

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,500

Open access books available

135,000

International authors and editors

165M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Understanding Aviation English: Challenges and Opportunities in NLP Applications for Indian Languages

Saptarshi Paul

Abstract

English is a language that is understood, spoken and used by citizens of a diverse array of countries. The speakers include both native and non-native speakers of English. NLP or Natural Language Processing on the other hand is a branch of computer science that deals with one of the most challenging aspect that a machine can process: dealing with Natural Languages. Natural languages which have evolved over centuries are complete, diverse and highly complex and thus are challenging for a computer system to understand and process. MT or Machine Translation is a more specific part of NLP that translates one natural language to another (English being one of the major researched and sought after languages among them). Though research in the field of NLP and MT has come a long way and many efficient translators are available, still Translation and other NLP applications in specialized domains such as aeronautics are still today a challenge for NLP researchers and developers to achieve. NLP applications are often used in education of English Language, and are therefore a continuous process for Non-Native speakers of English. Non-native English speakers take help of various NLP tools such as E-Dictionary, MT applications and others to better understand the English language and thus learn it better and faster. Aviation English poses a challenge to MT systems and understanding it as a whole requires specialized handling as it has own phonetic pronunciations and terminologies and constituent Out-Of-Vocabulary words. Dealing with Aviation English calls for teaming up of experts from Applied Linguistics, NLP and AI. As a result it becomes a cross-research discipline that covers situations that demand real time use of proper language, e.g. ATC communications. This Paper aims to discuss most recent research methodologies that deals with the Aviation English and reviews the problems posed by it. Being a specialized and structured form of English, the problems are faced by both native and non-native speakers of English Language. Discussion is carried out in the relevant and recent advances of methods in dealing with aviation English language challenges from both, the Human (ICAO/DGCA/AAI) as well as NLP angle. Lastly we have a look at how these challenges are linked to scope for development of applied technologies. Research in experiential Aviation English situations deals with both English for Specific Purposes - ESP (Aeronautics in our case) as well as situations in English as a Foreign Language i.e. EFL (English-Indian language pair).

Keywords: English, Aeronautics, NLP, AAI, ICAO, DGCA

1. Introduction

In order to deal with specialized domains such as Aviation, Aero-Science, Aeronautics manufacturing and maintenance, Translation applications till today depends on Human assisted machine translation systems rather than fully automated and autonomous ones. The real challenge is posed by the unique structured words, OOV words and phraseologies that these domains consist of. One such domain is Aviation. It has evolved completely based on technical OOV words and sentences and has constantly enriched itself with words that are used in lieu of general English-Language words. For each and every component in aviation is unique and thus, this domain has evolved as a specialized branch making use of many but not all English-Language words. These words constitute the sentences in aviation and are thus untranslatable through standard translators like Google-Translate and Microsoft Translator Bing. In this paper, discussions will aim to provide a detail look at the constituent parts of the Aviation language and how it is a challenge for NLP applications such as Machine Translation. The discussion details how the Airport Authority of India and Director General of Civil Aviation (DGCA), India deals with training, testing and certification for Aviation English Language proficiency. The paper also has a look at how this domain has remained unexplored for Indian languages till the recent past. In the next phase the paper goes through the work that has been carried out for English-Bengali/other language pair. India being a country of non-native speakers of English Language makes it a challenging task. The scope that arises out of these challenges for researchers and developers are also discussed in the last section. Apart from databases and data repositories there happens to be a need for a well-researched and authored collection of tools and applications in English (monolingual) and English-other language pair (bilingual) that addresses the need for understanding and dealing with aviation English.

2. “ABC” of aviation English

Aviation domain demands that both the native and non-native speakers of English speak and spell the aviation words and sentences in the same way. This has been made possible by ICAO/DGCA which has fixed a standard dialog that is to be followed by all aviators, ATC controllers, ground crew, and maintenance and operation staff among others. Aircrafts and airlines are identified by their country of origin (VT for India) by combination of English alphabets (Alpha-Bravo-Charlie for ABC). The phonetic pronunciations of these alphabets are also fixed by the ICAO. **Table 1** shows how the phonetic pronunciation of numbers in aviation, while **Table 2** show the English alphabets equivalent.

