Bilingual (English to Bengali) Technical E-Dictionary for Aviation OOV Words

Saptarshi Paul

Abstract: E-dictionaries, quite common today are available for multiple languages in monolingual, bilingual and multilingual forms. In NLP they form the core of a series of tools that are used to understand words, sentences and in turn the language itself. These E-Dictionaries work well for any language domain as a whole. For almost all languages E-dictionaries are available, but once specialized technical domains are encountered these E-Dictionaries are quite useless. Aviation is one such specialized domain for which no E-Dictionary, translation or transliteration tool exist. On the other hand the need for such tools for specialized domains are increasing. The tool discussed in this paper is an attempt to bridge the gap that currently exists between English and Bengali languages.

Keywords: Aviation, Air force, Aero-Space, Translation, E-dictionary.

I. INTRODUCTION

E-dictionaries are playing an important part in the field of Translation and Transliteration. E-Dictionary for daily use words are commonly available but it is hard to find E-Dictionaries/Translitrators/Translators for specialized domains like Medicine, Aviation etc. Aviation, being a highly specialized domain consists of out-of-vocabulary (OOVs) words, slangs and technical words which combine to form Aviation related instructions and sentences. Aviation sentences are mainly composed of structured words meant only for the aviation domain. Though related work has been done on preprocessing words before translation[1][2][4][5]and attempts to process and translate English to French and other aviation related words[2][3][6] The complete absence of E-Dictionaries and word translators, specially from English to Indian Languages has created a huge void. The E-Dictionary as created and discussed in this paper tries to fill up that void and tries to bridge the existing gap between English and Bengali Technical words.

II. THE BACKGROUND

The Indian aviation market with its tremendous growth is set to be the third largest market in aviation by 2030[7]. This trend is attracting exponential investment, thus creating huge job opportunities. The aviation domain is quite mysterious to the outside world.

People who venture into this attractive world of aviation find themselves in an alien environment. The aviation world is composed of hundreds of OOVs, slangs and structured words which then come together to form the aviation sentences (manual for ATC instructions, apron instructions, ground handling instructions etc). Directly or indirectly more and more people are getting involved with the aviation domain and they find themselves at a loss, unable to mix easily in the environment.

In aviation domain airports and aircrafts are not known by their standard English names. Airports are generally known by International Air Transport Association (IATA) and (ICAO) International Civil Aviation Organization codes and Aircrafts by their unique registration numbers. Country codes differ so does the various airliner codes. Let us look at some of the airport names of India:

<table>
<thead>
<tr>
<th>AIRPORT NAME</th>
<th>ICAO CODE</th>
<th>IATA CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mumbai</td>
<td>VABB</td>
<td>BOM</td>
</tr>
<tr>
<td>Bangalore</td>
<td>VOBL</td>
<td>BLR</td>
</tr>
<tr>
<td>Agartala</td>
<td>VEAT</td>
<td>IXA</td>
</tr>
</tbody>
</table>

Also almost all the aviation related infrastructure and procedures are composed and known by unique aviation codes which in turn is an OOV word in the standard language domain.

<table>
<thead>
<tr>
<th>Aviation codes</th>
<th>Meaning</th>
<th>Aviation codes</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>Air Base</td>
<td>NOTAM</td>
<td>Notice To Air Men</td>
</tr>
<tr>
<td>GEAR</td>
<td>Wheels of the Aircraft</td>
<td>ATC</td>
<td>Air Traffic Controller</td>
</tr>
</tbody>
</table>

From the viewpoint of Natural language Processing, Transliteration of aviation words are possible to some extent while no standard translation software exists for these terms. Till recently for India and Indian Languages there were no centralized storage and processing facility available. It is only recently in 2019 that the central ATFM (Air Traffic Flow Management Central command Centre) complex has been inaugurated. Also there is no standard tool such as E-Dictionary, translation and transliteration tool available for Indian Languages.
III. THE PROBLEM

Apart from the absence of English to Indian languages E-Dictionary (Aviation domain), no standard Translation software’s (Google Translate, Microsoft Bing Translate) are not able to translate/Transliterate the aviation terms and terminology leaving a huge VOID in the English-Indian Languages (Bengali as shown in the example) domain.

For Example: in the picture given below we can see that Microsoft BING translate is unable to transliterate/translate the term IXS (Silchar Airport).

The output of IXS has come out to be ইইসস while it should have been আইএক্সএস.

While Google Translate is completely unable to understand the term IXS and leaves it un-transliterated. some degree of transliteration can be expected from standard software, which does not seem to happen in such cases.

Fig.1: snapshot of Microsoft Bing translating aviation OOV words “IXS”

Also, there is no standard E-Dictionary (Online/Offline) or in paperback format that can address the English-Bengali Aviation OOV words and slangs.

IV. THE CHALLENGE

Bilingual E-Dictionary needs a database on which it can work on. Creation of the database consisting of aviation OOV words in absence of any supporting standard Indian Aviation Database System and finding out its corresponding meaning in Bengali is the real challenge.

V. THE METHODOLOGY

The design and working Methodology of the tool can be described through the following steps:

Step1: the collection of OOV words, slangs and Technical terms from Standard sources [8][9][10][11][12]

Step2: freezing the format of the tool (Structure and composition) (fig.4)

Step3: creation of a database that will act as the base for the tool (fig.4)

Step4: fixing the software needed for creation of the tool (WAMP in our tool)

Step5: the creation of an E-Dictionary based on the database and required structure.

VI. IMPLEMENTATION

WAMP has been used for the creation of the database as it includes MySQL, UTF-8 Unicode, Unicode Transformation Format has been used for Bengali letters. WAMP has also been used for the creation of the GUI.

