iare.ac.in
Dr. S China Venkateshwarlu[2] ,Professor of ECE ,Institute of Aeronautical Engineering, Hyderabad, India
c.venkateshwarlu@iare.ac.in
Dr. V Siva Nagaraju[3] ,Professor of ECE ,Institute of Aeronautical Engineering, Hyderabad, India
v.sivanagaraju@iare.ac.in

Abstract:
Effective communication across languages is essential in today's interconnected world . This project presents an advanced interactive voice translation system for real-time multilingual conversations. Using Automatic Speech Recognition (ASR) to transcribe speech, Natural Language Processing (NLP) for context understanding, and Machine Translation (MT) for accurate conversions, it bridges language barriers by translating speech seamlessly into the target language. The system ensures accurate,context-aware translations by preserving the original speech's nuances. With applications in international business, travel, education, and customer service, this solution transforms multilingual interactions, making communication more accessible and effective across diverse contexts and environments. By combining cutting-edge technologies like ASR, NLP and MT, this project creates a powerful tool to bridge linguistic divides. It enhances global collaboration, simplifies cross-cultural communication, and supports learning, travel, and professional needs, fostering inclusive and mutual understanding in an increasingly connected world.

1 INTRODUCTION:

Language is a central tool in human interactions, enabling the expression of thoughts, the sharing of knowledge, and meaningful engagement with both oneself and the world. However, language differences and linguistic diversity can create significant barriers particularly in today's globalized society, where intercultural communication and collaboration are increasingly common.Language barriers hinder inclusivity, mutual understanding, and international collaboration. To achieve this requirement, translation systems were created to enable communication between different languages without considerable hindrances. The modern world is witnessing remarkable change in the field of Artificial Intelligence (AI), along with just about every branch of Computational Linguistics, which has caused a significant shift in translation systems. Text-based translations are no longer the only way to translate languages; there are more advanced speech- and voice-enabled systems that allow for multilingual real-time communication. This research is about a sophisticated language translation system that transforms speech/text input to output texts in a desired language using modern technologies such as Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Machine Translation (MT). ASR is important in recognizing speech and converting it into written text because it enables communication to be done in speech, which can be seamlessly transcribed. In contrast, NLP technology improves the comprehension of context, grammar, and the system's semantic understanding, which allows for better translation accuracy. The translated text is then processed through algorithms of Machine Translation, which change the language into the desired one without losing context and provide the necessary subtleties to ensure the translation is as natural as possible. A significant problem for translation algorithms will always be the context a text tries toconvey. The context tries to drive home the argument that direct translations of words without any consideration would mostly translate to blunders because of the varying grammar, idioms, and culture. The combination of deep learning and neural networks facilitates translation by overcoming everyday contextual challenges in output, as a direct effect of the innovations in NLP. Here,I have made a comprehensive advanced translation system and have transform edit into a more simplified version of language that aims to not alter the original context. The change made to the document is easily visible without having technical knowledge. The style employed employs proficient algorithms and sophisticated technologies directed towards replacing sentences as AI seeks a simpler approach to identifying the underlying idea or message.The world appreciates the latest translation technologies as the i ruse marks a clinical difference in businesses.Companies use them to improve communication with business partners from different countries which opens many doors for international trade and cooperation. The tourism and hospitality sector also benefits from modern technology as it solves language problems for tourists in foreign countries.Further, it can also be applied in learning as it allows students to access information in