ICAO prescribed phonetic pronunciation for Numbers	0: ZE-RO	7: SEV-en
	1: WAN	8: AIT
	2: TOO	9: NIN-er
	3: TREE	Decimal: DAY-SEE-MAL
	4: FOW-er	Hundred: HUN-dred
	5: FIFE	Thousand: TOU-SAND
	6: SIX	—

Table 1.
Phonetics for roman numbers as prescribed by ICAO [1, 2].

A: ALFA (ALFA)	B: BRAVO (BRAVO)	C: CHARLIE (CHARLIE)
D: DELTA (DELTA)	E: ECHO (ECHO)	F: FOXTROT (FOXTROT)
G: GOLF (GOLF)	H: HOTEL (HOTEL)	I: INDIA (INDIA)
J: JULIET (JULIET)	K: KILO (KILO)	L: LIMA (LIMA)
M: MIKE (MIKE)	N: NOVEMBER (NOVEMBER)	O: OSCAR (OSCAR)
P: PAPA (PAPA)	Q: QUEBEC (QUEBEC)	R: ROMEO (ROMEO)
S: SIERRA (SIERRA)	T: TANGO (TANGO)	U: UNIFORM (UNIFORM)
V: VICTOR (VICTOR)	W: WISKEY (WISKEY)	X: XRAY (XRAY)
Y: YANKEE (YANKEE)	Z: ZULU (ZULU)	—

Table 2.
 Phonetics for alphabets as prescribed by ICAO [1, 2].

Going by the above example if a flight is registered as “VT-SCA”, where “VT” means “VICEROYs TERRORITY” that is, the aircraft is registered with DGCA (INDIA). If it is to be addressed by the ATC as then controller will pronounce it as “VICTOR TANGO -SIERRA CHARLIE ALFA”. The same pattern is followed for all documentation such as incident reports, maintenance manuals and paper works in the aviation domain.

3. Aviation “Out of Vocabulary” words and phraseologies

Aviation sentences consist of structured and standard OOV words and phraseologies combined with normal English words. Common example of such technical sentences is Notice to Airmen, better known as “NOTAM”. Aviation makes use of unique unheard of OOV words that are unique only to it. Example includes terms such as aircraft are known as ACFT, FL for Flight Level and such.

Name of airport	IATA code	ICAO code
Bagdogra	IXB	VEBD
Heathrow	LHR	EGLL
Los Angeles	LAX	KLAX
Patna	PAT	VEPT
Agartala	IXA	VEAT

Table 3.
 IATA and ICAO codes for a few airport names.

Aviation OOV	Meaning in English
Circuit	To circle at a particular altitude
Front Gear	Set of wheels at the front
ALT	Altitude
FLT	Flight
EBOUND	Going towards East

Table 4.
 Aviation OOV words and their meanings in English.

All documentations in aviation, airports are known by either IATA location identifier or by ICAO four letter arrangements which are unique only to that particular airport. These arrangements are used by people related to aviation irrespective of them being Native or Non-Native speakers of English. The IATA and ICAO code of a few airports are as depicted in **Table 3**. We can see in **Table 3**, while IATA codes are all 3 letter words, ICAO are 4 letters. IATA stands for International Air Transport Association, while ICAO means International Civil Aviation Organization. Arrangements as such make hundreds of OOV words and have formed the basic vocabulary for aviation related documents. **Table 4** presents with some examples of OOV words used in aviation.

4. Exploring the language aspect of AAI manual of air traffic services

India being a country of non-native speakers of the English language, Airport authority of India has made available-online the Manual of Air Traffic Services [3] for references of concerned parties. There are 17 chapters that can be listed in **Table 5**. In the table it can be observed that out of the 17 Chapters, Chapter 12 is concerned with the aviation phraseologies. Chapter 12 highlights how the OOV aviation words and general English sentences are combined together to form the various unique terminologies that are accepted all throughout the aviation domain. Chapter 14 emphasizes on the ATC-Pilot communications, which is again in aviation English. Chapter 15 highlights the procedures for communication over the communication channel. To get a better idea of phraseologies used in aviation let us take a look at some examples and their respective meanings (**Table 6**).

