CREATING AN ENGLISH TECHNICAL WORD LIST FOR THAI AIR TRAFFIC CONTROL OFFICERS: A CORPUS-BASED APPROACH

PAEVA ASAWASONGKRAM

Graduate School Srinakharinwirot University

2024

การสร้างรายการคำศัพท์ภาษาอังกฤษเฉพาะเพื่อเจ้าหน้าที่ควบคุมจราจรทางอากาศชาวไทยโดยใช้คลังข้อมูลภาษา

แพรวา อัศวสงคราม

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ศิลปศาสตรมหาบัณฑิต สาขาวิชาภาษาอังกฤษ
คณะมนุษยศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2567

CREATING AN ENGLISH TECHNICAL WORD LIST FOR THAI AIR TRAFFIC CONTROL OFFICERS: A CORPUS-BASED APPROACH

PAEVA ASAWASONGKRAM

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF ARTS

(English)

Faculty of Humanities, Srinakharinwirot University

2024

THE THESIS TITLED

CREATING AN ENGLISH TECHNICAL WORD LIST FOR
THAI AIR TRAFFIC CONTROL OFFICERS: A CORPUS-BASED APPROACH

BY

PAEVA ASAWASONGKRAM

HAS BEEN APPROVED BY THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE MASTER OF ARTS
IN ENGLISH AT SRINAKHARINWIROT UNIVERSITY

(Assoc. Prof. Dr. Chatchai Ekpanyaskul, MD.)
Dean of Graduate School

ORAL DEFENSE COMMITTEE

..... Major-advisor
(Dr.Pipittaporn Inpanich)

..... Chair
(Assoc. Prof. Dr.Woravit Kitjaroenpaiboon)

..... Committee
(Dr.Piyaporn Chaya)

Title	CREATING AN ENGLISH TECHNICAL WORD LIST FOR THAI AIR TRAFFIC CONTROL OFFICERS: A CORPUS-BASED APPROACH
Author	PAEVA ASAWASONGKRAM
Degree	MASTER OF ARTS
Academic Year	2024
Thesis Advisor	Dr. Pipittaporn Inpanich

Thailand's aviation sector is essential to the country's economic growth and international integration, and places high demands on Air Traffic Control Officers (ATCOs) to uphold both safety and operational efficiency. However, many Thai ATCOs encounter difficulties in comprehending International Civil Aviation Organization (ICAO) publications due to limited vocabulary knowledge of specialized English vocabulary. To address this issue, this research aimed to examine the lexical coverage required for Thai ATCOs to understand 95% of the text and to develop a technical word list tailored for Thai ATCOs to support their understanding of ICAO publications. A corpus-based approach was employed by compiling 19 ICAO Annexes into a written corpus of over 1.2 million tokens. The findings indicated that Thai ATCOs must possess knowledge of approximately 4,000 word families including proper nouns, marginal words, transparent compounds, and acronyms from the BNC/COCA word family lists to understand 95% of the Annexes. Additionally, the word list was constructed using five key criteria; frequency, range, lexical profiling, keyword analysis, and expert viewpoints which resulted in 139 technical words. This resource can help Thai ATCOs interpret ICAO publications clearly, resulting in more effective communication and greater safety in air traffic operations.

Keyword : technical word list, corpus-based approach, lexical coverage, corpus linguistics, air traffic control officers

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my research advisor, Dr. Pipittaporn Inpanich, for her unwavering support, insightful guidance, and endless patience throughout the entire course of this study. Her encouragement, expertise, and dedication have been invaluable to the completion of this thesis. I am truly grateful for the direction and reassurance she consistently provided.

I would also like to extend my heartfelt thanks to my parents, Mr. Sutthisakdi and Mrs. Puangpen Asawasongkram, whose unconditional love, support, and belief in me have been my greatest source of strength. I am deeply thankful for all the opportunities they have given me, without which this journey would not have been possible.

My sincere appreciation goes to the thesis committee members: Asst. Prof. Dr. Sakulrat Worathumrong, Dr. Aranya Srijongjai, Dr. Nattapra Wongsittikan, and Dr. Piyaporn Chaya. Their constructive feedback and insightful comments have significantly enriched this research. Additionally, I want to thank you Dr. Narathip Thumawongsa for her dedication to the faculty and for introducing me to my advisor, without her I would have not made it this far. I am also especially grateful to Assoc. Prof. Dr. Woravit Kitjaroenpaiboon, whose inspiration and valuable suggestions played an important role in shaping this study.

Lastly, I would like to thank my classmates for their constant encouragement and companionship throughout this journey. I am equally thankful to my colleagues and supervisor at work for their understanding and continuous support, which allowed me to balance my academic and professional responsibilities. I could not have accomplished this without them.

PAEVA ASAWASONGKRAM

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1
INTRODUCTION

This chapter provides an overview of the study by first explaining the background of the study and the reasons that motivated this research study. Then it outlines the research objectives and research questions that the study seeks to answer. The chapter also highlights the significance of the research by explaining how it contributes to academic knowledge and practical applications. Additionally, it defines the scope of the research and lists the important terms that are used throughout the study to ensure clarity and consistency in interpretation.

Background of the Study

The aviation industry is considered an important part of Thailand's economic and tourism growth as aviation transportation allows other countries to connect with Thailand. Therefore, the Airports of Thailand Public Company Limited (AOT) aims to position Thailand as a thriving global aviation hub to encourage tourism, generate economic prospects, and foster increased stability (PRD, 2024). A hub refers to a location where passengers can transfer to connecting flights to reach their destination (OAG, 2023). Hubs provide better accessibility for passengers and attract more traffic for airlines and airports (Yang & Liu, 2019). This results in more tourist traffic which is a key source of income generation for the country's economy (Lohmann et al., 2009).

Safety is a critical consideration for tourists when selecting a destination, as a lack thereof can lead to an unsatisfying experience (Fuchs et al., 2011). This includes safety in aviation transportation, as airplane crashes can impact tourists' travel behavior, economy, and destination image (Fuchs et al., 2011). Therefore, safety allows tourists to ensure that Thailand is a destination worth travelling to and is one of the main responsibilities of air traffic control officers (ATCOs). ATCOs are part of air traffic control units responsible for air traffic services (ATS). The purpose of ATS is to prevent collisions between aircraft and on the maneuvering area between aircraft and obstructions, facilitate the smooth and organized flow of air traffic, provide information

useful for the safe and efficient conduct of flights, and notify appropriate organizations regarding aircraft in need of search and rescue aid by providing assistance as required (ICAO, 2018).

In order to achieve air safety goals, ATCOs must consult the International Civil Aviation Organization (ICAO) publications regarding Air Traffic Management to ensure regular compliance, understand operational procedures, promote harmony and consistency, and engage in continuous professional development. ICAO publications are published to develop principles and techniques of international air navigation to foster the planning and development of air transport (ICAO, 2023). ICAO publications handle over 12,000 Standards and Recommended Practices (SARPs) that are essential for Contracting State (Dib, 2022). Contracting States are country members of ICAO that have agreed to work together to achieve a high level of consistency in regulations, standards, procedures, and organization pertaining to aircraft, personnel, airways, and support services by establishing uniformity whenever beneficial (ICAO, 2006). For example, Thailand is a Contracting State and is the location of the Asia and Pacific Region Office.

There are a total of nineteen Annexes that display the SARPs based on each specific topic. The Annexes further explain what SARPs are and how each Contracting State can achieve them. Each Contracting State is invited to apply the Annexes' text for their own national regulations. In addition, there are six Procedures for Air Navigation Services (PANS) which are documents that are used as supplements to Annexes. PANS may become SARPs when they are deemed necessary for adaptation. Moreover, technical manuals are documents that offer support and further details to ensure consistent implementation of SARPs for the Contracting States to follow for uniform safety air traffic control (ICAO, 2023).

These ICAO publications are published in six languages including Arabic, Chinese, English, French, Russian, and Spanish. In Thailand, as Thai ATCOs must consult the ICAO publications in English, they must have a high understanding of Aviation English. This refers to utilization primarily for aviation purposes between aviation

personnel, including ATCOs, flight crew members, and aircraft maintenance staff (Estival & Farris, 2016). Aviation English differs from other domains of English for specific purposes (ESP) as it extends beyond mere language proficiency to encompass statutory compliance and adherence to regulatory mandates.

Understanding these publications is an important part in ATCO training, as it is integral for them to achieve overall safety. In Thailand, training for Thai ATCOs implements a competency-based training programme to ensure standardized application of knowledge, skills, and attitude. The training is divided into three phases including initial training (basic training and rating training), unit training (pre-on-the-job training and on-the-job training), and continuation training (conversion training and refresher training) (ICAO, 2017). First, the initial training is provided for training the fundamental knowledge for specialized air traffic control training. Second, the unit training is provided for the operational work situation involving the usage of simulation facilities and live training with a supervisor. Third, the continuation training is for further training to improve competency in air traffic control (SKYbrary, n.d.). In the training, aviation personnel especially Thai ATCOs are trained based on the ICAO publications to achieve air safety goals.

English is important for ATCOs, especially those who are non-native speakers. However, non-native English speaking ATCOs especially those who have low English proficiency may face problems in their work that can lead to misunderstanding and dangerous accidents. A research study by Mekkaoui and Mouhadjer (2019) concluded that Algerian ATCOs lacked listening comprehension skills, which are essential for understanding entire conversations, and the vocabulary knowledge, which is needed to completely express their thoughts. As a result, ATCOs with limited proficiency may cause safety issues during communication. Similarly, Kaur (2021) found that aviation maintenance requires proficient writing skills with a basic grammar knowledge and good reading comprehension to follow specific instructions in maintenance manuals. The findings were in accordance with previous studies by Rahmat and As'ary (2017) and Sukma et al. (2019) whose studies suggest that adequate reading comprehension will

reduce misunderstandings and misinterpretations between aviation maintenances. As a result of limited English proficiency among aviation personnel, these misunderstandings and misinterpretations may result in lower safety levels. Therefore, English is an important part of ATCOs, especially Thai ATCOs.

Some Thai ATCOs have lower-than-expected scores on the ICAO Language Proficiency Test. Based on this serious issue, it might be possible for them to misunderstand ICAO publications, which leads to accidents in the future. Insufficient vocabulary knowledge constitutes a contributing factor to misunderstandings as it can affect language competency and utilization (Viera, 2017). Therefore, understanding vocabulary plays a pivotal role in acquiring proficiency in a foreign language which is critical for clear comprehension. A corpus is a tool that can help readers and listeners learn and understand vocabulary. A corpus is a systematically gathered collection of authentic spoken or written language which is stored electronically and analyzed for linguistic patterns using computer programs (Biber et al., 1998; Park & Nam, 2017). Each program provides a variety of methods for analysis, such as frequency lists, keyword lists, and concordance analysis (Ngula, 2017). For example, teachers can use a corpus in discovering vocabulary, grammar, and language usage of a foreign language *(Dazdarevic et al., 2015)*.

A type of analysis called a corpus-based approach aims to test linguistic hypotheses or theories, particularly those reliant on introspection rather than empirical corpus evidence, by examining corpus data to determine their support, modification, or refutation (Ngula, 2017). A corpus-based approach can be used in investigating the lexical coverage of a text. Lexical coverage is considered as the degree of knowledge readers must have to comprehend the text (Webb, 2021). Laufer and Ravenhorst-Kalovski (2010) argued that 95% of text coverage is enough for minimal comprehension. Therefore, it is crucial to measure the vocabulary load Thai ATCOs must possess to be able to understand ICAO publications with 95% of the text coverage.

To determine how many word families Thai ATCOs must know to comprehend text, a corpus should be developed, and it can also be used to create an English

technical word list. Furthermore, creating a technical word list can be used at the trainings to assist Thai ATCOs in enhancing their understanding of ICAO publications by enabling them to identify the technical terms. Subsequently, creating an English technical word list will help them to achieve the lexical coverage required to understand the text. Moreover, as Aviation English is a part of ESP, creating a technical word list can allow Thai ATCOs to distinguish between technical words and general English words.

In conclusion, Thai ATCOs need to understand written ICAO publications to ensure the safety of international air transport, so it is important for them to have adequate knowledge of vocabulary. Therefore, it is imperative to develop an English technical word list for Thai ATCOs. In previous research, there is an emphasis on a spoken corpus about communication between ATCOs, pilots, and aviation maintenance (e.g., ATC Corpus, ATCOSIM, ATCO2 Corpus, ATCC). Despite the abundance of a spoken corpus, there are not many studies related to a written corpus about ICAO publications. For this reason, the researcher has decided to create a written corpus using ICAO publications to develop an English technical word list for Thai ATCOs.

Overall Aim and Research Objectives

The overall aim of this research is to form an English technical word list for Air Traffic Management for Thai ATCOs. The research objectives are as follows:

1. To measure the vocabulary load Thai ATCOs must possess to be able to understand ICAO publications with 95% of the text coverage.

2. To create an English technical word list for Thai ATCOs.

Research Questions

1. How many word families are required to understand the vocabulary in ICAO publications with 95% of the text coverage?

2. What are the technical words to create an English technical word list for Thai ATCOs?

Significance of the Study

The results of the study will benefit Thai ATCOs, ATCOs' trainers, and instructors. Firstly, ATCOs must consult ICAO publications in English to understand the procedures related to safety. This technical word list can be used as a guide when reading ICAO publications to distinguish whether the terms are technical words or general English. It is important for ATCOs to know the technical words in order to understand and achieve safety after consulting ICAO publications. Secondly, the technical words found from the technical word list can be used as the teaching materials during competency-based training for ATCOs. Lastly, instructors can incorporate the word list for familarization so that they are able to use it in real-life situations and in professional practice.

Scope of the Study

This study selected the 19 ICAO Annexes to the Convention on International Civil Aviation. This resulted in 30 ICAO publications as some Annexes contain multiple volumes from the ICAO Products & Services Catalogue 2023, and these publications were used to build a corpus and create an English technical word list for Thai ATCOs. The researcher did not include restricted publications as they may contain sensitive information related to security and intellectual property, which may require limitations in access.

Definitions of Terms

1. **Technical Word List** is a collection of specialized vocabulary or terminology associated with a particular field. This list is designed to facilitate comprehension within specialized contexts by providing a standardized set of terms with precise meanings (Davis & Navarro, 2015; Smith & Johnson, 2018). In this study, the technical word list is developed based on the criteria suggested by Rungrueang et al. (2022) including frequency, range, lexical profiling, keyword analysis, and expert viewpoints. High frequency words from General Service List (GSL) and academic words from Academic Word List (AWL) are used for lexical profiling.

2. **Air Traffic Control Officers (ATCOs)** are individuals tasked with delivering air traffic control services and bearing the responsibility of overseeing the safe, orderly, and expeditious management of air traffic (Crocker, 2005; Jeppesen, 2012).

3. **Thai Air Traffic Control Officers (Thai ATCOs)** are air traffic control officers that are Thai nationals and work in Thailand.

4. **Lexical Coverage** refers to the extent of vocabulary understanding required for text comprehension (Webb, 2021). According to Laufer and Ravenhorst-Kalovski (2010), to achieve a 95% of text coverage suffices for basic understanding, which is imperative for readers to comprehend texts across various contexts.

5. **Vocabulary Load** indicates the number of word families needed in order to read a text (Coxhead & Demecheleer, 2018).

6. **International Civil Aviation Organization (ICAO)** constitutes a specialized agency within the framework of the United Nations (UN), established pursuant to the Convention on International Civil Aviation. Its mandate comprises of establishing standards and principles pertaining to safety, security, efficiency, and regularity in the realm of international air transport (ICAO, 2023).

7. **Contracting States** are countries that are members of ICAO that have consented in cooperating to achieve a high level of consistency in standards, regulations, procedures, and organizational structures related to airways, aviation personnel, aircrafts, and support services. This cooperation is aimed at promoting uniformity whenever deemed beneficial (ICAO, 2006).

8. **Aviation English** is defined as a specialized branch of the English language tailored for proficient communication amidst aviation professionals including pilots, ATCOs, flight attendants, and ground staff within the aviation industry. It encompasses specialized lexicon, phraseology, and communication protocols pertinent to aviation operations (Ragan, 1996).

9. **Standards and Recommended Practices (SARPs)** delineate the standardized procedures aimed at ensuring the safe, regular, and efficient provision of air services. These procedures are established, approved, and periodically revised through the

ICAO Annexes to The Convention on International Civil Aviation (Chicago Convention) (Florio, 2006). “Standards” are described as instructions Contracting States comply in correspondence to the Convention fostering uniformity in implementation for the safety or regularity of international air transport. “Recommended Practices” are recognized as beneficial for ensuring safety, regularity, or efficiency of international air transport (ICAO, 2018).