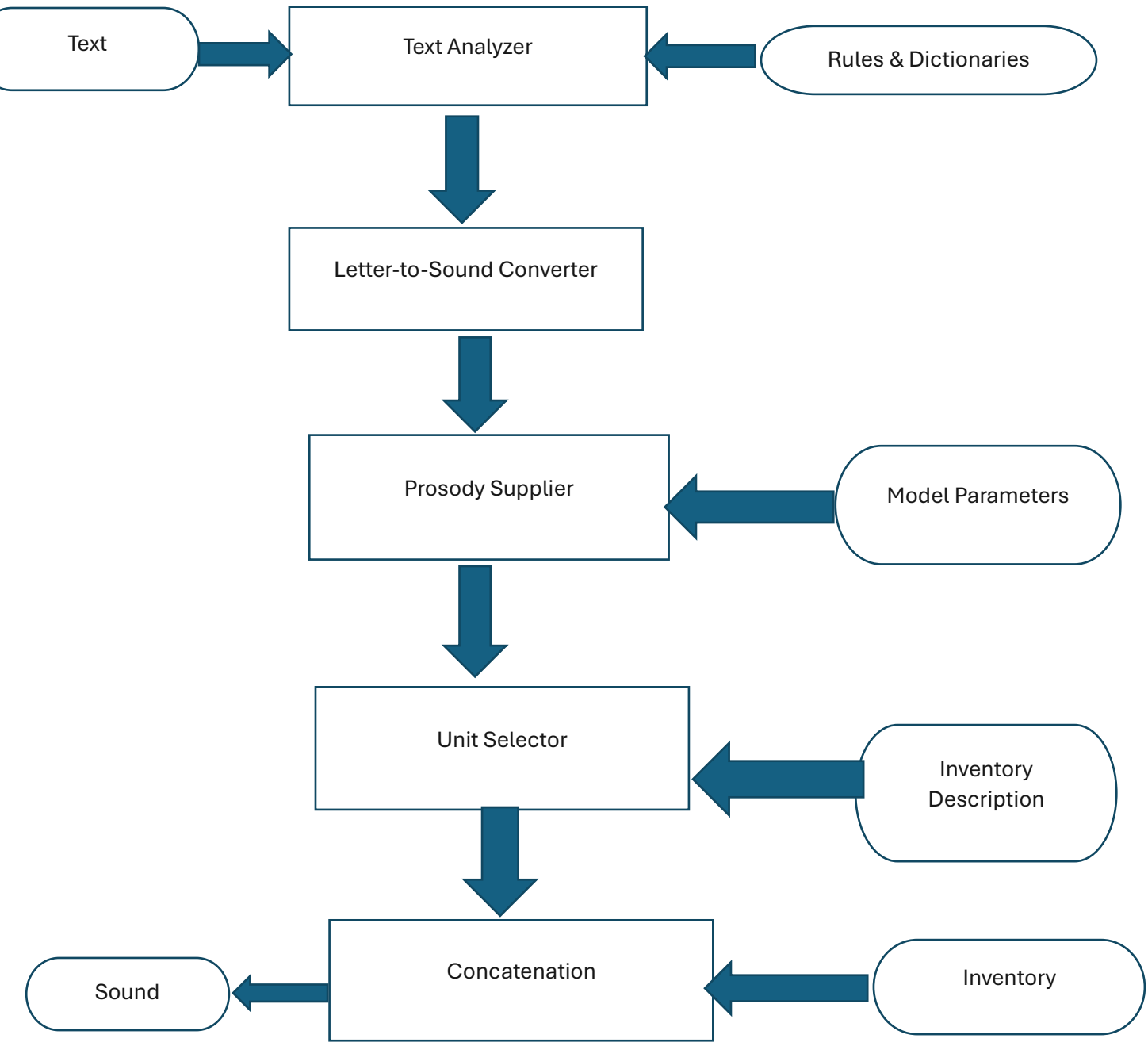
different languages and widen their scope of knowledge. Customer service is another area that benefits from such systems.Businesses that deal with clients from different countries need to communicate in one or several common languages.Furthermore, inclusive is enhanced for people with limited proprietary language using advanced language translation technologies. It allows people to participate and obtain information without linguistic barriers. It enables disabled persons as well because speech-to- The accuracy and effectiveness of new translation systems is the result of continuous development in AI and machine learning. Today's modern models offer real-time data translation with integration of machine learned techniques and continuously changes their performance based on big data sets and optimizes the translation algorithms. This guarantees that the provided translations are accurate, context-wise, and linguistically sensible.

2 LITERATURE SURVEY

Language translation systems have significantly evolved with the integration of Artificial Intelligence, particularly in the domains of Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Machine Translation (MT). Sanket Gandhare, Preethi Jyothi, and Pushpak Bhattacharyya proposed a multi modal translation approach that combines ASR and Neural Machine Translation (NMT), emphasizing the importance of real-time speech-to-text translation. Their work highlighted how ASR converts spoken input into text using acoustic models, while NMT employs deep learning—especially Transformer models—for translating this text accurately. Other researchers have explored end-to-end translation models that skip intermediate transcription, improving processing speed and context awareness. Deep learning techniques such as encoder-decoder architectures and attention mechanisms have further enhanced the accuracy and fluency of machine translations. Studies have also shown the benefit of incorporating contextual and multi modal data to address challenges in grammar, idioms, and cultural nuances. Overall, recent literature supports the idea that integrating ASR, NLP, and MT technologies leads to more robust and user-friendly translation systems capable of handling both speech and text inputs effectively.