Chapter number and name	Chapter number and name	Chapter number and name	Chapter number and name	Chapter number and name	Chapter number and name
1. Document Identification and control	2. Definitions	3. General	4. AirTraffic Services (ATS)	5. Separation Methods and Minima	6. Procedures in the vicinity of aerodromes
7. Aerodrome control Services	8. Radar Services and Procedures	9. Flight information Services	10. Coordination	11. ATS Messages	12. Phraseologies
13. ADS services	14. Controller-Pilot data link communications	15. procedures for communication failures, etc	16. Miscellaneous Procedures	17. ATS Safety Management	

Table 5.
AAI MATS layout.

Aviation Phraseologies /Phrases	Meaning
RVR NOT AVAILABLE	Runway Visible Range not available
CANCEL OFFSET	Instruction to rejoin cleared flight route or other information
MAINTAIN VMC	Maintain visual meteorological conditions

Table 6.
Example of AAI MATS CHAPTER 12 phraseologies.

There are hundreds of aviation phraseologies spanning from aircraft identification to holding pattern that has to be known to people working in the aviation sector, especially pilots and ATC controllers. These phraseologies are arrangements of specific aviation domain words and general English ones.

5. DGCA aviation English language proficiency training, test & certification

DGCA, India has devised a complete systematic procedure that provides training in real time application of Aviation English Language. It also covers radio-telephony English knowledge acquisition and communication skills over RF involving simulated and real time learning environments for English as a Foreign specialized language. As a case study the authors discuss in detail the way AEL is trained, tested and certified for the cadets/ candidates. The main aim of DGCA is to make sure that an applicant for Pilot license, ATC personnel, aircraft engineer and route navigator license to have the capability to communicate and understand the aviation English language used through RT to the level of required proficiency [4]. The CAR or better known as Civil Aviation Requirement is available according to the provisions of Aircraft rules of 1937-133A. It lays down the procedures for Training, Testing and Certification for Aviation English Language proficiency. The following **Table 7** gives us an idea of the candidates, evaluators, measures, metrics and measurements used to evaluate the Aviation English Language proficiency.

The multiple stages for proficiency in English language are as depicted in **Table 8**. The assessment parameters have 6 areas of specialization which are:

Applicable to	Evaluator	Measures	Measurements	Metrics
A holder of Pilot's (PP License, CP License, Airline Transport Pilot's License (Fixed and rotary wings) A holder of Flight Engineer's License Route Navigator's License	Interlocutor	Language proficiency skills	Descriptor	<i>Minimum Age:</i> 16 years
	Operational rater or Operational assessor		Rating scale	<i>Educational qualification:</i> Class 10 or equivalent
	Rater or Assessor			<i>Training:</i> 200 to 400 hours of aviation English language training
An holder of Air Traffic Services Personnel License				

Table 7.
 DGCA aviation English language proficiency testing layout.

Stage	Meaning	Acceptable?	Re-evaluate?-further action
1	Pre-elementary	No	More training suggested
2	Elementary	No	More training suggested
3	Pre-operational	No	More training suggested
4	Operational	Yes	Yes, after 6 years
5	Extended	Yes	Yes, after 6 years
6	Expert	Yes	No Re-evaluation required

Table 8.
 DGCA aviation English proficiency stages.

pronunciation, structure, vocabulary, fluency, comprehension and interactions. Together they determine the linguistic performance of the candidate.

6. Language loss & deterioration

It is commonly observed from experience and practical observation that for non-native speakers of English, language loss is quite common. Deterioration in language proficiency of candidates for whom English is not the 1st language is also a common trait. In cases as such, candidates' proper re-evaluation and assessment may be conducted according to ICAO norms for Aviation English. DGCA and other aviation regulatory bodies around the world endorse such progress.