10. Annexes are ICAO publications that are the framework for establishing and upholding international SARPs, which are foundational principles aligned with the Chicago Convention and ICAO’s mission and role (ICAO, 2023).

11. Procedures for Air Navigation Services (PANS) are ICAO publications that are recognized as additional regulations that, while not formally endorsed by the ICAO Council, have been authorized as a comprehensive guideline to be followed to the greatest extent feasible (Pepin, 1952).

12. Technical Manuals are ICAO publications that provide consistent application of international SARPs for guidance and information (ICAO, 2023).

CHAPTER 2
LITERATURE REVIEW

This chapter will contains a review of previous research and relevant literature to provide background information. The first section provides information on ICAO publications for air traffic control officers (ATCOs). The second section illustrates the classification of words. The third section describes the lexical coverage for people reading English. The fourth section explains corpus linguistics and corpus construction. The fifth section presents a corpus-based approach, followed by the sixth section about the criteria for developing a corpus. The seventh section defines what word lists are and how they are helpful in teaching vocabulary. The eighth section includes general service list (GSL) and academic word list (AWL). The nineth section describes the criteria for creating a word list. Finally, the last section explores the related studies on creating word lists.

ICAO Publications

The International Civil Aviation Organization (ICAO) is an agency under the United Nations (UN) that focuses on aiding Contracting States with the aim of cooperative utilization of airspace for mutual advantage (ICAO, n.d.-a). Through its active involvement, ICAO aids the 193 Contracting States to the Chicago Convention 1944 in the formulation and adoption of international standards, practices, and policies governing international civilian flights. The Chicago Convention 1944 was established to develop a foundation for the standards and procedures for harmonic air navigation worldwide (ICAO, n.d.-b). Therefore, Contracting States must comply with the ICAO Standards and Recommended Practices (SARPs) which are written in the ICAO Publications as it they provide the core foundation for safety and efficiency for international air transport.

Standards and Recommended Practices (SARPs)

ICAO Standards and Recommended Practices (SARPs) concern related matters that deal with safety, regularity, and efficiency of air navigation, such as communication systems and air navigation aids, rules of the air and air traffic control practices, and aircraft in distress and investigation of accidents. “Standards” are mandatory requirements that Contracting States of ICAO must follow to reach the minimum level of safety and operational effectiveness as they provide a baseline for aviation operations and are obligated to incorporate these standards in their national aviation laws. “Recommended Practices” are guidelines that Contracting States are encouraged to follow to enhance aviation safety and efficiency. These “Recommended Practices” are more flexible than “Standards”, but Contracting States must try to maintain consistency and global aviation standards. SARPs can be found in Annexes, PANS (Procedures for Air Navigation Services), and Technical Manuals. Further details are as follows.

1. Annexes are regulatory standards that legally bind Contracting States to incorporate into their legal framework for civil aviation and amendment process that they can participate to provide input on proposed changes.
2. Procedures for Air Navigation (PANS) are implementation guidance that is complementary to the Annexes and provide support to practical applications by being flexible while the fundamental objectives of the Annexes are maintained. These PANS do not have the same status as SARPs. However, PANS may eventually become a part of SARPs when they reach maturity and stability. This process takes around two years, contingent on the nature and priority of the proposal being evaluated.
3. Technical Manuals are documents that contain additional guidance as they further expand specific topics on the provisions of the Annexes and PANS.

Figure 1 illustrates ICAO publications which contain 19 Annexes, 6 PANS documents, and 11 areas of Technical Manuals. Each area consists of many documents. For example, AGA – Aerodromes includes 29 Technical Manuals while ATM – Air Traffic Management includes 30 Technical Manuals.

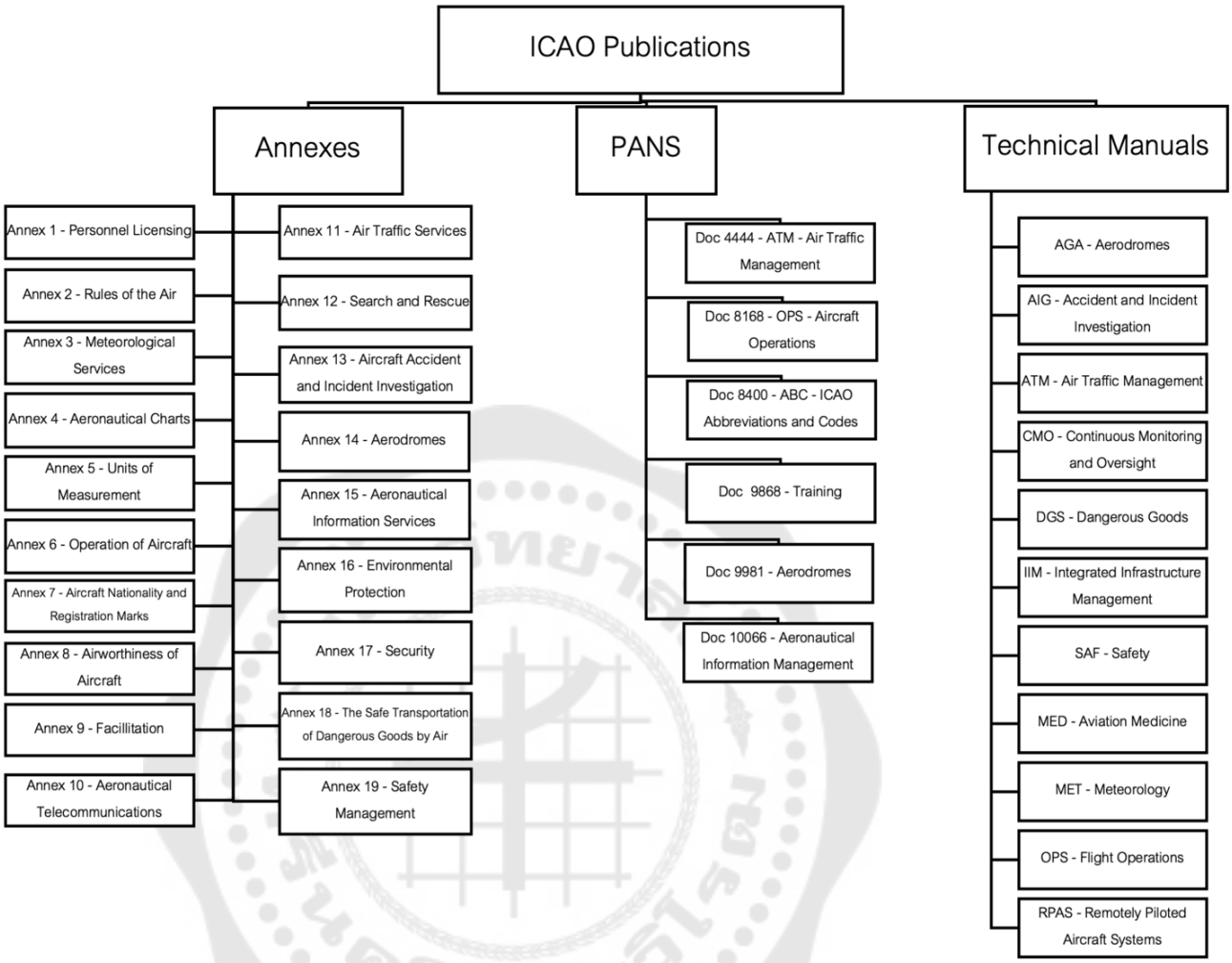
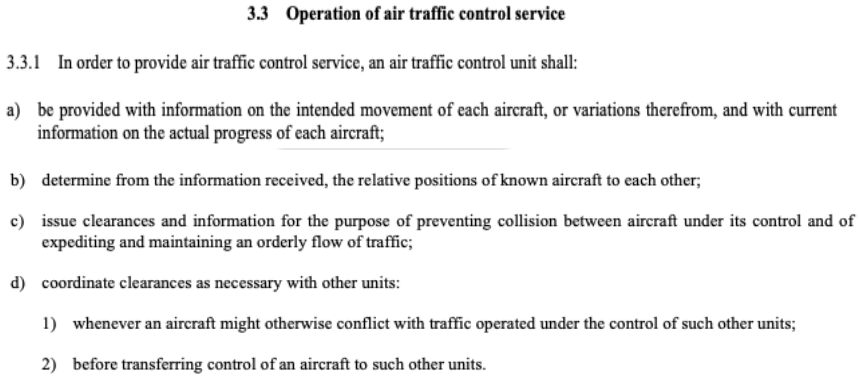


Figure 1 ICAO Publications

Source: ICAO. (2023). Products & Services Catalogue.

Figures 2 - 4 show examples of extracts from Annexes, PANS, and Technical Manuals. Figure 2 is an extract from Annex 11 – Air Traffic Services. It represents the responsibilities ATCOs have in order to provide air traffic services (ATS) to the Contracting States' airway. In the extracted text, ATCOs must be provided with information regarding the activities of the aircraft, corresponding positions of the aircraft, issue clearances and information for safe and orderly flow of air traffic, and coordinate with relative units. Figure 3 is an extract from PANS Doc 4444 – Air Traffic Management. It states the general provisions and procedures regarding ATS safety management. The extracted text explains that Contracting States should ensure ATS, and related procedures are standardized to maintain safety and follow regional air navigation agreements. Figure 4 is an extract from Doc 9432 – Manual of Radiotelephony. It provides a description of where phraseologies used for radiotelephony communications are printed in and where it can be consulted. According to the extracted text, language for radiotelephone communications is regulated by SARPs in Annex 10 – Aeronautical Telecommunications and PANS-ATM, which the understanding of phraseologies is crucial for their proper use.



3.3 Operation of air traffic control service

3.3.1 In order to provide air traffic control service, an air traffic control unit shall:

a) be provided with information on the intended movement of each aircraft, or variations therefrom, and with current information on the actual progress of each aircraft;

b) determine from the information received, the relative positions of known aircraft to each other;

c) issue clearances and information for the purpose of preventing collision between aircraft under its control and of expediting and maintaining an orderly flow of traffic;

d) coordinate clearances as necessary with other units:

1) whenever an aircraft might otherwise conflict with traffic operated under the control of such other units;

2) before transferring control of an aircraft to such other units.

Figure 2 Extract from Annex 11 – Air Traffic Services

Source: ICAO. (2016). Annex 11 – Air Traffic Services. p.53.

2.1 GENERAL

2.1.1 States shall ensure that the level of air traffic services (ATS) and communications, navigation and surveillance, as well as the ATS procedures applicable to the airspace or aerodrome concerned, are appropriate and adequate for maintaining an acceptable level of safety in the provision of ATS.

2.1.2 The requirements in respect of services, systems and procedures applicable to airspaces and aerodromes should be established on the basis of a regional air navigation agreement in order to facilitate the harmonization of ATS in adjacent airspaces.

2.1.3 To ensure that safety in the provision of ATS is maintained, the appropriate ATS authority shall implement safety management systems (SMS) for the air traffic services under its jurisdiction. Where appropriate, ATS SMS should be established on the basis of a regional air navigation agreement.

Figure 3 Extract from Procedural for Air Navigation Services – Air Traffic Management (Doc 4444)

Source: ICAO. (2016). Procedural for Air Navigation Services – Air Traffic Management p. 37.

3.2 AN EXPLANATION OF THE ROLE OF PHRASEOLOGIES AND PLAIN LANGUAGE IN RADIOTELEPHONY COMMUNICATIONS

3.2.1 The use of language in radiotelephony communications is governed by Standards and Recommended Practices (SARPs) and Procedures for Air Navigation Services (PANS) contained in Annex 10 — *Aeronautical Telecommunications* and the PANS-ATM. Specific language proficiency requirements are contained in Annex 1 — *Personnel Licensing*. ICAO phraseologies are published in Annex 10, Volume II — *Communication Procedures including those with PANS status* and the PANS-ATM. The phraseologies contained in these documents are not intended to be exhaustive, and both documents refer, in several instances, to the need for "additional phraseologies", or "appropriate subsidiary phraseologies", or "plain language". An explanation of the role of phraseologies in radiotelephony communications will clarify their appropriate use.

Figure 4 Extract from Manual of Radiotelephony (Doc 9432) (Technical Manual)

Source: ICAO. (2007). Manual of Radiotelephony (Doc 9432). p.34.

Standards and Recommended Practices for Air Traffic Control Officers

ICAO requires ATCOs to meet a number of criteria related to air traffic controller training that includes knowledge, skills, and practical experience requirements (ICAO, 2017), for example, air law (i.e., rules and regulations relevant to the air traffic controller) and operational procedures (i.e., air traffic control, communication, radiotelephony and phraseology procedures). Contracting States that provide ATS in their airspace must apply SARPs in a consistent manner to ensure flying internationally is accomplished under uniform conditions (ICAO, 2018).

ATCOs can access ICAO publications through various channels, including the official ICAO website, employer-provided libraries, and aviation training institutions. ICAO Annexes are essential resources for maintaining standardized global air traffic operations. ATCOs are required to consult these publications during situations, such as managing emergency scenarios, ensuring compliance with environmental regulations, and coordinating international flight operations. ICAO Annexes were selected for this study because they form the foundation of global aviation SARPs. These documents encompass critical areas of air traffic management, such as communication procedures, environmental sustainability, and safety measures, which directly impact ATCOs' responsibilities. Focusing on the ICAO Annexes allow the target of specific vocabulary Thai ATCOs need to understand and apply these regulations effectively to enhance operational safety and compliance within Thailand's aviation sector. However, PANS and Technical Manuals were not chosen as they provide additional guidance and details on operational effectiveness.

Figure 5 illustrates ICAO Annexes used in this research study to create an Air Traffic Management corpus and develop an English technical word list for Thai ATCOs.

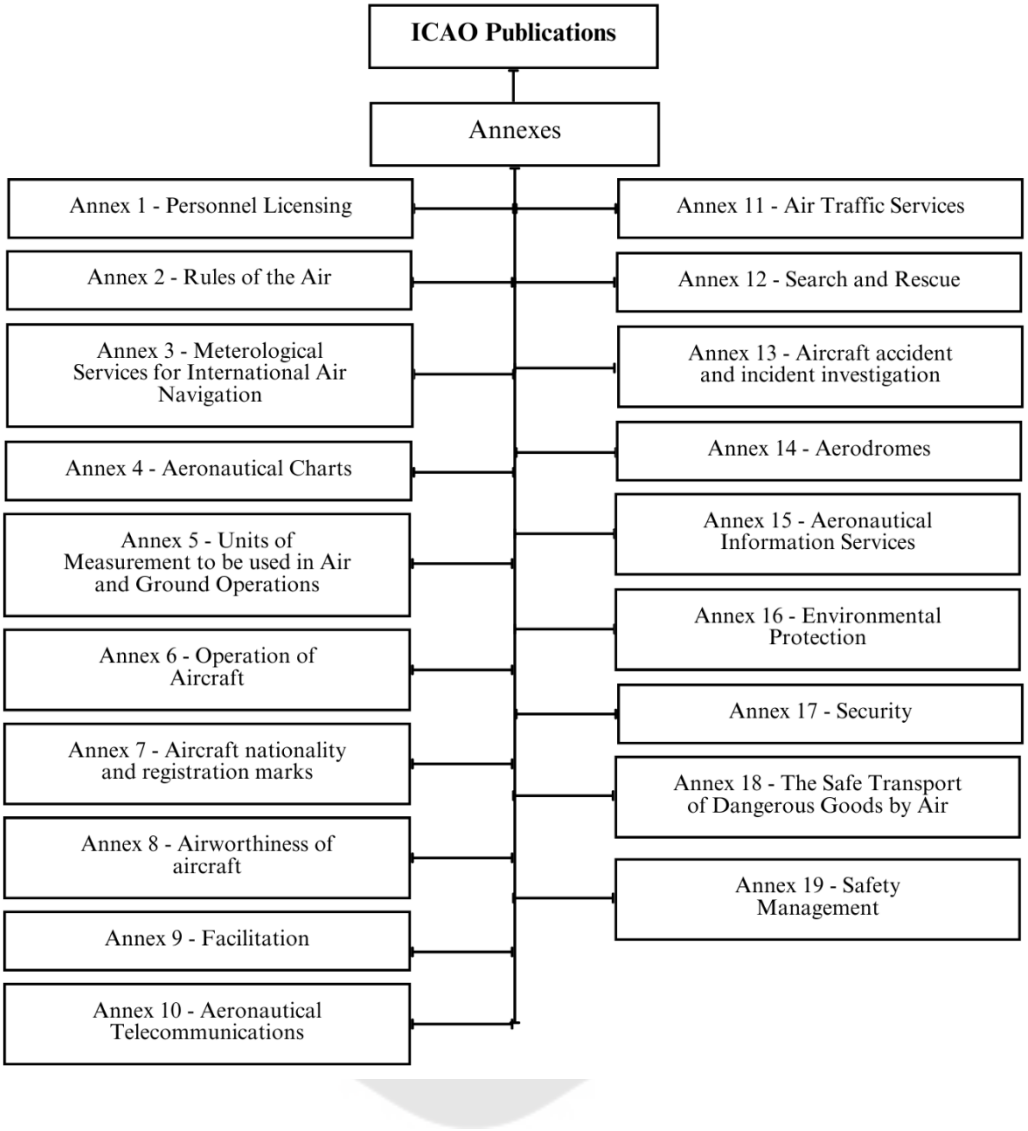


Figure 5 ICAO Annexes

Source: ICAO. (2023). Products & Services Catalogue.