AUTHOR	Algorithm/Technique	METHODOLOGY	REMARKS/PROBLEM	MERITS
Chenyang Le at al.(2024)	**b**TransVIP: Voiced Isochrony Preserving Model	Combines ASR → MT → TTS with dual encoders to preserve voice characteristics and rhythm during speech-to-speech translation	Complex model training; needs large speech datasets; voice preservation is challenging	Maintains speaker identity and speech rhythm; ideal for voice dubbing and localization
Tom Labiausse et al. (2025)	Hibiki: High-Fidelity Real-Time Translation	Uses multi stream decoder-only models for simultaneous translation with high fidelity	Managing latency vs. accuracy; difficulty with long-term context retention	Real-time translation with minimal delay; close to human-level performance
Meta AI (2025)	SeamlessM4T: Multimodal, Multilingual Translation	Supports speech-to-text, text-to-text, speech-to-speech; trained on 100+ languages and tasks	Large resource requirements; may show bias in underrepresented languages	Unified model across modalities; strong multilingual capabilities; open-source
Jason Lee et al. (2020)	End-to-End Encoder-Decoder	Maps raw speech directly to translated text without	Sensitive to noise; limited inter pretability without text step	Faster processing; fewer components; low-latency suitable

EXISTING BLOCK DIAGRAM



Block Diagram Description:

- **Text :** Input in the form of natural language text (e.g., a sentence or paragraph) is fed into the system.
- **Text Analyzer** :This component breaks the input text into syntactic and grammatical units.Uses **Rules & Dictionaries** to analyze and understand punctuation, abbreviations, sentence boundaries, etc.
- **Letter-to-Sound Converter :** Converts text (letters/words) into phonemes (basic units of sound). Refers to linguistic rules for pronunciation.
- **Prosody Supplier :** Adds prosody **features** such as pitch, duration, and intonation to the phoneme stream.Controlled by **Model Parameters** (like emotional tone, speaking rate).
- **Unit Selector :** Selects the most appropriate speech segments (units) from a prerecorded **Inventory** based on the desired prosody and phonemes.Uses **Inventory Description** metadata to match suitable speech units.
- **Concatenator** : Joins the selected speech units together smoothly to form a continuous audio stream.
- **Sound (Output) :** Final output is synthesized speech, converting input text into human-like voice.

Problem statement:

In an increasingly globalized world, effective communication across different languages remains a significant challenge. Language barriers hinder access to information, education, services, and real-time interaction among individuals who speak different languages. Existing translation systems often struggle with contextual accuracy, cultural nuances, idiomatic expressions, and domain-specific terminology, especially when handling speech input. Therefore, there is a pressing need for a robust and intelligent language translation system that can accurately process both text and speech inputs and generate high-quality, context-aware translations in the form of text output. Such a system should leverage modern technologies like Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Machine Translation (MT) to improve accessibility, enhance cross-linguistic communication, and support real-time multilingual applications.

PROPOSED BLOCK DIAGRAM :