7. The NLP angle

Natural Language Processing or in short NLP is an interesting branch of research that encompasses Artificial Intelligence, Neural Networks, Linguistics and an array of Natural Languages such as English, French etc. It aims to provide seamless translation from one natural language to another through translation and transliteration, among other applications such as part-of-speech tagging and E-dictionary. Though the use of monolingual NLP applications is found in the modern aviation services (IBM WATSON [5], AMRIT [6], BLEU [7] and PLUS [8]), it is hard and almost impossible to find bilingual translation services in regular real time use. The same goes for all Indian languages. Apart from incorporating the underlying rules of the concerned natural languages NLP has always strived to create monolingual and bilingual corpora that can assist in the translation and transliteration of the various natural languages. Specialized English applications such as Aviation / Aeronautics and similar streams have always posed a challenge to achieving the goals of NLP applications. The use of structured English words in aviation not only prevents proper translation but also Transliteration. Let us look at some examples of the mentioned problems.

8. Problem with direct translation and transliteration

The aviation OOV words cannot be directly translated and transliterated by standard translation applications like Google Translate and Microsoft translator

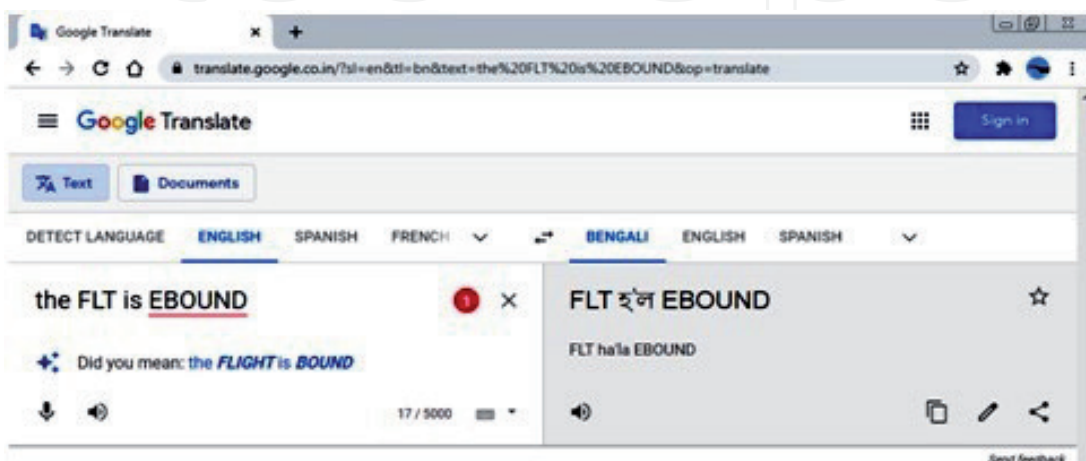


Figure 1.
Inability of Google TRANSLATE in handling aviation OOV words.