ICAO Annexes have been employed by various researchers as an important source of data, providing valuable information to answer their questions. For example, Garcia (2023) investigated the construct of the ICAO rating scale used in aviation English listening tests. The researcher selected Annex 1 as a source to analyze the ICAO rating scale and the ICAO policy. Annex 1, the primary source of this study, provided information about the requirements that Contracting States must follow. The results of this study helped test developers, especially those who would like to access pilots and ATCOs' listening skills, come to a better understanding of the ICAO policy. Similarly, Drayton and Coxhead (2023) analyzed the language proficiency requirements (LPRs) for language testing in aviation contained in Annexes 1 and 10. The results revealed that ATCOs' beliefs about the role of plain language and standard phraseology in radiotelephony communication should be considered for further language training.

In contrast, Nagy (2019) focused on analyzing the pragmatic uses of modal verbs in Annexes 1, 2, 10, 17, 18, and 19. Those ICAO Annexes were selected in the study because they are related to obligation and recommendation. The findings indicated that modal verbs were used differently based on different contexts. For example, *shall* and *should* were used differently as *shall* refers to the meaning of vital obligation while *should* refers to chance or possibility. The results of the study also suggested that non-native airspace users should be aware of meaning interpretation when they read the ICAO Annexes.

While these studies highlight the use of Annexes as the valuable source for providing insights related to ICAO language testing and pragmatism, there remains a significant gap in employing Annexes to create a technical word list to help improve ATCOs' English language proficiency. Given this gap, there is a need for research that suggests technical words that ATCOs should know. By addressing this need, ATCOs will be able to reach the ICAO requirement, providing effective communication and safety in air traffic management. In this research study, there is a need to create a word list for ATCOs using a corpus-based approach.

Classification of Words

A word is the smallest grammatical unit in a language and can be referred to as building-blocks of a language (Carstairs-McCarthy, 2002; Lardiere, 2006). According to Nation (2001), there are four classifications of words including high-frequency words, academic words, technical words, and low-frequency words.

High-Frequency Words

High-frequency words are the words that are frequently seen in written and spoken text and have the widest ranges in either formal or informal language. These words are the top two thousand to three thousand words which are mostly function words (e.g., prepositions and articles) and common content words (e.g., forest, boy, and computer). For example, the General Service List (GSL) is considered as the most distinguished word list by Michael West in 1953. GSL comprises 2,000 of the most commonly used words, accounting for 78-81% of written texts and 85% of spoken texts.

Academic Words

Academic words are the words that appear in academic textbooks or academic materials throughout different disciplines, such as linguistics, politics, and business. These words cover around ninety-eight percent of the texts to allow unassisted comprehension. The most well-known word list for academic words is the Academic Word List (AWL) by Coxhead (2000). These types of words are for learners who want to read and write academic text or those who study at university level (Corson, 1997).

Technical Words

Technical words or technical terms usually referred by Cabré and Sager (1999) are “domain-specific words” which means that they are related to only one specific area that may or may not require specialized knowledge to be understood (Chung & Nation, 2004). These terms include acronyms, abbreviations, chemical formulas, symbols, and multi-word phrases. Coxhead and Nation (2001) assumed that technical terms account for five percent of words in academic texts. However, Chung and Nation (2003)

indicated that technical terms cover one-third of the text, which covers twenty to thirty percent of the tokens. Technical terms can be determined by systematically limiting the topics or about one thousand entries in a technical dictionary.

Low-Frequency Words

Low-frequency words are large groups of words that barely occur or are rarely used terms in a text. These words are considered only five percent of the words in academic texts and can be proper nouns, infrequent words, and technical terms of other academic disciplines.

From the four categories above, this study focuses on technical words to create a technical word list because Aviation English is used in a specific area that requires specialized knowledge. When Thai ATCOs know the technical words from the technical word list, they will be able to understand the ICAO publications regarding safety.

Lexical Coverage

Lexical coverage refers to the amount of words in input that are known by English language listeners and readers (Nation, 2006; Webb, 2021). In reading context, lexical coverage is different from reading comprehension. To specify, lexical coverage refers to the percentage of words that readers know to understand the text while reading comprehension refers to the ability to understand the text's meaning (Yildiz, 2023). A reader's reading comprehension does not depend on only the lexical coverage, but it depends on other factors, such as prior knowledge. For example, the more prior knowledge the reader has, the higher the rate of reading comprehension the reader will have. According to Laufer and Ravenhorst-Kalovski (2010), the lexical coverage for sufficient comprehension are 95% and 98%. 95% of text coverage of a hundred words would mean that there will be one word out of twenty that is unknown, and 98% would equal to one word out of fifty words that is unknown (Nurmukhamedov, 2017). The required vocabulary knowledge of 4,000 – 5,000 word families will allow readers to understand 95%, while a knowledge of 8,000 word families will permit readers to

comprehend 98%. In addition, Nation (2006) stated that 6,000 – 7,000 word families were essential to understand 98% of spoken text, as well as 8,000 – 9,000 word families were necessary to understand 98% of written text. However, Hsu (2011) claimed that 95% of text coverage that amounted to 5,000 word families was an attainable goal for EFL learners.

It is important to examine the lexical coverage of Thai ATCOs as it is essential for them to fully understand ICAO publications during ATCOs training. To achieve this, ATCOs in training need to understand the Annexes to follow standard procedures for air traffic safety which is critical for their professional development. To support this, training instructors can develop these strategies to help engage students and motivate them in understanding the language used in these Annexes. For example, it can be integrated in vocabulary exercises, guided reading sessions, and discussion-based activities to practice important technical terms.

To determine the vocabulary load of the text, the following programs are used for analysis: Range, VocabProfile, and AntWordProfiler (VUW, n.d.). Firstly, the Range program is utilized to identify the quantity and types of vocabulary present in a specific text or collection of texts, based on the GSL/AWL lists or British National Corpus lists. Secondly, the VocabProfile program is similar to Range, but is a simpler, web-based version. Lastly, the AntWordProfiler program is used to measure the vocabulary load of the text by comparing the files against BNC/COCA word family lists. AntWordProfiler represents a contemporary vocabulary analysis tool equipped with multiple supplementary functionalities. The program is accessible through Laurance Anthony's website, which can be operated on macOS and is compatible with word lists in diverse languages.

The AntWordProfiler program offers a range of features that have been widely utilized by researchers, including analyzing the list of word types, tokens, and groups. It provides vocabulary load indexes based on word frequency and range for every 1,000 word families from the BNC/COCA word lists. Thiankasem (2018) employed the program to assess vocabulary load across 25,000 word families and additional supplemental

lists. For the technical words, AntWordProfiler analyzed the corpus using the GSL and AWL to determine word range, specialized occurences, and frequency for developing a preliminary word list.

Previous research studies such as Yildiz (2023) and Hong and Wen (2025) examined the lexical coverage required to reach 95% and 98% respectively for reading comprehension. For example, Yildiz (2023) analyzed English reading passages from The Higher Education Institutions Examination in Türkiye during 2012-2021 to identify word families to reach 95% and 98% lexical coverage. The researcher found that in order to reach a 95% lexical coverage, learners must have a vocabulary knowledge of 4,000 word families for minimal comprehension. While to reach a 98% lexical coverage, learners must have a vocabulary knowledge of 8,000 word families for optimal comprehension. Similarly, Yu and Wen (2025) explored the lexical coverage requirements for EFL students when reading science news published in the science publication called Scientific American (SA). The research study reported that 5,000 and 10,000 word families were required to reach 95% and 98% lexical coverage respecitvely.

In summary, previous research studies have extensively studied the lexical coverage in academic texts used in examination as seen in Yidiz (2023). Additionally, Yu and Wen (2025) studied the lexical coverage related to English for Specific Purposes (ESP). However, there is still a notable gap on studying the lexical coverage in ICAO publications. As a result, in this study, there is a need to investigate word families which are required to understand the vocabulary in ICAO publications, especially Annexes, with 95% of the text coverage.

Corpus Linguistics and Corpus Construction

“Corpus” is a collection of natural spoken or written language gathered systematically that is stored as an electronic database and relies on computer software to analyze linguistic patterns (Biber et al., 1998; Park & Nam, 2017). According to Bennett (2010), there are 8 types of corpus linguistics, but only four are commonly used

in a classroom setting including a generalized corpus, a specialized corpus, a learner corpus, and a pedagogic corpus.

Firstly, a generalized corpus, also known as a sample or reference corpus, is used to provide the big picture of a language that can be called general and used in everyday situations (e.g., American English and Birish English) (Bennett, 2010; Vaughan & O'Keefe, 2015). Examples of large, generalized corpora are the American English Corpus (BNC) and the Corpus of Contemporary American English (COCA). The collection of text can grow according to new avaliable data and become large in size (Dash, 2010).

Secondly, a specialized corpus contains distinctive ranges of language, dialect, and subject as it is used for a specific purpose (Dash, 2010). It can be used to answer a particular question in a specific area, such as the Michigan Corpus of Academic Spoken English (MICASE), as it only consists of spoken language in a university environment. This type of corpus is mostly used in English for Specific Purposes (ESP) classrooms (Bennett, 2010).

Thirdly, a learner corpus is a type of specialized corpus that can be used by students learning a specific language to study written and spoken contexts (Bennett, 2010). These include common errors, features of a language, and language used by non-native speakers (Vaughan & O'Keefe, 2015). The Active Learning of English for Science Students (ALESS) Learner Corpus is an example of a learner corpus that contains research papers by Japanese science majors (Gilquin, 2015).

Lastly, a pedagogic corpus is a specialized corpus that is used especially for language teaching and learning purposes which are collected from textbooks, online resources, learner essays, and spoken interactions (Biber & Reppen, 2002; Vaughan & O'Keefe, 2015). It allows teachers to identify relevant, frequent, and useful patterns related to grammar that can help language learners develop more natural and idiomatic language usage (Biber & Reppen, 2002; Meunier & Granger, 2008). For instance, the International Corpus of Learner English (ICLE) encompasses a compilation of written texts produced by students from international backgrounds that range from essays and

summaries (Granger et al., 2015). This allows teachers to identify how language learners write and what errors they make, which will be used in their teaching approaches.

In the air traffic management context, there are several spoken-based air traffic control corpora that have been carried out. Firstly, the Air Traffic Control Simulation Speech (ATCOSIM) corpus was developed by Hofbauer et al. (2008) as a spoken corpus specifically for ATC applications. The researchers created the corpus to address the need of a domain-specifc corpus for spoken language technologies where it captures non-prompted and clean speech of ATCOs. The ATCOSIM contains ten hours of non-prompted and clean ATC phraseology recorded during real-time simulations in a ATC room. The development of the corpus involved multiple methodologies beginning with recordings made in the EUROCONTROL Experimental Centre's (EEC) air traffic control room in France which provided realistic simulations and working conditions. This included a strict set of rules that omitted punctuation marks; apostrophes for possessives and contractions; and numbers, letters, navigational aids, and radio call signs definition based on aeronautical standards. After that, the corpus was organized into directories which included audio files, transcriptions, and metadata presented in HTML files for the easy sorting, reviewing, and replaying of utterances. Lastly, the ATCOSIM Corpus was validated through a two-step process where automatic tests were first conducted to check for the completeness and readability of the data, which was successful. The second step involved a manual inspection of the documentations, metadata, transcriptions, and lexicon. The transcription resulted in a 99.4% accuracy and contains 108,883 words.

Secondly, the ATCO2 corpus by Zuluaga-Gomez et al. (2022) aimed to encourage research on challenges in the air traffic control field as there is lack of annotated data to provide futher understanding on automatic speech recognition (ASR) and natural language understanding (NLU) of air traffic control communications. The ATCO2 corpus is separated into three sets of corpora called the ATCO2-PL-set corpus, ATCO2-test-set corpus, and a subset called ATCO2-test-set-1h corpus. Data collection depended on volunteers called "data feeders" to annotate ATC voice communications

collected from very-high frequency (VHF) radio receivers. After the data is collected, it undergoes a data processing pipeline that includes various stages of cleaning, filtering, transformation, analysis, and visualization. The first step was “segmentation and demodulation” which used a radio receiver software called RTL-SDR (software defined radio). The second step was “segement-based gain control” which involved using the software to increase the volume when the speakers’ voice is weak from weak signals causing spikes in the waveform. The third step was “signal-to-noise ratio filtering” which used a WADA-SDR (Waveform Amplitude Distribution Analysis) software to estimate the signal-to-noise ratio where non-spech parts are removed. The fourth step was “acoustic-based speaker diarization” which included identifying and separating the speakers’ voice in the audio recording based on acoustic characteristics to ensure the NLU processes from a single speaker. The fifth step was “automatic speech recognition” which required the ASR system to convert the recordings into written text. The sixth step was “English language detection” which employed an ELD (English language detection) system to isolate non-English utterances. The seventh step was “post-processing by NLP” which incorporated three tasks: callsign recognition, ATCOs/pilot classification, and ATC-Entity recognition. The last step was “dataflow statistics” which were the statistics of the data processing pipeline which can be viewed online at https://www.spokendata.com/atco2.

Additionally, there are previous studies that have conducted corpus-based research on aviation. For example, Terenzi (2021) created a corpus about aviation maintenance called the Corpus of Documents Related to Aviation Maintenace (CoDoRAM). The resercher aimed to specify the most important documents related with aviation maintenance for the corpus compilation and mentioned various analysis for solutions in understanding the documents, choosing authentic language for activities, and teaching inside the classroom. The CoDoRAM contains 44,253,149 tokens compiled nineteen relevant aviation maintenance publications. The researcher considered the different factors of aviation English for the selection criteria of the corpus. This included 1) the essential skill for professionals: reading, 2) the essential documents

related to aviation maintenance, 3) text availability, and 4) genre variety. The essential documents were broken down into three functions: Aircraft Maintenance Manuals, Component Maintenance Manual, and Aircraft Operating Manuals. Furthermore, the documents were based on usage and relevance, i.e., availability, source, publication date, and number of tokens and types. The researcher provided additional usage for the CoDoRAM to teach English vocabulary, such as using concordance lines to learn verbs, observing noun strings, and understanding prepositions from real examples.

Although there has been previous research on creating written corpora on aviation, research on creating written corpora based on ICAO publications is still needed to be conducted. Consequently, in this study, the researcher created a written corpus based on ICAO Annexes in order to measure the vocabulary load Thai ATCOs must possess to be able to understand ICAO publications with 95% of the text coverage and to create an English technical word list for Thai ATCOs.

Specialized Corpus

From the four types of corpora, this research focuses on a specialized corpus. A specialized corpus comprises of spoken or written texts that are representative of a language belonging to a specific area (Fuertes-Olivera, 2008). According to Flowerdew (2004), there are various parameters that can define the specialization of a corpus including purpose, genre, size, discourse type, and variety of the language. The purpose infers to how a specialized corpus is compiled to study a particular grammatical, discoursal, or lexical item which using a general corpus may be unsuitable as it contains a broader category of texts (Flowerdew, 2004; Fuertes-Olivera, 2008; Koester, 2010). Specific genres are not included in general corpora and are usually excluded as it is not accessible, such as scientific and academic writing (Koester, 2010). The size of a Language for Specific Purposes (LSP) is intended to be smaller than those made for general language purposes (Bowker & Pearson, 2002). The discourse or text type are where the text is collected from, such as textbooks or conversations (Flowerdew, 2004). The variation of the language can be monolingual or

multilingual, learner, or standard (Bowker & Pearson, 2002; Flowerdew, 2004). To summarize, in this study, the researcher created a specialized corpus to develop an English technical word list for Thai ATCOs.