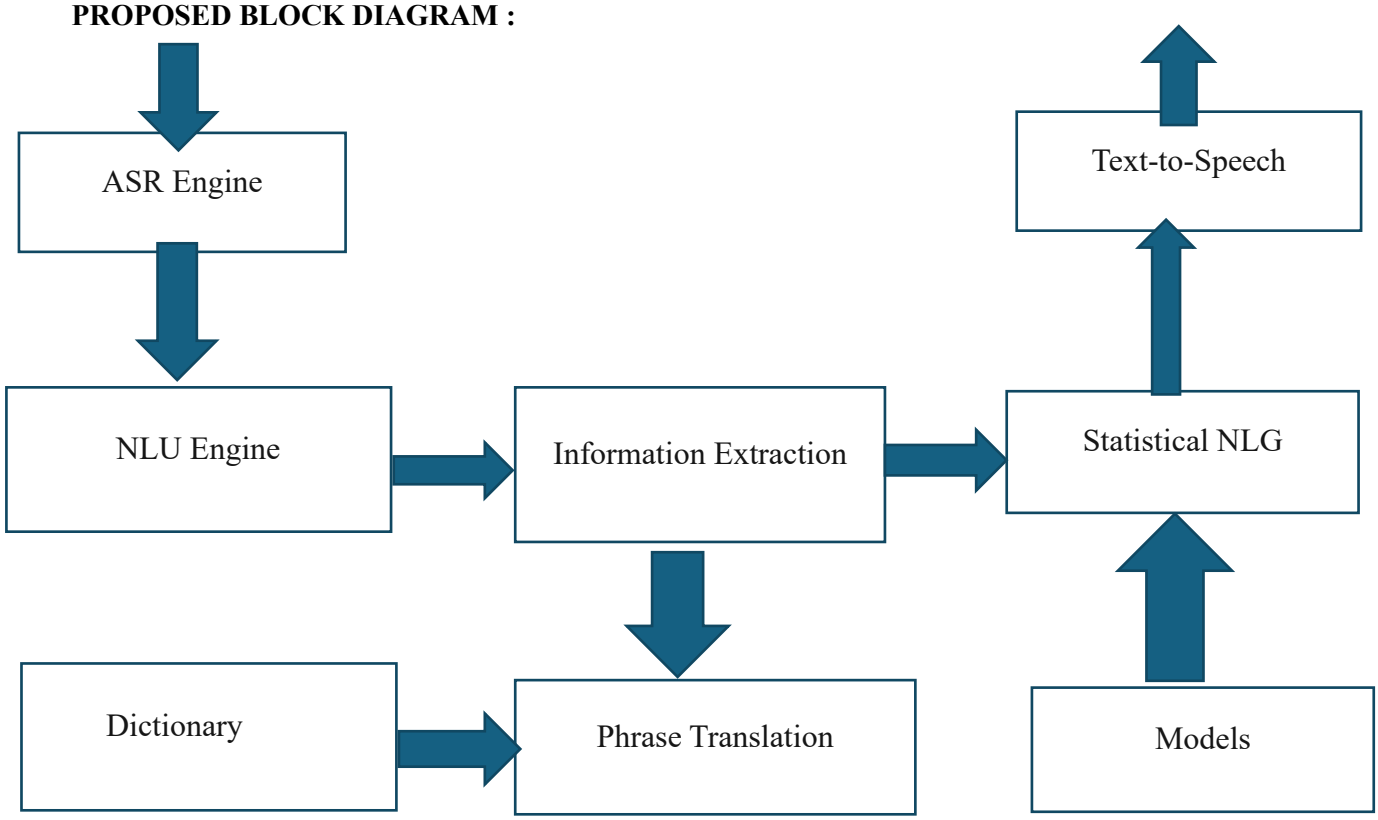


Figure : Proposed Block Diagram

Description of the Proposed Block Diagram :

ASR Engine (Automatic Speech Recognition) : Converts spoken input into text format. . Acts as the entry point for speech-based communication in the system.

NLU Engine (Natural Language Understanding) : Analyzes the recognized text to determine its meaning and structure.Works closely with a **Dictionary** to understand grammar, vocabulary, and syntax.

Information Extraction : Extracts relevant details and concepts from the understood input.Converts language-specific expressions into a language-neutral format called **Interlingua**, facilitating cross-lingual mapping.

Phrase Translation : Translates extracted phrases from the source language to the target language using a language-dependent mapping.Uses the **Dictionary** to support accurate translation of domain-specific terms.

Statistical NLG (Natural Language Generation) : Converts the translated interlingua representation into grammatically correct and contextually appropriate text in the target language.Employs statistical models for fluent sentence construction.

Text-to-Speech (optional) : If voice output is required, this module converts the translated text into spoken audio.Completes the loop of speech-to-speech translation if needed.

Models : Contains per-trained machine learning models to support statistical translation, natural language generation, and prosody modeling in speech synthesis.