A. The Database

the database has been created using Php which consists aviation OOV words, its corresponding meaning, Parts of speech, Bengali transliteration, Bengali meaning and an example sentence to explain how it is used. The database is an ever growing one, encompassing new OOV words as we come across them.

B. Structure and Composition of the database

In Fig.4, shown below, we see that the database consists of Seven columns. Which can be summarized as follows:

Column1 : sequence id
Column 2: the aviation OOV word
Column 3: English meaning of the OOV word
Column 4: part of speech
Column 5: Bengali transliteration of the equivalent aviation OOV word
Column 6: Bengali meaning of the OOV word
Column 7: An example sentence explaining how the OOV word is being used with respect to its part of speech.
VII. RESULTS AND OUTPUT

The results have been depicted using Snap-shots of the E-dictionary that addresses the current existing gap.

A. Example snapshots of the Software

Some examples of the E-dictionary are provided to explain the working of the tool:

If we consider the OOV word “IXS” again as in the earlier examples then our tool provides us with the following output:

Fig. 3: Snapshot showing result/output of the tool for OOV word IXS

The output as shown above can be listed as follows:
1. English meaning: Silchar Airport
2. Bengali Transliteration: আইএক্সএস
3. Part of Speech: NOUN
4. Bengali Translation: শিলচর বিমানবন্দর
5. Example sentence: Silchar Airport is the primary airport in Barak valley region of Assam

Taking another example for the Aviation OOV word “BS”, on entry into our E-Dictionary tool it provides us with the following output:

Fig. 4: Snap shot of the database that has been created to support the E-Dictionary
Fig. 5: Snapshot showing result/output of the tool for OOV word BS.

The output can be listed as follows:
The output as shown above can be listed as follows:

1. English meaning: Broadcast Station
2. Bengali Transliteration: বি এস
3. Part of Speech: NOUN
4. Bengali Translation: সম্প্রচার সেবাসদন
5. Example sentence: Broadcast Station (Commercial) provides various information related to aviation

VIII. RESULTS AND OUTPUT ANALYSIS:
Comparing fig.1 and fig.2 we can see that while both Google Translate and Microsoft Bing are unable to translate the term aviation OOV word “IXS”, fig.3 and fig.5 shows that our tool not only transliterate the term “IXS”, It also provides the corresponding translation, part of speech and an example of its use according to the part of speech. For OOV words that have different part of speech (fig.6) the tool provides different example sentences.

Fig. 6: Snapshot showing result/output of the tool for OOV word BS for its different parts of speech (adjective and noun) and its example sentences.

A. Benefits of the E-Dictionary:
The English to Bengali E-Dictionary not only provides the meaning in Bengali, its Bengali transliteration but also the meaning of the OOV in English its part of speech and how it can be used accordingly. This type of E-Dictionary in the aviation domain is unique not only

B. Uses of the Tool
The tool can be used as a base tool during the creation of any Bilingual technical corpus for English to Bengali and also to some extent for English to any other language. This tool can act as a guide to anyone working on English-Bengali corpus or for people who wants to refer and understand aviation related language better. This tool can also be used to guide and introduce the newcomers in the world of aviation. This tool is a unique one and first of its kind, its uniqueness lies in the technical aspect of aviation which in turn is based on OOV words.

C. Uniqueness of the tool
This tool as described in the paper is an unique one with no other contemporary E-Dictionary, currently available that address the aviation domain in any Indian languages.

D. The multifold benefits of the bilingual database
The bilingual database created for this E-dictionary can be used to form Bilingual corpus. This corpus can be a technical English to Bengali corpus in aviation that can be an unique one and which can act as an benchmark for any other future technical corpus in Space science, Aero-science, defense sector (Air force) and loads of other technical domains. The uniqueness of such bilingual dictionary and corpus is unparallel

E. The Source of the database
The source of the database includes standard data sources such as
1. Airport Authority Of India (AAI) - Manual of Air traffic Services [8]
2. Directorate General of Civil Aviation, (DGCA) manuals and accident/incident reports [9]
3. International Civil Aviation Organization (ICAO) reports on safety/incidents.[10]
4. National Transportation safety board-Aviation accident data summery also data from NASA-ASRS incident reports, aviation newsletters and aviation related blogs has been included .[11][12]

IX. CONCLUSION AND FUTURE WORKS
The tool as discussed in the paper is a unique tool that addresses the English-Bengali technical gap that exists in the vocabulary. While no other current tool is capable of such explanation of Aviation OOV words, the discussed tool is an simple but efficient one. Future work includes increasing the number of OOV words in the database of the E-Dictionary and expanding the scope to include Aero-Space and Air force terms and terminologies.
REFERENCES:


5. Vishal Gupta and Gunpreet Singh Lehali,Processing Phase of Punjabi Language Text Summarization, Information Systems for Indian Languages, ICISIL 2011, pp 250-253


11. https://asrs.arc.nasa.gov/search/reportsets.html

AUTHORS PROFILE

SAPTARSHI PAUL, Assistant Professor, Computer Science Department, Assam University, Silchar. Masters of Technology in Information Technology (Networking) from Vellore Institute of Technology University, Vellore, and Masters of Science in Computer Science from Assam University, Silchar. Specialization and research work is in the field of Natural Language Processing. Ten (10) years of Work experience as lecturer in CMR IT at Bangalore and as Assistant Professor(Grade I and II) at Department of Computer Science, Assam University, Silchar. Publications in the field of Remote Sensing, GIS and Natural Language Processing involving transliteration and translation dealing with Aviation related vocabulary. A member of Asian Language Tree bank project.