Corpus-Based Approach

A corpus-based approach is defined as a linguistic research method that focuses on the application of the language, such as the real-life and language teaching function (Faxriddinovna, 2021). The key features of a corpus-based approach include 1) real-world observational usage, 2) extensive representations of selected texts, 3) the application of computers and concordance programs for analysis, 4) both statistical and qualitative text analysis, and 5) focus on practical applications and results. In applied linguistics, corpus-based methods are indispensable for developing educational materials, advancing language acquisition research, and improving translation pedagogy through the utlization parallel corpora.

Criteria for Developing a Corpus

According to Sinclair (2005), criteria for developing a corpus is described as the “first major step” when building a corpus. Common criteria includes: 1) mode of the text (e.g., spoken mode, written mode, electric mode); 2) type of text, where the text can be found (e.g., books, journals); 3) domain of the text, topic or genre of the text (e.g., academic, contemporary); 4) language varieties of the corpus; 5) location of the text (e.g., British English or Australian English); and 6) date of the texts. Corpus creators should choose simple criteria to avoid difficulty when employing them as errors in classification can lead to invalid results.

Word Lists and Vocabulary Teaching

Vocabulary is fundamental to building English language proficiency, particularly in the area of reading. When learners grow their vocabulary knowledge, they are able to improve their understanding capacity and interpretation of the texts with greater ease. For instance, the General Service List (GSL) provide learners with approximately 80%

coverage of words appeared in general written materials, thus supporting their comprehension and reading fluency (Kwary & Jurianto, 2017; Nesi, 2002).

A range of techniques can support the growth of learners' vocabulary, including strategies for extensive reading and employing technology to explore language patterns (Liu & Zhang, 2018; Siddiq et al., 2021). Another useful method is developing word lists, which are effective for systematically expanding vocabulary. These word lists aim to pinpoint keywords that learners should be familiar with. The goal of word lists in language learning is to highlight frequently used or specialized words in that field which can align with the learners' needs (Kwary & Jurianto, 2017).

Word lists are essential in English language teaching for several reasons, including designing course, creating teaching materials, developing tests, and buidling specialized dictionaries. Nation (2016) points out that educators can rely on word lists to structure teaching programs and assessments related to vocabulary. Gardner and Davies (2003) stated that word lists can be applied in defining objectives for vocabulary acquisition, evaluating vocabulary comprehension and development, examining the complexity of the text, developing, or adapting materials for reading, devising vocabulary resources, and identifying vocabulary elements within an academic syllabus. These word lists can assist instructors shape course content and identify key terms that need to be taught. For instance, they can help prioritize high-frequency or field-specific words in order for learners to focus on language that follows their needs. Instructional materials such as textbooks, reading seletions, and vocabulary activities can be developed based on these word lists to match the learners' language levels. Assessments like tests and quizzes can be designed to assess the learners' mastery of important and technical terms. Moreover, specialized dictionaries can be tailored to specific groups of learners to highlight crucial words and can include usage examples to support deeper comprehension and encourage effective vocabualry usage (Nesi, 2002).

Taking these factors into account, it is clear that more research is required to create a specialized word list that is suitable for ATCOs in training. To support training

ATCOs, a specialized word list would form a guideline for the training curriculum and gain a strong grasp of the essential technical terms for precise and standardized communication. This approach highlights the crucial role that word lists play in effective vocabulary learning, especially in specialized areas like air traffic control where it requires accuracy.

General Service List (GSL) and Academic Word List (AWL)

The most well-known word lists are the General Service List (GSL) by West (1953) and the Academic Word List (AWL) by Coxhead (2000).

General Service List (GSL)

General Service List (GSL) was formed by a five-million-word corpus resulting in two thousand of the most frequent words that covers 78 – 81% of written texts and 85% of spoken texts. The selection criteria for the word list West (1953) used were both qualitative and quantitative and consisted of four characteristics: frequency, universality, utility, and usefulness (Palinkasevic, 2017). The qualitative criteria were 1) ease of learning (similar word forms), 2) necessity and cover (covers important ideas/concepts), and 3) stylistic and emotional neutrality (neutral ideas). However, the GSL is not enough to understand average texts as Nation (2006) has explained that 98% of the text should be understood. This has resulted in the formation of the New General Service List (NGSL) by Brezina and Gablasova resulting in a 2,122-word list.

Academic Word List (AWL)

Academic Word List (AWL) comprises of 570 word families which targets university students who need to use English with their studies to provide better understanding of their learning materials and a manageable vocabulary load. It was developed from a 3.5-million-word corpus and covered four disciplines: arts, commerce, law, and science. The selection criteria for the word list were 1) exclusion of GSL words, 2) occurrence frequency of one hundred times or more, 3) ranged from 15 or more subject areas, and 4) uniformity in occurrence of 10 times or more in the four disciplines.

In this study, GSL and AWL were employed in the third criterion for Lexical Profiling as referent word lists using the AntWordProfiler program to omit the words. The words that do not appear in these referent word lists would pass the third criterion.

Criteria for Developing Word Lists

Scholars use various criteria to compile word lists methodically to follow the needs of the different fields. For example, Kim and Lee (2019) developed the Linguistic Academic Vocbaulary List (LAVL) to assist undergraduate EFL students to build the lexical knowledge for academic reading and writing. The researchers assembled the Linguistics Textbooks Corpus (LTC) and applied three criteria from Hsu (2014) including specialized occurrence, range, and frequency. The specialized occurrence criterion focused on the word families that appeared beyond the first 2,000 word families. The range criterion required a word to appear in at least three out of the five textbooks selected. The frequency criterion meant that the word must show up at least 17 times throughout the textbooks in the LTC.

Moreover, to construct a word list, Rungrueang et al. (2022) mentioned that there are five criteria used including 1) frequency, 2) range, 3) lexical profiling, 4) keyword analysis, and 5) expert viewpoints based on multiple scholars including Coxhead (2000), Lei and Liu (2016), Laosrirattanachai and Ruangjaroon (2020), Muñoz (2015), Tangpijaikul (2014), Watson Todd (2017), West (1953), and Yang (2015).

1. Frequency

Based on the AWL by Coxhead (2000), the frequency threshold for each word in the list was set to a minimum 100 times within the 3.5 million running word corpus. Therefore, the equation used for calculating the frequency is presented below.

$$x = 100 \times \frac{corpus\ tokens}{3{,}500{,}000}$$

According to the equation above, x is frequency. Words that amount to the frequency will pass this criterion, which is calculated by using the AntWordProfiler

program. However, this criterion alone does not suffice to generate an authentic word list as it may be influenced by longer text segments (Laosrirattanachai & Ruangjaroon, 2021; Rungrueang et al., 2022).

2. Range

The range is used to decrease the size and eliminate bias words. According to Rungrueang et al. (2022), it is important to examine how many sources a word appears. The wider the words appear in a variety of sources, the wider the range will be. A wide range means a word that appears in 50% of the sources. Words with a wide range will be evaluated for inclusion into the word list. Some words may appear often in one source and not in other sources (Nation, 2016, as cited in Rungrueang, 2022). Therefore, these words are considered as a narrow range meaning it does not pass the criterion.

3. Lexical profiling

Lexical profiling is utilized to remove words that are deemed irrelevant and already present in referent word lists, such as GSL, AWL, and other word lists. Consequently, the words not found in these referent lists meet this criterion.

4. Keyword analysis

Keyword analysis is used to recall words that have emerged in the referent word list, but have particular meaning related to the word list being generated. It can be calculated by following Watson Todd's (2017) research. In Watson Todd's study, a word list was created by comparing the meanings of words in context with their definitions in online dictionaries. The keyword analysis utilized a log-likehood statistic to identify words significantly more frequent in the target corpus than in the benchmark corpus. The log-likehood statistic was calculated by using the Key-BNC program. The results from the program displayed the highest to lowest log-likelihood (LL) values. In his study, he focused only on the 500 words with the highest LL values derived from a corpus size of 1,150,000 tokens. Therefore, the minimum frequency ratio of his study was the first 500 words.

In this study, to establish the minimum frequency ratio, the equation used to calculate is provided below.

$$x = \frac{500 \times corpus\ tokens}{1{,}150{,}000}$$

According to the equation above, x refers to the minimum frequency ratio. The words calculated using this equation will pass the criterion, which is calculated by using the Key-BNC program.

5. Expert viewpoints

Expert viewpoints involve the help of experts in a specific field to provide opinions and feedback on which words relevant in the word list using a four rating scale. These words can be rated using a scale from 1 to 4 proposed by Chung and Nation (2004) as shown in Table 1.

Table 1 Expert Viewpoints Four Rating Scale

Description	Score
Words unrelated to the field	1
Words pertaining minimal relevance to the field	2
Words pertaining high relevance to the field	3
Words completely relevant to the field	4

Unlike the criteria used by Kim and Lee (2019), the criteria outlined by Rungrueang et al. (2022) incorporates three more important criteria, including lexical profiling, keyword analysis, and expert viewpoints. These additional criteria help ensure that the word list is more technical and specific to the field. As a result, the researcher adopted Rungrueang et al.'s (2022) criteria to develop a specialized vocabulary list for Thai ATCOs.

Related Studies

Multiple researchers have created word lists within various industries, such as food service, computer science, and pharmacy. The first research was conducted by Chen and Lei (2019) where the researchers endeavored to discern the techincal words employed in research studies regarding computer science throughout ten sub-disciplines between 2010-2014 and distinguish the distribution of technical words in terms of range and frequency. The Corpus of Research Articles in Computer Science (CRACS) contained 10,450,035 running words, which followed Sinclair's (2005) criteria on 1) representatives, 2) corpus specificity, 3) whole document usage, and 4) electronic availability. The researchers employed the Range program to analyze the coverage of four lists: first thousand GSL words, second thousand GSL words, AWL, and Not found in any list. This resulted in 1,262 word-types following the selection criteria: range, frequency, and type. Additionally, the researchers chose to count word types instead of word families for not all words in a family are counted as technical words. Subsequently, the researchers autonomously evaluated each word to ascertain the high frequency and computer science specific technical words from GSL, New General Service List (New GSL), AWL, and Academic Vocabulary List (AVL). The Technical Word List for Research Articles in Computer Science (TWLRACS) is assembled of 769 word-types. As for the coverage of the technical words, more than half appeared in eight or more sub-discipline areas and ranged between 100-299 times. Due to this, the technical words in the TWLRACS are not listed alphabetically, but by range and frequency.

The second research was conducted by Heidari et al. (2020) where the researchers constructed the Pharmacy Academic Word List (PAWL) from a corpus of over 3 million tokens collected from research articles related to pharmacy across four subfields that were published in 2016 and 2017. After data collection, the researchers standardized the texts which eliminated everything unrelated to the text, such as references and charts. WordSmith, AntWordProfiler, and Familizer program were applied to arrange the word families according to the four selected criteria: 1) specialized occurrence outside of the first 2,000 words, 2) range at least in 3 of the 4

subfields, 3) dispersion of a minimum of 40% index in Wordsmith, and 4) frequency of at least 100 times in the corpus. The PAWL resulted in 750 word families after being reviewed by expert pharmacy researchers. Furthermore, the PAWL was compared to the AWL, Medical Academic Word List and BNC using Coverage Calculator to ensure domain-specificity.

The third research was conducted by Rungrueang et al. (2022) where the researchers aimed to create the Food Service Word List (FSWL) and discover the proportion of the Food Service Corpus (FSC) coverage in FSWL as there is a lack of technical word list in the food serve industry. The FSC contained 1,871,271 running words that were compiled between July-August 2021 from four relevant websites: BBC Food, TasteAtlas, Escape.com.au, and CNN Food & Drink. The researchers employed 3 main criteria to develop the FSWL: 1) keyword analysis, 2) lexical profiling, and 3) expert viewpoints. The study did not rely on frequency and range as there may be related words that do not appear in many sources, but is significant in the food service industry. First, the keyword analysis adopted the cut-off point from Watson-Todd (2017) where the top 500 words with the highest log-likelihood value (LL) from a 1,150,000 token corpus was analyzed. This resulted in the first 814 words passing the first criterion. Second, lexical profiling using AntWordProfiler program was distributed into six referent word lists: 1) first thousand words from GSL (102 words), 2) second thousand words from GSL (180 words), 3) AWL (14 words), 4) FWL (15 words), 5) AL (1 word), and 6) PNL (8 words). The program eliminated 290 words distributed in the six referent word lists, resulting in 524 words for consideration based on experts' viewpoints to be included into the FSWL. Third, expert viewpoints were from three individuals who have each worked for a span of four years while serving as an assistant manager and utilized English in communicating with foreign customers. The FSWL resulted in 261 words rated 3 or 4 in the questionnaire. When compared to other related word lists in the food service industry, the corpus coverage reached 8.64%, exceeding the recommended 5% coverage advised for a technical word list. This demonstrates the effectiveness of FSWL and assists food service industry learners.

In Thailand, there are two research studies that created a word list about aviation. Firstly, Laosrirattanachai and Ruangjaroon (2021) created three specialized word lists: the Tourism Business Word List (TBWL), Hotel Business Word List (HBWL), and Airline Business Word List (ABWL). The ABWL, which focuses specifically on aviation, was created to address a major challenge for ESP learners where there was a lack of technical vocabulary in foreign settings. The word lists derived from an overall corpus of 52,079,960 running words, consisting of the Tourism Business Corpus (TBC), Hotel Business Corpus (HBC), and Airline Business Corpus (ABC). The selection criteria for the word list called Six Filters (6Fs) was meticulously chosen as it is arranged systematically. The first filter resulted in words appearing the ABC at least 444 times respectively. For the second filter, the researchers used the program AntWordProfiler to extract words that appeared in at least fifty percent of the total sources. For the third filter, irrelevant words that have appeared in the GWL, AWL, Function Word List (FWL), Abbreviation List (AL), and the Proper Name List (PNL) were eliminated by using AntWordProfiler. For the fourth filter, the researchers compared the ABC to a reference corpus, the British National Corpus (BNC), based on the highest log-likelihood in the Key-BNC program resulting in 6,758 words. For the fifth filter, five experts working for more than five years each in the airline business rated whether the words had relevant meaning or not. The results from the fifth filter constituted 245 words in the ABWL. For the last filter, the VocabProfile program was used to divide the three main word lists into sub-word lists according to the words' difficulty. The remaining words were allocated into the sub-word lists of each field with 30 words each, which resulted in the ABWL comprised of eight sub-word lists.

Secondly, a research study conducted by Thiankasem (2018) aimed to measure the vocabulary load cabin crew need in order to understand 95% of the text coverage in Cabin Crew Manuals used in Thailand and investigate the technical words that occur outside of GSL and AWL. The Cabin Crew Manual (CCM) Corpus was developed by gathering a total of 7 cabin crew manuals by using the AntWordProfiler and AntFileConverter. This resulted in 1,053,941 running words and were separated into

three sub-corpora: 1) Airbus sub-corpus containing 461,891 running tokens, 2) Boeing sub-corpus containing 488,736 running tokens, and 3) CCM sub-corpus containing 103,314 running tokens. The selection criteria used to develop the word list included: 1) range as the technical words had to occur at least two out of three sub-corporas, 2) categorization of word families, and 3) three subject matter experts. The study found that in order for cabin crew members to understand 96.73% of the text coverage, they should know 4,629 word families including proper nouns, marginal words, transparent compounds and acronyms. As for the techinical words, the researcher developed the CREW-related English Word (CREW) List with 590 word families that was based on two criterias: specialized occurrence and range.

There have been several previous research studies that developed a word list for aircraft maintenance, airline operations, and cabin crew communication. Yet, there has been little focus towards techincal words required by ATCOs. Although some studies have examined ATCO communication through standard phraselogy, it still leaves a significant gap as it is essential for ATCOs to have clear and accurate understanding. To bridge this gap, the current research study aims to create a specialized word list for the Air Traffic Management (ATM) sector. This research study seeks to provide Thai ATCOs the linguistic tool needed to increase their understanding by employing a corpus-based analysis. Ultimately, the results are expected to support safer and more efficient air traffic control operations. Chapter 3 will provides an comprehensive overview and analysis of the research methodology.

CHAPTER 3
METHODOLOGY

This chapter demonstrates the data selection, instruments, data collection, and data processing. The researcher employed the criteria for developing a corpus by Sinclair (2005) for data selection and data collection. Afterward, the criteria for developing a word list suggested by Rungrueang et al. (2022) were employed for data processing.