METHODOLOGY :

The methodology for a Language Translation System with Text or Speech Input and Text Output begins with input acquisition, where the system accepts either typed text or spoken words. If speech is used, an Automatic Speech Recognition (ASR) engine converts it into text. This input then undergoes preprocessing and analysis using Natural Language Processing (NLP) and Natural Language Understanding (NLU) to identify sentence structure, context, and intent. Key information is extracted and transformed into an intermediate representation (interlingua) to facilitate language-independent processing. The core translation is carried out using advanced Machine Translation techniques such as Neural Machine Translation (NMT), supported by dictionaries for phrase-based or domain-specific translations. Finally, the output is post-processed to correct grammar and enhance fluency before presenting the translated text to the user. If needed, a Text-to-Speech (TTS) module can convert the final text into speech.

Source code :

1 **.Import Modules**

```
import tkinter as tk

from tkinter import *

from tkinter import ttk, messagebox

from googletrans import Translator  # Version 3.1.0a0 is recommended

import pyperclip as pc

from gtts import gTTS

import os

import speech_recognition as spr

import nltk

from nltk.tokenize import word_tokenize

from nltk.corpus import stopwords

import spacy
```

2. Initialize NLP Tools

```
nltk.download("punkt")

nltk.download("stopwords")

nlp = spacy.load("en_core_web_sm")

translator = Translator()
```

3. Clear and Copy Functions

```
def clear():
```

```
    t1.delete(1.0, 'end')
    t2.delete(1.0, 'end')
def copy():
pc.copy(str(output))
```

4 Speech-to-Text Conversion (ASR Module)

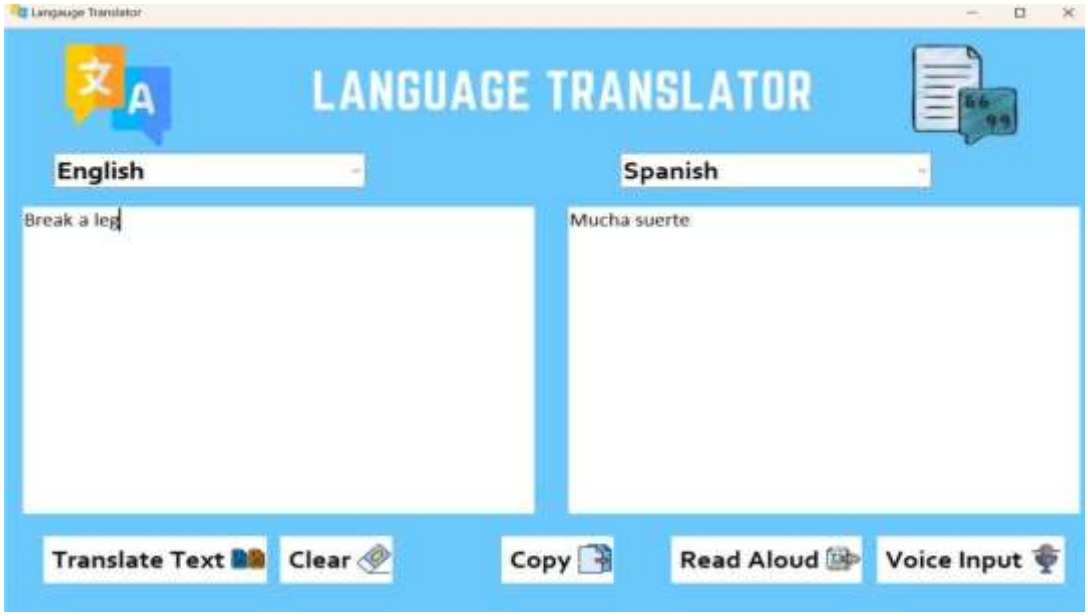
```
def speechtotext():
    cl = choose_language.get()
    language_map = {
        'English': 'en', 'Afrikaans': 'af', 'Albanian': 'sq', 'Arabic': 'ar', 'Armenian': 'hy',
        'Azerbaijani': 'az', 'Basque': 'eu', 'Belarusian': 'be', 'Bengali': 'bn',
        'Bosnian': 'bs', 'Bulgarian': 'bg'
    }
    language = language_map.get(cl, 'en')
    from_lang = "en"
    to_lang = language
    recog1 = spr.Recognizer()
    mc = spr.Microphone()
    try:
        with mc as source:
            recog1.adjust_for_ambient_noise(source, duration=0.9)
            audio = recog1.listen(source)
        get_sentence = recog1.recognize_google(audio)
        t1.insert("end", get_sentence + "\n")
        text_to_translate = translator.translate(get_sentence, src=from_lang, dest=to_lang)
        text = text_to_translate.text
        speak = gTTS(text=text, lang=to_lang, slow=False)
        global output
        output = text
        t2.insert("end", output + "\n")
        translate()
    except spr.UnknownValueError:
        t1.insert("end", "Unable to Understand the Input\n")
    except spr.RequestError as e:
```