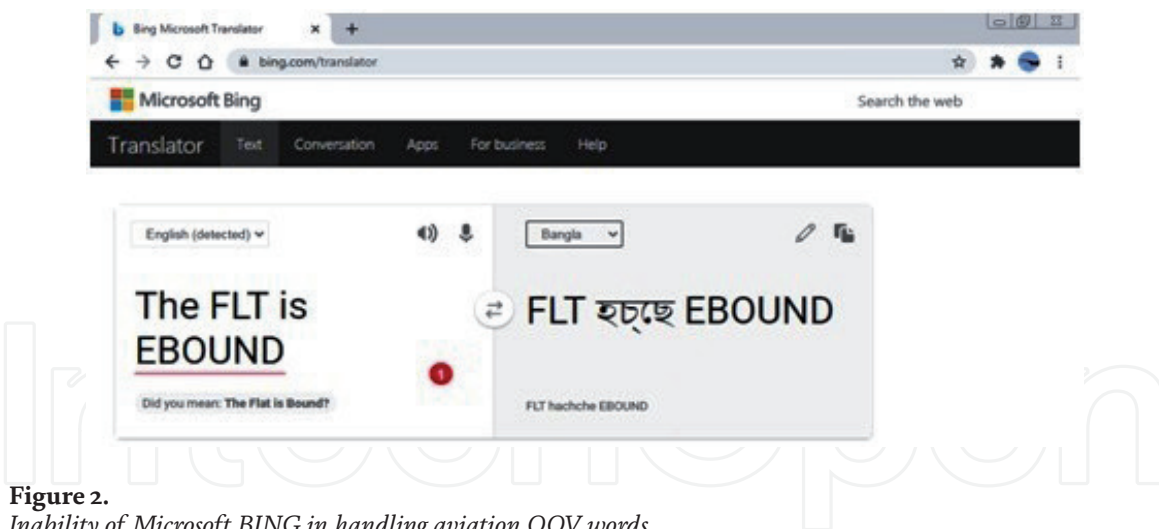


Figure 2.
Inability of Microsoft BING in handling aviation OOV words.

Bing. Multiple attempts to translate and transliterate them have resulted in failure. **Figures 1** and **2** shows the inability of standard translators to transliterate aviation OOV word.

9. Unavailability of E-dictionary and standard translation work

While standard E-Dictionaries are available in both online and offline form and in soft and hard copies it is special domains such as aviation where E-Dictionaries are not available. Before 2017 apart from TUAM-AVIATION [9] no translation/transliteration work was taken up for aviation maintenance manuals, navigation manuals or such. This makes the work of dealing with aviation sequences all the more challenging.

10. Complex situation with Indian languages

Though in European countries and the USA (ASRS) [<https://asrs.arc.nasa.gov/>] some attempts has been made to document maintenance, guidance and incident reports and manuals in the soft format and online versions, In India, no such attempts has been made. Although DGCA [<http://dgca.nic.in/>] maintains incident and accident summaries involving those happening in Indian Airspace, these reports are saved in PDF format. No attempt has been made to store these reports in a centralized repository, that can be used for further research or mining. These reports are neither categorized nor segmented, thus making them unsuitable for research and development purposes. Though TDIL (<https://tdil.meity.gov.in/>) holds a wide variety of Indian corpora and tools still resources for aeronautics and aero-space domain are completely absent. Making the matter more complex no database or corpora in India along with options for corresponding pronunciation, phonetic representation and meaning in any Indian languages are available for the aviation domain. Transliteration and Translation applications for Indian languages in aviation and aeronautics domain are non-existent. Though English is the medium of instruction in many institutions in India for non-native speakers of English, the availability of corresponding terms of these abbreviations in Indian languages and their meaning are a different aspect altogether. For many who wish to make into the lucrative career of maintenance, ATC, support staff and aviation-related jobs these transliteration and translation will be of much help. With the huge number of MRO and aircraft manufacturing companies starting production factories in

India, Transliteration and Translation tools will be helpful in introducing potential candidates to the Aviation domain.

11. Unavailability of Monolingual & Bilingual Corpora

MT systems such as SMT and NMT are parallel corpora dependent. The absence of parallel corpus for any particular domain means that the MT application will not be able to translate the words and sentences (**Figures 1 and 2**). Before 2020 there were no parallel corpora available in the aviation domain for any Indian language. The first known corpora was proposed and completed by the faculties of Department of Computer Science, Assam University for English-Bengali Language pair. The corpus was developed, keeping in mind the complexity of the aviation domain and the vocabulary size was determined through OpenNMT while training the NMT system. Given its uniqueness, the corpus consists of hundreds of aviation OOV words and phraseologies. The Corpus was made to go through pre-processing steps and thus cleaned, tokenized and lemmatized (for both English and Bengali languages). The source of the English aviation sentences ranges from NASA ASRS reports to AAI and DGCA reports.