Data Selection

ICAO publications are specialized documents that are apportioned among stakeholders including Contracting States, government agencies, airlines, industry manufacturers, and aviation simulation training centers. In Thailand, AEROTHAI is assigned by the Thai government responsible for air traffic services (ATS) within Thailand, which include ATCOs (AEROTHAI, 2023). ATCOs are required to consult ICAO publications to guarantee uniform compliance, operational procedures knowledge, and uniformity of SARPs as they are a part of ICAO's Contracting States. Annexes are official documents by ICAO that expands on the Standards and Recommended Practices (SARPs) which Contracting States of ICAO are expected to adhere to ensure consistent regulations, standards, and procedures (Dib, 2022; ICAO, 2006). SARPs are Standards and Recommended Practices that Contracting States, such as Thailand, are required to follow any specifications for the uniform administration for the safety or regularity of international air navigation (ICAO, 2006).

The data in this research study were selected from the ICAO Products & Services Catalogue 2023 under the section of Annexes to the Convention on International Civil Aviation. 19 ICAO Annexes were chosen for their crucial role in maintaining global aviation safety and efficiency. Since the Annexes incorporate SARPS, Contracting States must adopt them under the Chicago Convention (1944) to promote uniform aviation standards. These Annexes cover key topics such as aviation safety, airworthiness, air traffic management, and other elements related to air transport.

Contrastingly, the PANS and Technical Manuals are the two other ICAO publications that serve as supplementary resources where they offer guidance rather than mandatory obligations. ICAO Annexes serve as the essential foundation for global aviation safety, regulatory consistency, and standardized implementation, while PANS and Technical Manuals are important for operational effectiveness.

The ICAO Annexes comprise a total of 30 ICAO publications. These documents were sourced online from the Federal Office of Civil Aviation (FOCA) website (https://www.bazl.admin.ch) and the International Aviation Organization website (https://www.icao.int). Some Annexes contained only one volume, while others contained more than one volume, as detailed in Table 2. For example, Annex 6 is is organized into three separate volumes that addresses specfic operations of aircraft, such as International Commercial Air Transport – Aeroplanes, International General Aviation – Aeroplanes, and International Operations – Helicopters. Thai ATCOs need to be familiar with these volumes to maintain safe and internationally compliant operations. Annex 6, for instance, is crucial for understanding aircraft operations which ATCOs can use it to maange efficient flight operations for smooth flow of traffic considering factors like aircraft performance, wake turbulence, and fuel efficiency. Overall, a thorough understanding of these Annexes will enable Thai ATCOs to align with global aviation standards issued by ICAO. For this reason, the researcher chose to investigate all 19 Annexes in this research study.

Table 2 ICAO Annexes and Volumes

Annex	Title	Volume
1	Personnel Licensing	-
2	Rules of the Air	-
3	Meteorological Services	-
4	Aeronautical Charts	-
5	Units of Measurement	-
6	Operation of Aircraft	I International Commercial Air Transport - Aeroplanes II International General Aviation - Aeroplanes III International Operations - Helicopters
7	Aircraft Nationality and Registration Marks	-
8	Airworthiness of Aircraft	-
9	Facilitation	-
10	Aeronautical Telecommunications	I Radio Navigation Aids II Communication Procedures including those with PANS status III Communication Systems IV Surveillance and Collision Avoidance Systems V Aeronautical Radio Frequency Spectrum Utilization VI Communication Systems and Procedures Relating to Remotely
11	Air Traffic Services	-
12	Search and Rescue	-
13	Aircraft Accident and Incident Investigation	-
14	Aerodromes	I Aerodrome Design and Operations II Heliports
15	Aeronautical Information Services	
16	Environmental Protection	I Aircraft Noise II Aircraft Engine Emissions III Aeroplane CO_2 Emissions IV Carbon Offsetting and Reduction Scheme for International Aviation (CORSIA)
17	Security	-
18	The Safe Transportation of Dangerous Goods by Air	-
19	Safety Management	-

Instruments

The research instruments of this study consist of AntFileConverter version 2.0.2 (Anthony, 2022), AntConc version 4.2.0 (Anthony, 2023a), AntWordProfiler (Anthony, 2023b), the Key-BNC program, and BNC/COCA word family lists.

1. AntFileConverter

AntFileConverter version 2.0.2 was downloaded online from Anthony's website (https://www.laurenceanthony.net/software/antfileconverter/) and is used to convert electronic files of the ICAO publication from pdf files to plain text or ".txt" files. The program includes the input and output files function for the user to upload the pdf files in input and receive the plain text as output, as shown in Figure 6.

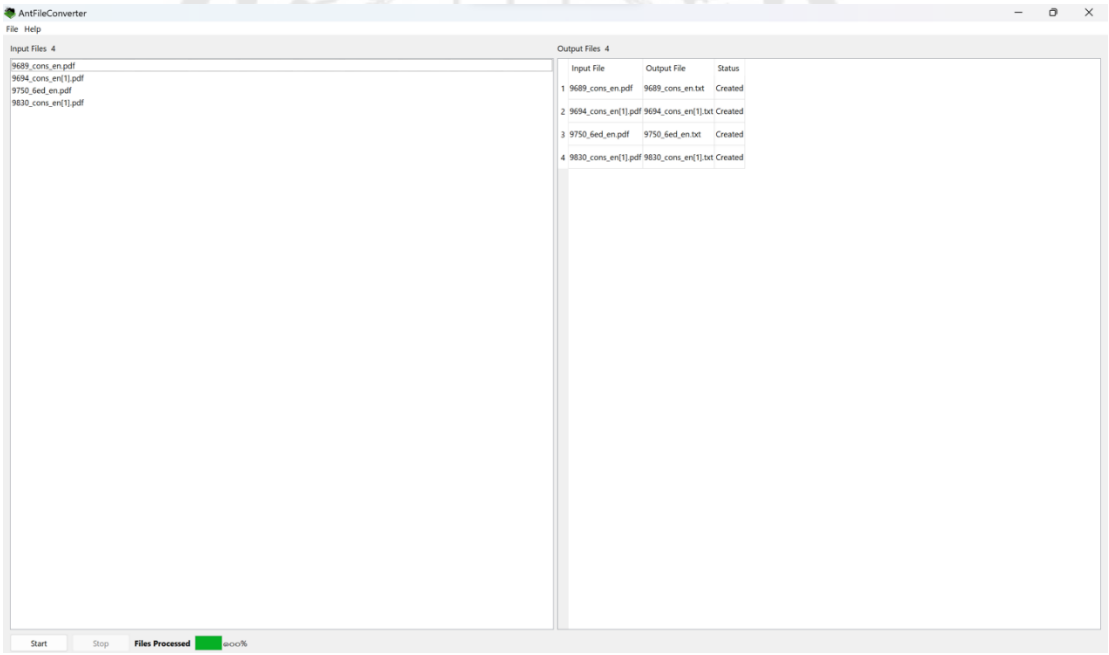


Figure 6 AntFileConverter (Version 2.0.2) Computer Software (Anthony, 2022)

Source: Laurence Anthony. (2022).

https://www.laurenceanthony.net/software/antfileconverter/

2. AntConc

AntConc version 4.2.0 was downloaded online from Anthony's website (https://www.laurenceanthony.net/software/antconc/) and is used for analyzing the frequency of air traffic management words found in ICAO publications. The program contains functions to compute the amount of word types, word tokens, and search hits as shown in Figure 7.

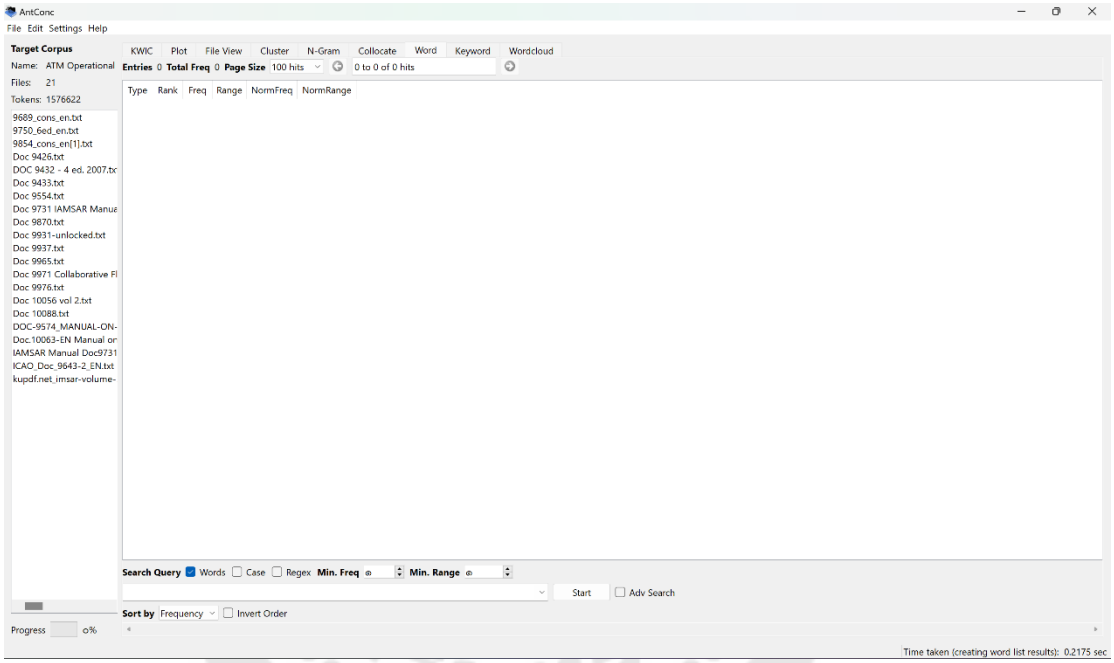


Figure 7 AntConc (Windows Version 4.2.0) Computer Software (Anthony, 2023)

Source: Laurence Anthony. (2023). https://www.laurenceanthony.net/software/antconc/

3. Range

The vocabulary load was analyzed using the Range program to answer the first research question. The ATM corpus contained 1,245,440 tokens and to be able to analyze the vocabulary load, it needed to be compared with the BNC/COCA word family lists in the AntWordProfiler program. The researcher first downloaded the BNC/COCA word family lists from Paul Nation's resources (https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-

programs) that contain the Range program and 25 word family lists and four additional lists named “basewrd 1 - 34”. The lists “basewrd 1 - 25” contain the most common 25,000 words, each level comprises of 1,000 words. The lists “basewrd 26 - 30” include nonsense words. The lists “basewrd 31 - 34” incorporate the four additional word lists, including proper nouns (e.g., names and countries), marginal words (e.g., alphabet letters and swear words), transparent compounds (e.g., compound words that can be easily understood from the meaning of each part), and acronyms (e.g., ICAO and ESP) respectively. As the lists “basewrd 1 -34” cannot be directly imported into the AntWordProfiler program, they needed to be made into “headwords 1- 34”.

In order to measure the vocabulary load, only “headwords 1 - 25” and “headword 31 – 34” from the BNC/COCA word family lists were selected and imported into the AntWordProfiler program as reference lists. As a result, the Range program was used in this study to compare the ATM corpus with the BNC/COCA word family lists.

4. BNC/COCA Word Family Lists

Developed by Paul Nation, the BNC/COCA word family lists consitute of a comprehensive compilation of words organized into categories based on their occurrence frequency of 1,000-word bands. The list includes 25,000 of the most prevalent words in English. The BNC/COCA lists are crafted to support individuals learning English as a second language that prioritize vocabulary pertinent to international travel, academic pursuits in English-speaking contexts, and digital communication within the initial 2,000 words. Moreover, these lists have been incorporated in the creation of graded readers and textbooks that facilitate in language acquisition through controlled vocabulary exposure (Smith, 2023).

5. Key-BNC Program

The Key-BNC program is a tool that conducts comparative keyword analysis by calculating statistics against the BNC word list for linguists who lack direct access to the BNC. It employs the log-likelihood (LL) metric, a statistical measure of probability that assesses the frequency of word occurrence across two distinct corpora. A high LL value

indicates a notable difference in the relative frequencies of a word according to the sizes of the two corpora. Consequently, when the relative proportions of words align, those with higher absolute frequency tend to exhibit higher LL values which are typically common words. This clarifies why keyword lists ordered by LL tend to emphasize more frequently used words (Watson Todd & Pojanapunya, n.d.).

6. AntWordProfiler

AntWordProfiler is a contemporary program to analyze vocabulary with a variety of additional features. The program is available on Anthony's website (https://laurenceanthony.net/software/antwordprofiler/) which is used for profiling the level of vocabulary and the complexity of the texts as shown in Figure 8.

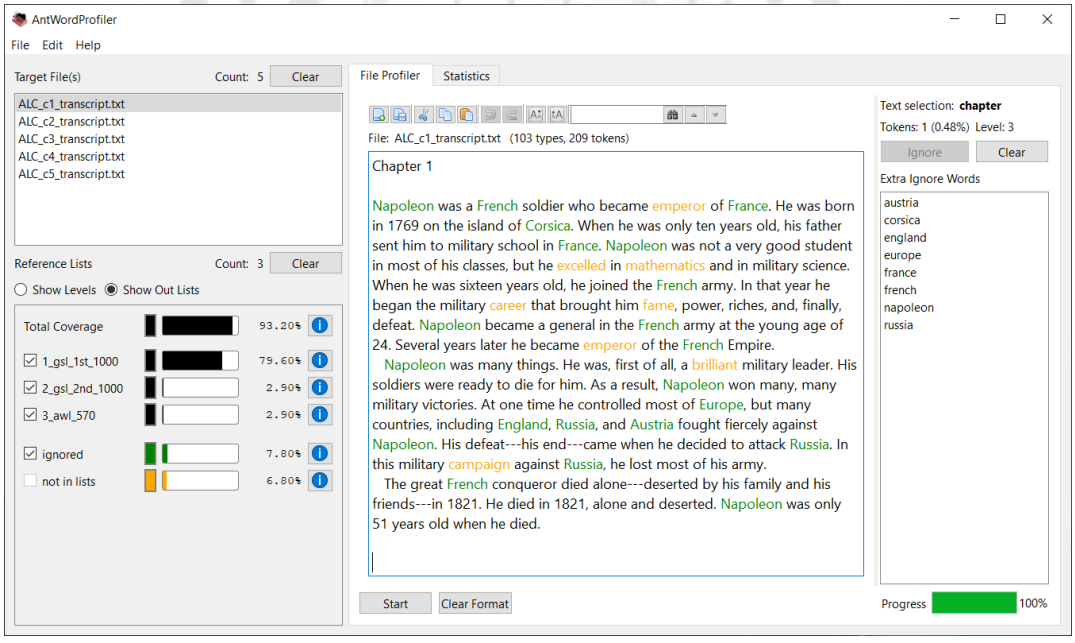


Figure 8 AntWordProfiler Profile Window (Anthony, 2023)

Source: Laurence Anthony. (2023). www.laurenceanthony.net

Data Collection

The thirty ICAO publications were selected from the following two websites:

1. The Federal Office of Civil Aviation (FOCA) (https://www.bazl.admin.ch) is a department under the Federal Department of the Environment, Transport, Energy, and Communications (DETEC) in Switzerland that focuses on the civil aviation activities and sustainable development in the country.
2. International Civil Aviation Organization (https://www.icao.int) is the official website for ICAO, a United Nations specialized agency established in 1944 to unite aviation.

The publications were downloaded online from both websites as pdf files and converted them into .txt files using the AntFileConverter program into a single folder. The folder was then imported into the AntConc program to build a corpus. The resulting Air Traffic Management corpus contained a total of 1,245,440 tokens and 13,681 word types.

Table 3 Online Sources for Downloading ICAO Publications

No.	Source	ICAO Publications	No. of ICAO Publications
1	Federal Office of Civil Aviation (FOCA) (https://www.bazl.admin.ch)	Annex 3, Annex 4, Annex 5, Annex 6 Vol. I, Annex 6 Vol. II, Annex 6 Vol. III, Annex 7, Annex 8, Annex 9, Annex 10 Vol. I, Annex 10 Vol. II, Annex 10 Vol. III, Annex 10 Vol. IV, Annex 10 Vol. V, Annex 13, Annex 14 Vol. I, Annex 14 Vol II., Annex 15, Annex 16 Vol. I, Annex 16 Vol. II, Annex 16 Vol. III, Annex 16 Vol. IV, Annex 17, Annex 18, Annex 19	25
2	International Civil Aviation Organization (https://www.icao.int)	Annex 1, Annex 2, Annex 10 Vol. VI, Annex 11, Annex 12	5
	Total		30

The thirty ICAO publications were accumulated online from two websites and downloaded the most recent editions. In Table 2, the researcher searched for each publication in a pdf form and downloaded them from the two websites. Then the pdf files of each publication were uploaded into AntFileConverter to become .txt files.