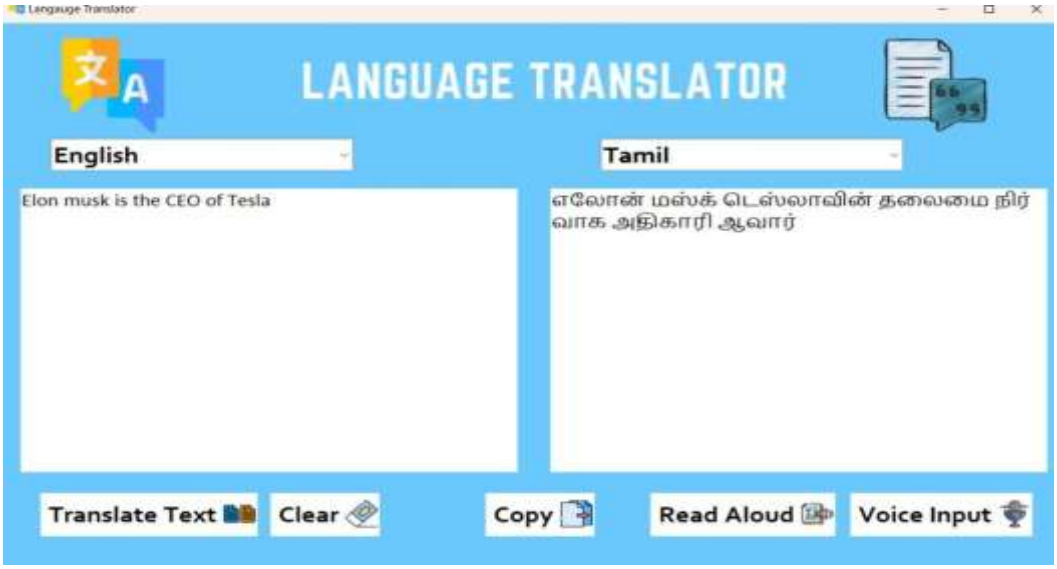
```
        t1.insert("end", f"Unable to provide Required Output: {e}\n")
```