12. Scope in MT and NLP

The challenges posed by aviation domain in the field of NLP, especially for Machine Translation also create a huge scope for researchers and developers. It is an unexplored avenue that needs immediate attention. It can open up huge opportunity for researchers in the form of creating monolingual and bilingual corpora, both Preprocessing and post-processing tools and E-dictionaries among others. Some important work that has been carried out by faculties of Computer Science, Assam University, Silchar for the English-Bengali language pair in the aviation domain are as listed in the following **Table 9** [1, 2, 10, 11].

Thus scope of work exists in implementation of E-Dictionary, Pre-processing tools, Post-Processing tools, text analyzers, and MT systems between English and Indian native languages for aviation, aero-space and other specialized and technical domains.

Name of the paper	Contributions	Published in
NLP tools used in civil aviation, A survey. (S. Paul B.S. Purkyastha, P. Das)	Survey and Identification of NLP tools used in the aviation Domain	IJARCS, Vol 9., No.2, March –April 2018, ISSN: 0976-5697, Page:110-114
English to Bengali Transliteration tool for OOV words common in Indian civil aviation (S. Paul, B.S. Purkaystha)	Design and Development of 1st known transliteration tool of aviation OOV words for English-Bengali language pair	JADMS, Volume 6, issue 1, April 2019, ISSN: 2393-8730
Bilingual Technical E-Dictionary for aviation OOV words. (S. Paul, B.S. Purkhyastha)	Design and Development of 1st known E-Dictionary for Aviation OOV words for English-Bengali pair	IJEAT, Volume 9, issue 2, December 2019, ISSN: 2249-8958
Handling Aviation OOV words for Machine translation and corpus creation	Design and Development of Corpus Creation tool. Translation Assistance tool for Bengali language	IJCSE, Volume 11, No.5 Sep-Oct-2020, ISSN: 0976-5166

Table 9.

List of work in the field of aviation domain for English-Bengali language pair.

13. Importance of achieving phonetic equivalence

Aviation/Aeronautics English consists of OOVs, Phrases and Phraseologies in their hundreds, as a result in order to create a parallel corpus for English-Indian language requires us not only to create the native language equivalent translation but also the phonetic equivalent terms of the aviation OOV words. Phonetic equivalence of OOV words can be created through use of a standard phonetic keyboard for that particular language. For Bengali the AVRO keyboard is a handy tool. **Figure 3** depicts the phonetic layout for the Avro Keyboard. Phonetic equivalent words are useful for development of transliterated words. Transliterated words play a huge role in creation of aviation and technical corpora / database.

14. Conclusion and future scope

The native Indian languages present a huge scope for researchers to work on, specially for technical and unexplored domains. Though untouched till 2017, but work has begun on the aviation domain for Indian languages (English-Bengali pair) and we can conclude with the following points: The first known implementation of English-Indian language NMT based MT system for the aviation domain has been carried out and published with satisfactory results, is documented as “Detailed analysis of successful implementation of aviation NMT system and the effects of aviation post-processing tools on TDIL tourism corpus”, Saptarshi Paul, Bipul Syam Purkaystha, Journal of KING SAUD university-computer and information sciences [12].

Scope exists for development of English-other native Indian languages MT systems and NLP tools.

Development of MT systems for specialized and technical domains such as Aero-Space, Aircraft Maintenance manuals etc. are still unexplored.

Future Scope of research and work exists for NLP applications for Foreign Languages- Indian Languages pair as well as Indian- Indian Languages pair too. We can list some of them as follows:

Creation of Chat bots for the aviation domain in Indian languages: This application can help in reaching out to target audience and potential customers, chat bots are already in use for various tourism sites and can be easily extended for aviation related applications and WebPages.

Development of E-dictionary for Indian languages: As Indians are all non-native speakers of English so huge scope prevail in creation of aviation / aero-science E-Dictionaries.