Data Processing

According to the research objectives, the researcher would like to measure the vocabulary load ATCOs need in order to understand 95% of text coverage of ICAO publications related to Air Traffic Management and create an Air Traffic Management word list.

1. **Measuring the Vocabulary Load**

To measure the vocabulary load, AntWordProfiler, a program devised by Laurence Anthony (2023b) specializes in examining the vocabulary and the creation of lexical profiles. The initial step involves inputting the designated text for analysis. The program generates a detailed report encompassing information on the occurrence of each word, its grammatical classification, and additional linguistic attributes. This process identifies prevalent words as well as less common or specialized vocabulary terms. The ATM Corpus will be compared with the BNC/COCA word family lists to show the percentage of the text coverage.

2. **Creating a Word List**

The AntWordProfiler software was used to develop the word list following the five key criteria outlined by Rungrueang et al. (2022) including frequency, range, lexical profiling, keyword analysis, and expert viewpoints. Firstly, frerquency ensures the words that are commonly used are prioritized for supporting better comprehension and language fluency. Secondly, range helps prevent bias towards word selection as it requires words to appear across multiple sources to improve their representativeness and practicla relevance. Thirdly, lexical profiling filters our unnecessary words already included in word lists that are already established such as General Service List (GSL)

(West, 1953), the Academic Word List (AWL) (Coxhead, 2000), and the Function Word List (FWL). Fourthly, keyword analysis identifies domain-specific words related to air traffic management by isolating relevant vocabulary from the GSL and AWL. Finally, expert viewpoints play a crucial role as professionals in the related field can assess the practical relevane of each word based on their experience to ensure the word list actually reflects the language used during operations.

1. Frequency: The researcher followed Coxhead's (2000) guideline on the development of the Academic Word List (AWL) that a word must appear at least 100 times in a 3.5 million word corpus. For this study, the researcher used the equation as presented below.

$$x = 100 \times \frac{corpus\ tokens}{3{,}500{,}000}$$

$$x = \frac{100 \times 1{,}245{,}440}{3{,}500{,}000}$$

$$x = 35.584$$

A word needs to occur at least 36 times in order to pass this criterion. The AntWordProfiler program was used to identify the frequency of the words which will then be assessed under the next criterion.

2. Range: The researcher aimed to reduce the list size and remove biased words by checking how broad a word appeared across different sources. As suggested by Rungrueang et al. (2022), a word needs to appear in at least 50% of the sources to qualify. As a result, the word must appear in a minimum of 15 different sources. Importantly, these words have to also meet the frequency criterion of 36 occurences. The words that pass both the frequency and range criteria would considered for further selection.
3. Lexical profiling: The researcher utilized the AntWordProfiler program to exclude unrelated words that appear in reference word lists including

GSL, AWL, and FWL. The words that appear in the reference word lists would be excluded, while the words that did not appear in the reference word lists will pass this criterion and kept for further analysis.

4. Keyword analysis: The researcher recalled the words that have appeared in the reference word lists. Keyword analysis uses a minimum frequency threshold to identify words significantly more frequent in the target corpus than in the benchmark corpus. The log-likehood statistic was calculated by using the Key-BNC program. The equation used to establish the minimum frequency ratio was adapted from Watson Todd's (2017) research study.

$$x = \frac{500 \times corpus\ tokens}{1{,}150{,}000}$$

$$x = \frac{500 \times 1{,}245{,}440}{1{,}150{,}000}$$

$$x = 541$$

The first 541 words based on the calculation would pass the criterion using the Key-BNC program.

5. Expert viewpoints: The researcher enlisted three experts in the field of air traffic management to provide opinions and feedback on the significance of the words in the word list. They provided opinions and feedback on which words relevant in the word list using a four rating scale proposed by Chung and Nation (2004). The chosen experts have been working in the aviation industry for more than 5 years and are licensed ground instructors for air traffic services. Two of them work as air traffic management instructors and another one is an air traffic controller.

CHAPTER 4
FINDINGS

This chapter displays the findings of the vocabulary load Thai ATCOs must possess to be able to understand ICAO publications with 95% of the text coverage and the technical word list created for Thai ATCOs. The results were separated into two sections according to the research questions. The first section displays the vocabulary load results. The second section displays the techincal word list created using a corpus-based approach.

Measuring the Vocabulary Load

In order to analyze the vocabulary load Thai ATCOs must possess to be able to understand ICAO publications with 95% of the text coverage, the ATM corpus which contains 1,245,440 tokens was compared with the BNC/COCA word family lists in the AntWordProfiler program. The results showed that the words in the corpus that appeared in the BNC/COCA word family lists resulted in 98.97%. However, the words that did not appear in the lists were 1.03%, also referred to words out of list.

Table 4 shows that the first 1,000-word level resulted in 65.11% and the 2,000-word level (1,001 - 2,000 words) resulted in 7.07%. This ultimately resulted in an accumulated percentage of 72.18%. However, if the readers know only the most frequent 2,000 word families, it was still not enough to understand the ICAO Annexes as the readers need to know 95% of the text coverage in order to understand the publications.

Next, when considering the most frequent 3,000 word families, it resulted in an accumulated percentage of 78.90%, which was still not enough to understand the ICAO Annexes. Then after considering the most frequent 4,000 word families, it resulted in 80.38%. However, it was still not enough to understand the Annexes.

Therefore, to achieve 95%, the readers must know the most frequent 4,000 word families and other words, including proper nouns, marginal words such as alphabet

letters, swear words, and exclamations, transparent compounds, and acronyms. This resulted in an accumulated percentage of 95.68%.

Additionally, Table 4 illustrates the top three highest percentage of words found in the BNC/COCA word family list which were the 1,000-word level, 2,000-word level, and 3,000-word level. The 1st 1,000-word level contained 810,964 tokens (65.11%). The 2nd 1,000-word level contained 88,005 tokens (7.07%). The 3rd 1,000-word level contained 83,693 tokens (6.72%). Interestingly, the tokens in the word list show a pattern of decreasing tokens as the level increases. However, it can be seen in the 9th 1,000-word level that the number of tokens increases (2,324 tokens, 0.19%), as well as the 13th 1,000-word level that the number of tokens increases (2,512 tokens, 0.20%).

In comparison of the four additional word lists (proper nouns, marginal words, transparent compounds, and acronyms) in the BNC/COCA word family lists, the proper nouns were found the most (144,676 tokens, 11.62%). It might be because the ICAO Annexes often mention languages, countries, and cities. For example, Annex 1 - Personnel Licensing found the word “English” used often as this specific Annex describes the standard procedures to receive license and the ICAO language proficiency test, which pilots and ATCOs must take in order to be qualified to work. Another example can be found in Annex 9 - Facilitation, where most words found were related to cities such as “Chicago” and “Manila”. This refers to the Chicago Convention 1944 and contains the amendments of the Annex and where the meeting was held.

Table 4 Tokens at Each Word Level for the ICAO Annexes

Word List	Tokens	%	Cumulative Coverage (%)
Proper Nouns	144,676	11.62	11.62
Marginal Words	26,259	2.11	13.73
Transparent Compounds	4,091	0.33	14.06
Acronyms	15,446	1.24	15.30
1st 1,000	810,964	65.11	80.41
2nd 1,000	88,005	7.07	87.48
3rd 1,000	83,693	6.72	94.20
4th 1,000	18,463	1.48	95.68
5th 1,000	12,852	1.03	96.71
6th 1,000	9,428	0.76	97.47
7th 1,000	4,259	0.34	97.81
8th 1,000	1,905	0.15	97.96
9th 1,000	2,324	0.19	98.15
10th 1,000	1,127	0.09	98.24
11th 1,000	1,164	0.09	98.33
12th 1,000	249	0.02	98.35
13th 1,000	2,512	0.20	98.55
14th 1,000	2,192	0.18	98.73
15th 1,000	980	0.08	98.81
16th 1,000	282	0.02	98.83
17th 1,000	181	0.01	98.84
18th 1,000	140	0.01	98.85
19th 1,000	536	0.04	98.89
20th 1,000	69	0.01	98.90
21st 1,000	375	0.03	98.93
22nd 1,000	219	0.02	98.95
23rd 1,000	63	0.01	98.96
24th 1,000	72	0.01	98.97
25th 1,000	62	0.00	98.97
Not in the list	12,852	1.03	100
Total	1,245,440	100	100

To answer the first research question in order to understand the ICAO Annexes, Thai ATCOs must reach a lexical coverage of 95%. The 1st 1,000-word level combined with the proper nouns, marginal words, transparent compounds, and acronyms has a cumulative percentage of 80.41% which was not enough to understand the ICAO Annexes. When considering the most frequent 2,000 word families and the four additional lists, the cumulative percentage was 87.48% which was not enough to understand the text. Also, when considering the most frequent 3,000 word families, the cumulative percentage was 94.20% which was not enough to comprehend the written text. Therefore, to have a vocabulary load of 95%, the readers must know the most frequent 4,000 word families including the four additional lists. A cumulative percentage of the first 4,000 words including the four additional lists is 95.68%.

Creating the Air Traffic Management (ATM) Word List

1. Frequency

The frequency analysis resulted in 2,811 words which passed the criterion of occurring at least 36 times by using the AntWordProfiler program. Within the first 50 highest-frequency words, the words that occurred the most were a combination of function words and content words.

Below are examples of function words in the first 15 highest-frequency words:

the	*of*	*and*	*to*	*by*
at	*are*	*which*	*from*	*this*
when	*all*	*may*	*any*	*than*

Below are examples of content words in the first 10 highest-frequency words:

aircraft	*flight*	*annex*	*note*	*data*
runway	*information*	*state*	*chapter*	*air*

The word “aircraft” was one of the first 50 highest-frequency words as it was the most common content word that appeared 6,790 times in the corpus. Below is an example of this word with context in ICAO publications.

Example: Aircraft - any vehicle, with or without an engine, that can fly, such as a plane or helicopter (Cambridge, n.d.)

Annex 1: Similarly the expression “flight crew member” has the same meaning as the expressions “member of the operating crew of an “aircraft” and “operating personnel” used in the Convention while the expression “personnel other than flight crew members” includes the expression “mechanical personnel” used in the Convention.

Annex 5: The establishment and withdrawal of and changes to facilities, services and procedures affecting aircraft operations provided in accordance with the Standards and Recommended Practices specified in this Annex should be notified and take effect in accordance with the provisions of Annex 15.

Annex 18: Articles and substances which would otherwise be classed as dangerous goods but which are required to be aboard the aircraft in accordance with the pertinent airworthiness requirements and operating regulations, or for those specialized purposes identified in the Technical Instructions, shall be excepted from the provisions of this Annex.

However, the frequency criterion is not enough as words with high frequency need to be filtered to be able to find words that appear in all the ICAO publications.

2. Range

Coxhead (2000), referenced in Laosrirattanachai and Ruangjaroon (2021), stated that the words that occurred in at least 50% of the sources will pass the range criterion. Thus, the words that passed the range criterion must occur in at least 15 publications from the overall 30 publications. A total of 247 words met the criteria after considering both frequency and range. The words both included function words, such as “the”, “of”, “and”, and “to”, and content words.

The content words that passed this criterion are the following examples. Three content words that occurred in 15 sources were “message”, “lights”, and “elevation”, which were found in 50% of the sources. Two content words that occurred in 16 sources were “fuel” and “instrument”, which were found in 53% of the sources. Additionally, four content words that occurred in 17 sources were “aeroplanes”, “errors”, “height”, and “messages”, which were found in 57% of the sources.

However, some words that are found in many sources still cannot be considered as technical words as there may be some words that the readers already know. Therefore, there needs to be another filter to cut out these words.

3. **Lexical Profiling**

A total of 247 words passed the frequency and range criteria and were categorized using the AntWordProfiler program, as illustrated in Table 5.

Table 5 Percentage of Words that Passed the Frequency and Range Criteria

Word List	Number of Words	Percentage
1st 1,000 GSL	95	38.46%
2nd 1,000 GSL	21	8.50%
AWL	40	16.19%
FWL	60	24.30%
Out of List	31	12.55%
Total	**247**	**100%**

In the table above, the words that most people should know occur in the 1st 1,000 words in the GSL while the second highest are words that occur in the FWL which are function words that people should know.

The words that met this criteria were words not found in the referent word lists including GSL, AWL, and FWL. The third criteria resulted in 31 words, which occur in situations where ATCOs need to communicate with other units and pilots, know the standards for air safety, and recommended procedures in case of an emergency.

Below are examples of content words that passed the Lexical Profiling criterion:

aeronautical	*aircraft*	*annex*	*aviation*
icao	*navigation*	*guidance*	*pilot*
traffic	*aerodrome*	*aerodromes*	*emergency*
altitude	*vertical*	*meteorological*	

According to Rungrueang et al. (2022), 31 out of list words are still not enough to form a word list, so the researcher has to reconsider the words that appear in GSL and AWL.

4. Keyword Analysis

The fourth criterion reassessed the words that were featured in the GSL and AWL as there may be some words with multiple meanings that have significant relevance for ATCOs to be able to understand the ICAO Annexes. After considering the first 541 words calculated by the Key-BNC program, only 124 words appeared in both GSL and AWL.

The 31 words that passed the third criterion, Lexical Profiling, were also included with the 124 words that passed the fourth criterion, Keyword Analysis, making it a total of 155 words. However, there some alphabet letters found in the 155 words and cut from the list as these letters were included in conditions and equations used for calculation. The alphabet letters that were excluded included “b”, “c”, “d”, “e”, “f”, “g”, “h”, “l”, “k”, “m", “n”, “s”, “t”, and “x” which were a total of 14 letters. Nevertheless, it was found that some abbreviations remained in the word list as these abbreviations might be significant to understand the ICAO publications. Examples of these abbreviations include “att”, “app”, and “icao.”

The total was reduced to 141 words after eliminating the letters, and these words were then checked by experts in the fifth criterion.

Below are examples of content words that passed the Keyword Analysis criterion:

aeroplane	*approach*	*authority*
centre	*communication*	*equipment*
flight	*ground*	*instrument*
lights noise	*operation*	*pressure*
radio	*service*	

Interestingly, the word “instrument” does not only mean the musical instrument, but it can also mean tools that are used in the cockpit to control the aircraft. As well as the word “ground”, which does not just refer to the ground we stand on or the surface of the earth, but it can also mean the aircraft movements on the ground such as landing, taxiway, and runway.

5. Expert Viewpoints

The fifth criterion included three experts’ points of view on the significance of the 141 words included in the word list (see Appendix A). Among the three experts, two experts currently work as air traffic management instructors, while the third expert has experience working as an ATCO for almost a decade and currently hold an ICAO language proficiency level 5. ATCOs that hold an ICAO language proficiency level 5 means they have a native or bilingual proficiency which they can communicate with high performance and effectively. The researcher sent the rating form to the three experts by email and asked them to rate the words based on a 4-point rating scale.

The researcher provided the three experts with a 4-point rating scale. “1” means words that are unrelated to the field, “2” means words that hold slight relevance to the field, “3” means words that hold high relevance to the field, and “4” means words that are fully relevance to the field. The average score of each word was calculated as shown below.

$$\frac{score\ of\ expert\ 1 + score\ of\ expert\ 2 + score\ of\ expert\ 3}{3}$$

The average score of each word was categorized into four levels as follows.

Table 6 The Interpretation of the Average Score for Each Word

Average Score	Description
1.00 – 1.75	Words unrelated to the field
1.76 – 2.50	Words pertaining minimal relevance to the field
2.51 – 3.25	Words pertaining high relevance to the field
3.26 – 4.00	Words completely relevant to the field

Of the 141 words, there were two words that received a score below the average of 2.51 because these words hold slight relevance to the field or are unrelated to the field. This resulted in a total of 139 words that passed the criterion of receiving an average of 2.51 points (see Appendix B).

Below are examples of the words that passed the Expert Viewpoints criterion.

Example 1: Aerodrome – a defined area on land or water (including any building, installations and equipment) intended to be used either wholly or in part for the arrival, departure and surface movement of aircraft (ICAO, 2005).

Annex 2: Pre-flight action for flights away from the vicinity of an aerodrome, and for all IFR flights, shall include a careful study of available current weather reports and forecasts, taking into consideration fuel requirements and an alternative course of action if the flight cannot be completed as planned.