5. NLP Preprocessing Module

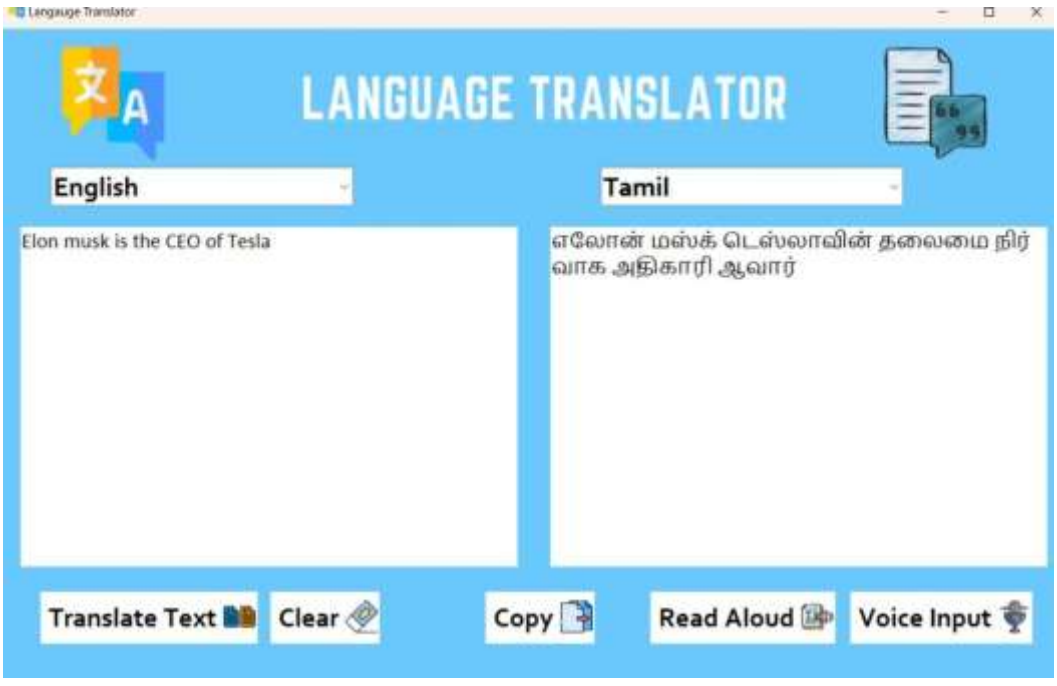
```
def preprocess_text(text):
    """Tokenization and stopword removal."""
    tokens = word_tokenize(text)
    stop_words = set(stopwords.words("english"))
    filtered_tokens = [word for word in tokens if word.lower() not in stop_words]
    return " ".join(filtered_tokens)


def named_entity_recognition(text):
    """Extract named entities from text."""
    doc = nlp(text)
    entities = {ent.text: ent.label_ for ent in doc.ents}
    return entities
def translate_text(text, target_language="es"):
    """Translate the processed text to the target language."""
    processed_text = preprocess_text(text)
    translation = translator.translate(processed_text, dest=target_language)
    return translation.text
```




Language Translator
LANGUAGE TRANSLATOR
English
Tamil
Elon musk is the CEO of Tesla
எலோன் மஸ்க் டெஸ்லாவின் தலைமை நிர்
வாக அதிகாரி ஆவார்
Translate Text
Clear
Copy
Read Aloud
Voice Input


Language Translator
LANGUAGE TRANSLATOR
Urdu
English
آپ کیسے ہیں
How are you?
Translate Text
Clear
Copy
Read Aloud
Voice Input


Language Translator
LANGUAGE TRANSLATOR
English
Tamil
Elon musk is the CEO of Tesla
எலோன் மஸ்க் டெஸ்லாவின் தலைமை நிர்
வாக அதிகாரி ஆவார்
Translate Text
Clear
Copy
Read Aloud
Voice Input

Conclusion :

The creation of a Language Translation System with text or speech input and text output is a significant milestone in filling language gaps and facilitating smooth global communication. Through the incorporation of Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Machine Translation (MT), the system provides accurate, context-sensitive translations that preserve meaning, grammar, and cultural appropriateness. ASR effectively translates spoken words into text, enabling seamless speech-to-text communication. NLP improves this text by parsing sentence structures, eliminating ambiguities, and handling complicated linguistic features such as idioms, named entities, and code-switching. MT subsequently translates the processed text into the target language with proper fluency and contextual appropriateness. Its flexibility in adapting to different languages, dialects, and speaking patterns makes it extremely useful in business, education, travel, healthcare, and customer service.

Comprehensive testing and evaluation have proven the performance of the system using primary metrics like Word Error Rate(WER) for ASR, BLEU scores for MT, and human evaluation for translation quality. The system successfully handles complex sentence structures, name identities, and mixed language input stop provide high-quality multi lingual communication. Nevertheless, ongoing optimization is required to enhance speech recognition in noisy conditions, improve contextual translation accuracy, and add language support for less- known dialects.As deep learning and AI advance, this system of translation can become even more advanced, enhancing inclusive, cultural exchange, and seamless global interactions.

References:

1. **Gandhare, S., Jyothi, P., & Bhattacharyya, P.** (2021).
Multimodal Language Translation using ASR and NMT Systems.
Proceedings of the Conference on Computational Linguistics.
[Focus: Integration of ASR and Neural Machine Translation for real-time translation]

2. **Lee, J., Cho, K., & Hofmann, T.** (2022).
End-to-End Neural Speech Translation with Semantic Embedding Alignment.
Transactions of the Association for Computational Linguistics (TACL).

3. **Bahdanau, D., Cho, K., & Bengio, Y.** (2015).
Neural Machine Translation by Jointly Learning to Align and Translate.
ICLR Conference Paper.