Development of aviation Machine Translation systems: This application can find its use with travelers as well as people related with aviation industry. The



Figure 3.
Avro keyboard layout.

ability to translate aviation related sentences from English/French to an array of Indian languages can also help people engaged in aviation sector in faster and better understanding the maintenance manuals of Boeing/Airbus/ATR etc. The above mentioned applications can all find its way into maintenance, repair, operations, and aeronautics training institutes. Apart from the mentioned points, academic interests may include development of Email filters, Smart assistants, Predictive analysis Digital phone calls, Data analysis and Text analytics among others for aviation sentences.

Though only a handful of NLP tools have been developed for English and Indian languages, the number of Translation tool is restricted to only one [13], that too at an experimental level.

Translation Tools between English and Indian Languages can not only be helpful for travelers but also for aspiring candidates appearing for various airlines regulatory bodies such as DGCA and AAI.

The non-native speakers of English can greatly benefit from these tools and MT applications in enhancing their skills and thus improving their chances of clearing the various tests and ultimately fulfilling their dream of reaching out to the sky.

IntechOpen

Author details

Saptarshi Paul

Computer Science Department, Assam University, Silchar, India

*Address all correspondence to: paulsaptarshi@yahoo.co.in

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Saptarshi paul, Bipul shyam purkhyastha, "NLP tools used in civil aviation: A Survey", International journal of advanced research in computer science, volume 9, Number 2, March-April-2018
- [2] Saptarshi Paul, Bipul Shyam Purkhyasta. English to Bengali Transliteration Tool for OOV words common in Indian Civil Aviation. Journal of Advanced Database Management & Systems. 2019; 6(1): 23-32p.
- [3] AAI MATS, (1995). Available from: <https://www.aai.aero/en/system/files/resources/1%20Commercial%20Manual%202019.pdf> [Accessed: 2020/11/27]
- [4] DGCA reports, (1999). Available from: http://dgca.nic.in/accident/reports/contents_acc_rep.html [Accessed: 2020/07/18]
- [5] IBM WATSON, (2010), Available from: <https://www.ibm.com/watson/stories/aviation.html> [Accessed: 2017/10/10]
- [6] NLP in AIRBUS (2016). Available from: <http://www.intelligentiaerospace.com/pt/2016/10/21/applying-natural-learning-processing-to-airbus-helicopters.html> [Accessed: 2021/02/27]
- [7] NLP in MRO (2015). Available from: <http://aviationweek.com/aviation-maintenance-and-support-software/applying-natural-learning-processing-airbus-helicopters>[Accessed: 2020/12/02]
- [8] Peter Clark and Phil Harrison, "Boeing's NLP System and the Challenges of Semantic Representation" WA 98124, ACL, STEP '08 Proceedings of the 2008 Conference on Semantics in Text Processing, Pages 263-276, Venice, Italy, September 22 - 24, 2008 .
- [9] Pierre Isabelle and Laurent Bourbeau, "TAUM-AVIATION: its Technical Features and some experimental Results", Computational Linguistics, Volume 11, Number 1, January-March 1985. Pages: 18-27.
- [10] Saptarshi paul, Bipul Shyam Purkhyasta "Bilingual (English to Bengali) Technical E-Dictionary for Aviation OOV Words", International journal of Engineering and advanced technology, volume 9, issue 2, December 2019
- [11] Saptarshi Paul, Bipul Shyam Purkhyasta. English to Bengali Transliteration Tool for OOV words common in Indian Civil Aviation. Journal of Advanced Database Management & Systems. 2019; 6(1): 23-32p.
- [12] Saptarshi Paul, Bipul Syam Purkaystha, "Detailed analysis of successful implementation of aviation NMT system and the effects of aviation post-processing tools on TDIL tourism corpus", Journal of KING SAUD university-computer and information sciences, ISSN: 1319-1578, publisher: ELSEVIER <https://doi.org/10.1016/j.jksuci.2021.01.016>, 2021.
- [13] Bipul Syam Purkaystha, Saptarshi Paul "English to Bengali Neural Machine Translation System for the Aviation Domain", Infocomp Journal of Computer Science E-ISSN: 1982-3363, Vol. 19, No. 2, December 2020, pp. 78-97