Annex 10 Volume III: Non-aircraft transponders that are installed on aerodrome surface vehicles, obstacles or fixed Mode S target detection devices for surveillance and/or radar monitoring purposes shall be assigned 24-bit aircraft addresses.

Example 2: Aeronautical – relating to the science of designing, building, and operating aircraft (Cambridge, n.d.)

Annex 11: To ensure that aeronautical information services units obtain information to enable them to provide up-to-date pre-flight information and to meet the need for in-flight information, arrangements shall be made between aeronautical information services and air traffic services authorities responsible for air traffic services to report to the responsible aeronautical information services unit, with a minimum of delay.

Annex 19: Such personnel include, but are not limited to: flight crews; air traffic controllers; aeronautical station operators; maintenance technicians; personnel of aircraft design and manufacturing organizations; cabin crews; flight dispatchers, apron personnel and ground handling personnel.

Example 3: Airworthiness – the status of an aircraft, engine, propeller or part when it conforms to its approved design and is in a condition for safe operation (SKYbrary, n.d.)

Annex 1: Rules and regulations relevant to an aircraft maintenance licence holder including applicable airworthiness requirements governing certification and continuing airworthiness of aircraft and approved aircraft maintenance organization and procedures.

Annex 8: A Certificate of Airworthiness shall be renewed or shall remain valid, subject to the laws of the State of Registry, provided that the State of Registry shall require that the continuing airworthiness of the aircraft shall be determined by a periodical inspection at appropriate intervals having regard to lapse of time and type of service or, alternatively, by means of a system of inspection, approved by the State, that will produce at least an equivalent result.

These 139 words passed the fifth criterion as all three experts who have experience in air traffic management have rated the words that they find relevance to the field. By using the five criteria, it can be seen that there are words with multiple meaning in general English and can be used technically for air traffic management.

CHAPTER 5
DISCUSSION AND CONCLUSION

This concluding chapter synthesizes the study's findings on the vocabulary load Thai ATCOs must possess to understand ICAO publications with 95% of text coverage and what technical words should be included to create a technical word list for Thai ATCOs. Moreover, this chapter provides the discussions, recommendations, and limitations of the research study.

Summary of the Study

As there have been an abundance of a spoken corpus related to air traffic mangement, it is important to create a written corpus. Creating a written corpus fulfils the goal of understanding and analyzing written text which is mostly formal and regulatory contexts unlike a spoken corpus which captures real-time usage and is often informal and abbreviated communication. In this study, a written corpus allows the researcher to systematically examine the official language used in the ICAO publications to understand the SARPs.

The study explored the vocabulary load Thai ATCOs must possess to understand 95% of the text coverage of ICAO publications and what technical words would be included in a technical word list for Thai ATCOs. The corpus was created using the AntConc program by importing the 30 ICAO publications taken from the 19 ICAO Annexes. The resulting corpus was compared with the BNC/COCA word family lists in the AntWordProfiler program. In order to understand ICAO Annexes, Thai ATCOs must know the first 4,000 words combined with the proper nouns, marginal words, transparent compounds, and acronyms from the BNC/COCA word family lists. Instructors for training courses must have teaching materials, such as vocabulary worksheets, flashcards, reading prompts, so that the students know these 4,000 words combined with the proper nouns, marginal words, transparent compounds, and acronyms. These four additional lists are not too difficult for them especially proper

nouns which are general knowledge such as cities and countries. The results also indicated that the four additional lists accounted for a relatively significant amount of coverage (15.29 %). This highlights their importance for gaining comprehension. Particularly, proper nouns provided the highest amount of coverage, showing their relative significance for comprehension while reading Annexes. However, proper nouns might not be too difficult for Thai ATCOs as most of them are vocabulary of cities and countries, such as “Miami” and “Manila”.

The word list was created using the five criteria suggested by Rungrueang et al. (2022) including, frequency, range, lexical profiling, keyword analysis, and expert viewpoints. The word list resulted in a total of 139 words that showed the linguistic features of the ICAO Annexes. Example of words in the word list include “aerodrome”, “level”, “pressure”, “state”, and “traffic”. By focusing on a written corpus, the technical word list ensures the relevant words for Thai ATCOs when studying for air traffic control licensing and certification, interpreting ICAO procedures, and ensuring compliance with safety procedures.

Discussion

This word list intends to support Thai ATCOs in effectively interpreting ICAO publications by helping them recognize words which are essential for improving understanding and promoting ssafety. Unlike previous research studies that mainly focused on spoken corpora, such as the ATCOSIM corpus created by Hofbauer et al. (2008) from live simulation recordings, this research study developed a corpus using written sources.

Thai ATCOs should have vocabulary load of 4,000 word families in order to understand 95.68% of the ICAO publications. Similarly, Hue’s (2023) research study investigated the vocabulary load of engineering textbooks for EFL students and developed a corpus with 4,575,413 running words using 100 college engineering textbooks that covered 20 areas related to engineering such as aerospace, electrical, and medical. The study found that EFL students should know 5,000 word families

combined with proper nouns, apparent compounds, and acronyms to have a lexical coverage of 95%. The cumulative percentage of the first 5,000 words including proper nouns, apparent compounds, and acronyms was 95.53%.

The ATM word list resulted in 139 words. This one is specifically designed to highlight vocabulary crucial for ATCOs to interpret the ICAO Annexes accurately unlike other word lists related to the aviation industry. For example, the Airline Business Word List (ABWL) by Laosrirattanachai and Ruangjaroon (2021) focused on the hospitality aspect specifically in the airline business context. The ABWL contained a total of 245 words created specifically for ESP learners in the airline business context such as “traffic”, “airline”, “flight”, “lounge”, and “baggage”. It can be seen that words like “flight”, “route”, “instrument”, and “service” exist in both of the ABWL and the ATM word lists. However, in the ABWL, the word “traffic” would mean the flow of passengers and airline while in this research study, “traffic” would refer to the air traffic movement monitored by ATCOs. Even though there are some words that overlap between both lists, they both serve different purposes. Moreover, some words in the ATM word list were not found in the ABWL, for example, “aerodrome” which refers to a defined area intended to for the movement of an aircraft.

Another example is the Tower Aviation Radiotelephony Technical Vocabulary List (TARTVL) by Drayton and Coxhead (2023) which focused on spoken standard phraseology between ATCOs and other aviation personnel. The TARTVL is useful when studying about radiotelephony, phraseology, and spoken communication. However, they often focus on short commands and phrases, and may lack the variety of vocabulary for reading comprehension. The words found in the list such as “clear”, “taxi”, “approach”, “caution”, and “approved” are used to lower misunderstanding between ATCOs and pilots. Conversely, words that were commonly found such as “appropriate”, “ensure”, “necessary”, “recommended”, and “standard” indicate the administrative and authoritative tone of the written context used in the ICAO publications.

This technical word list increases Thai ATCOs' understanding and comprehension of ICAO Annexes. For instance, words like "area", "authority", "flight", "instrument", and "position" can be found in Annex 2 - Rules of the Air. These words allow Thai ATCOs to understand the contents of Annex 2 about general rules such as visual flight and instrument flight rules. Additionally, words like "accordance, procedures, point, route, traffic" can be found in Annex 11 - Air Traffic Services. These words allow Thai ATCOs to understand the standards and recommended practices of air traffic services worldwide including communication with pilots and managing air traffic to prevent mid-air collisions while promoting an orderly flow of air traffic.

Interestingly, there are several common words such as "defined", "figure", "note", "shown", and "table" that appeared in the word List. This likely occurs from the structure of the ICAO Annexes which are written formally about standards that Contracting States must follow. These terms enhance precision and uniformity as these publications contain definitions, figures, notes, and tables to present and clarify complex information.

Nonetheless, applying the five criteria suggested by Rungrueang et al. (2022) may have some drawbacks as certain words may have been mistakenly excluded. Some words that are general words and used as everyday vocabulary were excluded even though they may carry specific technical meanings in the aviation context. For example, "report" was filtered out from the word list because of the conditions of some criteria. Although in the ICAO Annexes, it refers to situations where pilots must communicate their position to the ATCOs in order for them to check their position, this word is generally linked with submitting written documents. As a result, this word was excluded from the word list.

Recommendations for ATCO Training and Air Traffic Management Programs

Thailand's aviation safety standards was identified by ICAO of having 572 safety-related deficiencies in 2015. As a result, Thailand received a red flag from ICAO (Shoowong, 2025) which led to the dissolution of the Department of Civil Aviation and

the establishment of the Civil Aviation Authority of Thailand in its place. From this situation, Thai-registered airlines had limit on expanding their international routes which took a heavy blow to the country's travel and tourism sectors (Mahitthirook, 2015). For this reason, it is important to improve aviation personnel's understanding of ICAO standards to prevent similar issues from occurring in the future. This study would contribute to the goal of offering a specific word list for air traffic management aimed to help Thai ATCOs to better understand ICAO Annexes. In addition, this technical word list can be valuable tool for training, especially in air traffic management programs to strengthen Thai ATCOs skills in aviation English. Instructors are able to focus on the key technical words so that they are able to prepare training materials to align with ICAO guidelines and ensure Thailand remains compliant with global regulations.

This word list can also benefit students enrolled in undergraduate programs about air traffic management. While some words from the word list like "area", "centre", "level", and "recommended" may be seen as vocabulary in general English, their meanings have a specific purpose in the context of air traffic management. Instructors can help students futher understand by guiding them on how to use these words in real-life operations. According to Douglas (2014), instructors should create materials and teaching techniques to train ATCO and pilot students, especially those who are non-native speakers, in communication strategies. An effective teaching strategy for instructors is to combine the vocabulary learning with real-life materials which allows students to see the actual functions of air traffic management communications. This allows students to not only build familarity with the vocabulary but can also help them to understand the working application. By highlighting the specific context in which these words are used, instructors can support students to have a deeper understanding of specific vocabulary used in air traffic management which can help them prepare professional communications.

In addition, the word list can be incorporated into English Specific Purposes (ESP) courses, such as courses related to Aviation English or English used in air traffic management. These courses can support students within English majors to be

familiarized with the technical terms needed for roles in the aviation industry and help build a strong foundation for their future careers related to the field. This research emphasizes the need for training initiatives in Thailand to enhance ATCOs' ability to understand technical terms. By strengthening their knowledge and command of these technical terms, they are able to interpret and adhere to international standards. The word list was developed to target the authentic language ATCOs use when adhering to the ICAO SARPs. According to Tokar and Fainman (2018), lexical competence enhances comprehension skills, ensuring proficiency in interaction in aviation. Therefore, this word list can be incorporated into training which may help minimize language-related misunderstandings that could compromise safety. Thai ATCOs can improve their knowledge in these technical terms which can reduce the risk of of misinterpretation. This specific linguistic tool not only contributes to the goal of aviation safety, but also better prepares Thai ATCOs for effective international collaboration. Additionally, the word list addresses existing language barriers, and it has potential to drive long-term enhancements to the training in Thailand. Moreover, the outcomes of this study could benefit other non-native English speaking countries that aims to strengthen their ATCOs' comprehension of ICAO publications to promote secure air traffic management communication. For example, ATCOs from Mongolia and Bhutan come to Thailand for training, and this word list can be used with these students on an international level.

Limitations and Suggestions for Further Research

The aviation industry is constantly growing and progressing which may result in some SARPs to be updated or changed. This may result in irrelevant words in the word list. Therefore, the word list should be updated from time to time to ensure relevancy and incorporated in real world training to examine the its effectiveness. This research study can also be extended to other ICAO publications, such as Procedures for Air Navigation Services (PANS), Technical Manuals, and Circulars. For instance, Doc 8168 - PANS - OPS (PANS), Doc 1010 - Manual on Space Weather Information in Support of

International Air Navigation (Technical Manuals), and Cir 314 - Threat and Error Management (TEM) in Air Traffic Control (Circulars). These ICAO publications can provide a deeper understanding in other areas of the aviation industry that includes pilots, aircraft maintenance, and ground crew. Researchers can also examine from safety reports or incident reports of actual accidents to see if vocbaulary is a reason for errors in operations. By expanding the horizon of this research study, it will allow better perception towards vocabulary related to the aviaiton industry and improvement in training for increased safety.

REFERENCES

AEROTHAI. (2023). *Role and operation: Air traffic services*. https://www.aerothai.co.th/en/services/air-traffic-services

Anthony, L. (2022). *AntFileConverter (Version 2.0.2)*. In Waseda University. https://www.laurenceanthony.net/software

Anthony, L. (2023a). *AntConc (Version 4.2.0)*. In Waseda University. https://www.laurenceanthony.net/software

Anthony, L. (2023b). *AntWordProfiler (Version 2.1.0)*. In Waseda University. https://www.laurenceanthony.net/software

Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. University of Michigan Press ELT. https://doi.org/10.3998/mpub.371534

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, *24*, 199-208. https://doi.org/10.1017/S0272263102002048

Bowker, L., & Pearson, J. (2002). *Working with specialized language*. Routledge. https://doi.org/10.4324/9780203469255

Cabré, M. T., & Sager, J. C. (1999). *Terminology: Theory, methods, and applications*. J. Benjamins Publishing Company. https://doi.org/10.1075/BABEL.50.1.15CHU

Carstairs-McCarthy, A. (2002). *An introduction to English morphology: Words and their structure*. Edinburgh University Press. https://doi.org/10.1515/9781474428989

Chen, H., & Lei, G. (2019). Developing a technical words list for research articles in computer science discipline. *English Language Teaching*, *12*(10), 131-141. https://doi.org/10.5539/elt.v12n10p131

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a*

Foreign Language, *15*.

Corson, D. (1997). The learning and use of academic English words. *Language Learning 47*(4), 671-718. https://doi.org/10.1111/0023-8333.00025

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238. https://doi.org/10.2307/3587951

Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of plumbing. *English for Specific Purposes*, *51*, 84-97. https://doi.org/10.1016/j.esp.2018.03.006

Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. In J. F. M. Peacock (Ed.), *Research Perspectives on English for Academic Purposes*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524766.020

Crocker, D. (2005). *Dictionary of aviation*. A&C Black Publishers.

Dash, N. S. (2010). Corpus linguistics: A general introduction. *CIIL*.

Davis, S., & Navarro, M. (2015). Building a technical vocabulary: A case study in teaching technical communication. *IEEE Transactions on Professional Communication*, *58*(4), 367-383. https://doi.org/10.1109/tpc.2015.2487759

Dazdarevic, S., Zoranic, A.-L., & Fijuljanin, F. (2015). Benefits of corpus-based approach to language teaching. *BADEN Newsletter*, *7*.

Dib, C. (2022). *Publication spotlight: The ICAO annexes to the convention on international civil aviation*. https://unitingaviation.com/news/safety/publication-spotlight-the-icao-annexes-to-the-convention-on-international-civil-aviation/#:~:text=Annexes%20set%20out%20the%20Standards,must%20do%20to%20achieve%20them.

Drayton, J., & Coxhead, A. (2023). The development, evaluation and application of an aviation radiotelephony specialised technical vocabulary list. *English for Specific Purposes*, *69*, 51-66. https://doi.org/10.1016/j.esp.2022.10.001

Estival, D., & Farris, C. (2016). Aviation English as a lingua franca. In D. Estival, C. Farris, & B. Molesworth (Eds.), *Aviation English: A lingua france for pilots and air traffic*

controllers. Routeledge. https://doi.org/10.4324/9781315661179

Faxriddinovna, J. G. (2021). Corpus linguistics and corpus-based approach in foreign language teaching. *World Bulletin of Management and Law (WBML)*, *3*.

Florio, F. D. (2006). *Airworthiness: An introduction to aircraft certification*. Elsevier Ltd.

Flowerdew, L. (2004). The argument of using English specialized corpora to understand academic and professional language. In *Discourse in the Professions: Perspectives from Corpus Linguistics* (pp. 11-33). John Benjamins. https://doi.org/10.1075/scl.16.02flo

Fuchs, G., Pizam, A., & Wang, Y. (2011). The importance of safety and security for tourism destinations. In K. Marshall (Ed.), *Destination marketing and management: Theories and applications*. Scitus Academics. https://doi.org/10.1079/9781845937621.0300

Fuertes-Olivera, P. A. (2008). Pedagogical application of specialized corpora in ESP teaching: The case of the UVaSTECorpus. *Scripta Manent*, *3*(2), 68-81.

Garcia, A. C. M. (2023). Investigating the construct of aeronautical English listening testing: A qualitative analysis of the ICAO rating scale. *Journal of Teaching English for Specific and Academic Purposes*, *11*(1), 69-86. https://doi.org/https://doi.org/10.22190/JTESAP230220007G

Gilquin, G. (2015). From design to collection of learner corpora. In S. G. Granger, Gaetanelle; Meunier, Fanny (Ed.), *The Cambridge handbook of learner corpus research* (pp. 9-34). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414.002

Granger, S., Gilquin, G., & Meunier, F. (2015). *The Cambridge handbook of learner corpus research*. Cambridge University Press. https://doi.org/10.1017/CBO9781139649414

Heidari, F., Jalilifar, A., & Salimi, A. (2020). Developing a corpus-based word list in pharmacy research articles: A focus in academic culture. *International Journal of Society, Culture & Language*, *8*(1), 1-14.

Hong, Y., & Wen, J. (2025). Lexical coverage in science popularization discourse: The

case of popular science news from Scientific American. *English for Specific Purposes*, *77*, 45-55. https://doi.org/https://doi.org/10.1016/j.esp.2024.10.001

Hsu, W. (2011). The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes*, *30*(4), 247-257. https://doi.org/10.1016/j.esp.2011.04.005

ICAO. (2006). *Convention on international civil aviation (Doc 7300)*. ICAO.

ICAO. (2017). *Manual on air traffic controller competency-based training and assessment (Doc 10056)*. ICAO.

ICAO. (2018). *Annex 11 - Air Traffic Services*. ICAO.

ICAO. (2023). *Products & services catalogue 2023*. ICAO.

ICAO. (n.d.-a). *About ICAO*. https://www.icao.int/about-icao/Pages/default.aspx

ICAO. (n.d.-b). *History of ICAO*. https://www.icao.int/about-icao/History/Pages/default.aspx

Jeppesen. (2012). *The aviation dictionary: For pilots and aviation maintenance technicians*. Jeppesen Maintenance.

Kaur, R. (2021). Identifying aircraft maintenance trainees' English language learning needs. *Malaysian Journal of ELT Research*, *18*(2), 19-31. https://doi.org/10.52696/IJCB2913

Koester, A. (2010). Building a small specialised corpora. In *The Routledge Handbook of Corpus Linguistics* (pp. 66-79). Routledge. https://doi.org/10.4324/9780203856949.CH6

Laosrirattanachai, P., & Ruangjaroon, S. (2021). Corpus-based creation of tourism, hotel, and airline business word lists. *LEARN Journal: Language Education and Acquisition Research Network*, *14*(1), 50-86.

Lardiere, D. (2006). Attainment and acquirability in second language acquisition. *Second Language Research*, *22*(3), 239-242. https://doi.org/10.1191/0267658306sr267ed

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*, 15-30.

Lohmann, G., Albers, S., Koch, B., & Pavlovich, K. (2009). From hub to tourist destination -

An explorative study of Singapore and Dubai's aviation-based transformation. *Journal of Air Transport Management*, *15*(5), 205-211. https://doi.org/10.1016/j.jairtraman.2008.07.004

Mahitthirook, A. (2015). *ICAO red-flags Thailand's aviation safety standard*. The Bangkok Post. https://www.bangkokpost.com/thailand/general/596708/icao-red-flags-thailand-aviation-safety-standard

Mekkaoui, G., & Mouhadjer, N. (2019). Addressing air traffic controllers' English language proficiency needs: Case of Zenata airport. *Global Journal of Foreign Language Teaching*, *9*(3), 167-183. https://doi.org/10.18844/gjflt.v9i3.4245

Meunier, F., & Granger, S. (2008). *Phraseology in foreign language learning and teaching*. John Benjamins Publishing Company. https://doi.org/10.1075/Z.138

Nagy, D. (2019). *Pragmatism of modal verbs: Case study of 'ICAO' requirements.* International Conference RCIC'2019: Redefining Community in Intercultural Context, https://www.afahc.ro/ro/rcic/2019/rcic'19/volum_2019/131-136%20Nagy.pdf

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes*, *63*, 59-82. https://doi.org/10.1353/cml.2006.0049

Ngula, R. (2017). Corpus methods in language studies. In D. D. Kuupole (Ed.), *Perspectives on conducting and reporting research*. University of Cape Coast Press.

Nurmukhamedov, U. (2017). Lexical coverage of TED Talks implications for vocabulary instruction. *TESOL Journal*, *8*(4), 768-790. https://doi.org/10.1002/tesj.323

OAG. (2023). *What are airport hubs and what is their significance in the travel industry today?* https://www.oag.com/blog/what-are-airport-hubs-and-what-is-their-significance-in-the-travel-industry-today

Palinkasevic, R. (2017). Specialized word lists - Survey of the literature - Research

perspective. *Research in Pedagogy*, *7*(2), 221-238. https://doi.org/10.17810/2015.61

Park, H., & Nam, D. (2017). Corpus linguistics research trends from 1997 to 2016: A co-citation analysis. *Linguistic Research*, *34*, 427-457.

Pepin, E. (1952). ICAO and other agencies dealing with air regulations. *Journal of Air Law and Commerce*, *19*(2), 152-165.

PRD. (2024). *Airports of Thailand is pushing for developing Thailand as a global aviation hub*. https://thailand.prd.go.th/en/content/category/detail/id/48/iid/264816

Ragan, P. H. (1996). Aviation English: An introduction. *Journal of Aviation/Aerospace Education & Research*, *7*(2). https://doi.org/10.15394/jaaer.1997.1189

Rahmat, A., & As'ary, M. A. (2017). The application of EGP materials to ATC students of Casea Makassar. *LEKSEMA: Jurnal Bahasa Dan Sastra*, *2*(1), 65-75. https://doi.org/10.22515/ljbs.v2i1.630

Rungrueang, T., Boonprasert, P., Poempongsajaroen, S., & Laosrirattanachai, P. (2022). Corpus-based approach to generate a word list for food service. *THAITESOL Journal*, *35*(1), 57-76.

Shoowong, M. (2025). *Air safety standards scrutinised*. The Bangkok Post. https://www.bangkokpost.com/business/general/2947182/air-safety-standards-scrutinised

Sinclair, J. (2005). Corpus and text - Basic principles. In *Developing Linguistic Corpora: A Guide to Good Practice* (pp. 1-16). Oxbow Books. https://doi.org/10.1007/978-3-030-46216-1

SKYbrary. (n.d.). *ATCO training overview*. SKYbrary. https://skybrary.aero/articles/atco-training-overview#:~:text=A%2F%20Initial%20training%20%28called%20also%20Institutional%20training%29%20is,purpose.%20It%20comprises%20%27basic%20training%27%20and%20%27rating%20training%27.

Smith, J., & Johnson, R. (2018). Development and evaluation of technical word lists for engineering students. *International Journal of Engineering Education*, *34*(6), 1754-

1763.

Smith, S. (2023). *BNC/COCA lists*. https://www.eapfoundation.com/vocab/general/bnccoca/

Sukma, M. M., Rochmawati, L., & Fatmawati, F. (2019). The methods and learning design of English for specific purpose for aircraft maintenance engineering subject in aviation English. *Jurnal Penelitian*, *4*(2), 60-69. https://doi.org/10.46491/jp.v4e2.297.60-69

Terenzi, D. (2021). Overcoming challenges in English for aviation maintenance': Technical publications selection for the construction of a corpus and its use to teach language aspects considering learners needs. *Open Journal of Applied Sciences*, *11*, 1122-1134. https://doi.org/10.4236/ojapps.2021.1110084

Thiankasem, T. (2018). Investigating the technical vocabulary in cabin crew manuals: A corpus-based study [Master's Thesis, Thammasat University]. https://doi.org/10.14457/TU.the.2018.1441

Vaughan, E., & O'Keefe, A. (2015). Corpus analysis. In *The International Encyclopedia of Language and Social Interaction* (pp. 252-268). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118611463.wbielsi168

Viera, R. T. (2017). Vocabulary knowledge in the production of written texts: A case study on EFL language learners.

VUW. (n.d.). *Vocabulary analysis programs*. Victoria University of Wellington. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs

Watson Todd, R., & Pojanapunya, P. (n.d.). *Key-BNC*. http://crs2.kmutt.ac.th/Key-BNC/

Webb, S. (2021). Research investigating lexical coverage and lexical profiling: What we know, what we don't know, and what needs to be examined. *Reading in a Foreign Language*, *33*(2), 287-302.

Yang, H., & Liu, W. (2019). Dual-hub connectivity: A case study on China Eastern Airlines in Shanghai *European Transport Research Review*, *11*(25), 1-9. https://doi.org/10.1186/s12544-019-0364-6

Yildiz, M. (2023). Lexical coverage required for minimal and optimal levels of reading comprehension in the English tests of the higher education institutions examination. *rEFLections*, *30*(3), 695-711. https://doi.org/https://doi.org/10.61508/refl.v30i3.268077

Zuluaga-Gomez, J., Vesely, K., Szoke, I., Blatt, A., Motlicek, P., Kocour, M., Rigault, M., Prasad, K. C. A., Sarfjoo, S. S., Nigmatulina, I., Cevenini, C., Kolčárek, P., Tart, A., & Cernocky, J. (2022). ATCO2 corpus a large-scale dataset for research on automatic speech recognition and natural language understadning of air traffic communications. *SSRN*, 1-29.

APPENDICES

APPENDIX A

Expert Viewpoints

No.	Words	Expert 1	Expert 2	Expert 3	Mean
1	accordance	4	4	4	4.00
2	aerodrome	4	4	4	4.00
3	aerodromes	4	3	4	3.67
4	aeronautical	4	4	4	4.00
5	aeroplane	4	4	4	4.00
6	aeroplanes	4	3	4	3.67
7	air	4	4	4	4.00
8	aircraft	4	4	4	4.00
9	airworthiness	4	4	4	4.00
10	altitude	4	4	4	4.00
11	annex	4	4	4	4.00
12	app	2	4	4	3.33
13	appendix	4	4	4	4.00
14	applicable	4	3	4	3.67
15	application	4	4	4	4.00
16	approach	4	4	4	4.00
17	appropriate	4	3	4	3.67
18	approved	4	4	4	4.00
19	area	4	4	4	4.00
20	associated	4	4	4	4.00
21	att	1	4	1	2.00
22	attachment	1	4	4	3.00

No.	Words	Expert 1	Expert 2	Expert 3	Mean
23	authority	4	4	4	4.00
24	available	4	4	4	4.00
25	aviation	4	4	4	4.00
26	centre	4	4	4	4.00
27	certification	4	4	4	4.00
28	chapter	4	4	4	4.00
29	characteristics	4	4	4	4.00
30	cid	1	4	1	2.00
31	code	4	4	4	4.00
32	communication	4	4	4	4.00
33	communications	4	3	4	3.67
34	conditions	4	4	4	4.00
35	contained	3	3	4	3.33
36	contracting	4	4	4	4.00
37	control	4	4	4	4.00
38	coverage	3	4	4	3.67
39	crew	4	4	4	4.00
40	data	4	4	4	4.00
41	defined	4	3	4	3.67
42	design	3	3	4	3.33
43	distance	4	4	4	4.00
44	doc	4	4	4	4.00
45	elevation	4	4	4	4.00
46	emergency	4	4	4	4.00

No.	Words	Expert 1	Expert 2	Expert 3	Mean
47	engine	4	4	4	4.00
48	ensure	4	3	4	3.67
49	equipment	4	4	4	4.00
50	error	4	4	4	4.00
51	established	4	4	4	4.00
52	field	4	4	4	4.00
53	figure	3	3	4	3.33
54	flight	4	4	4	4.00
55	following	4	4	4	4.00
56	frequency	4	4	4	4.00
57	ft	4	4	4	4.00
58	fuel	4	4	4	4.00
59	ground	4	4	4	4.00
60	guidance	4	4	4	4.00
61	height	4	4	4	4.00
62	icao	4	4	4	4.00
63	include	4	3	4	3.67
64	including	4	3	4	3.67
65	information	4	4	4	4.00
66	Instrument	4	4	4	4.00
67	intended	4	3	4	3.67
68	international	4	4	4	4.00
69	landing	4	4	4	4.00
70	level	4	4	4	4.00

No.	Words	Expert 1	Expert 2	Expert 3	Mean
71	lights	4	4	4	4.00
72	line	3	3	4	3.33
73	link	3	3	4	3.33
74	manual	4	4	4	4.00
75	mass	4	4	4	4.00
76	material	3	4	4	3.67
77	maximum	4	3	4	3.67
78	measurement	4	4	4	4.00
79	message	4	3	4	3.67
80	messages	4	4	4	4.00
81	meteorological	4	4	4	4.00
82	minimum	4	4	4	4.00
83	mode	4	4	4	4.00
84	navigation	4	4	4	4.00
85	necessary	4	4	4	4.00
86	noise	4	4	4	4.00
87	note	4	4	4	4.00
88	operating	4	4	4	4.00
89	operation	4	4	4	4.00
90	operational	4	4	4	4.00
91	operations	4	3	4	3.67
92	operator	4	4	4	4.00
93	part	4	4	4	4.00
94	pilot	4	4	4	4.00

No.	Words	Expert 1	Expert 2	Expert 3	Mean
95	point	4	4	4	4.00
96	position	4	4	4	4.00
97	practices	4	4	4	4.00
98	pressure	4	4	4	4.00
99	procedure	4	4	4	4.00
100	procedures	4	4	4	4.00
101	protection	4	3	4	3.67
102	provide	4	3	4	3.67
103	provided	4	4	4	4.00
104	provisions	4	3	4	3.67
105	radio	4	4	4	4.00
106	range	4	4	4	4.00
107	recommendation	4	4	4	4.00
108	recommended	4	4	4	4.00
109	reference	4	4	4	4.00
110	related	4	4	4	4.00
111	required	4	4	4	4.00
112	requirements	4	4	4	4.00
113	route	4	4	4	4.00
114	runway	4	4	4	4.00
115	safety	4	4	4	4.00
116	service	4	4	4	4.00
117	services	4	4	4	4.00
118	shown	4	3	4	3.67

No.	Words	Expert 1	Expert 2	Expert 3	Mean
119	signal	4	4	4	4.00
120	specific	4	4	4	4.00
121	specifications	4	4	4	4.00
122	specified	4	3	4	3.67
123	speed	4	4	4	4.00
124	standard	4	4	4	4.00
125	standards	4	4	4	4.00
126	state	4	4	4	4.00
127	states	4	3	4	3.67
128	station	4	4	4	4.00
129	surface	4	4	4	4.00
130	system	4	4	4	4.00
131	systems	4	4	4	4.00
132	table	4	3	4	3.67
133	traffic	4	4	4	4.00
134	transmission	4	4	4	4.00
135	type	4	4	4	4.00
136	unit	4	3	4	3.67
137	units	4	4	4	4.00
138	value	4	3	4	3.67
139	vertical	4	4	4	4.00
140	visual	4	4	4	4.00
141	volume	4	4	4	4.00

APPENDIX B

Air Traffic Management (ATM) Word List

1	accordance	2	aerodrome
3	aerodromes	4	aeronautical
5	aeroplane	6	aeroplanes
7	air	8	aircraft
9	airworthiness	10	altitude
11	annex	12	app
13	appendix	14	applicable
15	application	16	approach
17	appropriate	18	approved
19	area	20	associated
21	attachment	22	authority
23	available	24	aviation
25	centre	26	certification
27	chapter	28	characteristics
29	code	30	communication
31	communications	32	conditions
33	contained	34	contracting
35	control	36	coverage
37	crew	38	data
39	defined	40	design
41	distance	42	doc
43	elevation	44	emergency
45	engine	46	ensure
47	equipment	48	error
49	established	50	field
51	figure	52	flight
53	following	54	frequency
55	ft	56	fuel

57	ground	58	guidance
59	height	60	icao
61	include	62	including
63	information	64	instrument
65	intended	66	international
67	landing	68	level
69	lights	70	line
71	link	72	manual
73	mass	74	material
75	maximum	76	measurement
77	message	78	messages
79	meteorological	80	minimum
81	mode	82	navigation
83	necessary	84	noise
85	note	86	operating
87	operation	88	operational
89	operations	90	operator
91	part	92	pilot
93	point	94	position
95	practices	96	pressure
97	procedure	98	procedures
99	protection	100	provide
101	provided	102	provisions
103	radio	104	range
105	recommendation	106	recommended
107	reference	108	related
109	required	110	requirements
111	route	112	runway
113	safety	114	service

115	services	116	shown
117	signal	118	specific
119	specifications	120	specified
121	speed	122	standard
123	standards	124	state
125	states	126	station
127	surface	128	system
129	systems	130	table
131	traffic	132	transmission
133	type	134	unit
135	units	136	value
137	vertical	138	visual
139	volume		

VITA