THE JOURNAL OF TEACHING ENGLISH FOR SPECIFIC AND ACADEMIC PURPOSES
Vol. 11, Nº 1, 2023, pp. 41–56

UDC: 811.111'276.6:358.4 https://doi.org/10.22190/JTESAP230103005C

Review research paper

LANGUAGE OF AIRCRAFT DOCUMENTATION: USING A MULTILINGUAL CORPUS IN COLLABORATIVE WORK IN LEXICOGRAPHY AND PEDAGOGY

Malila Carvalho de Almeida Prado[1], Daniela Terenzi[2], Patrícia Tosqui-Lucks[3], Ziqian Zhou[1]

[1]BNU-HKBU United International College, China
[2]Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Brazil
[3]Airspace Control Institute (ICEA), Brazil

Abstract. *This paper describes collaborative work by two higher education institutions in Brazil and China that joined forces to build a corpus with a view to investigating the language of aircraft maintenance documentation for teaching and lexicography purposes. The corpus consists of three languages, including English, with Chinese and Portuguese still under construction. The paper opens with an overview of language-related studies in the aircraft maintenance field, drawing attention to this specific niche and its demands. It discusses trials already performed with the English and Chinese sub-corpora, challenges faced, and the importance of task-sharing in multidisciplinary teams. It emphasizes the need to combine linguistics and professional expertise not only to develop materials for the aviation industry but also to assist the development of trainees who can experience a professional setting during their education. The corpus will be used in (among others) preparing pedagogical material for teaching not only aircraft maintenance personnel but also pilots and air traffic controllers. The corpus will also serve as a lexicographic resource for designing a visual dictionary with collocates and constructing trilingual glossaries.*

Key words: *multilingual corpus, aviation English, teaching, lexicography*

1. INTRODUCTION

Aviation English in a broad sense consists of the language used by all professionals involved in aviation. Significant attention is given to pilot and air traffic controller communication because of the implementation of the Language Proficiency Requirements (LPRs, ICAO 2010), which put forth practices related to a minimum level of English proficiency for licensing these professionals. However, this recognition often marginalizes other professional domains and genres within the area of aviation.

One such domain concerns documentation such as manuals, checklists, charts, reports, and logbooks, mostly intersecting aircraft maintenance engineers, flight dispatchers, pilots, flight attendants, and other professionals operating in airlines or private aviation companies. The use of such documentation goes beyond reading skills as professionals need to use it when

Submitted January 31st, 2023, accepted for publication March 5th, 2023
Corresponding author: Malila Carvalho de Almeida Prado, BNU-HKBU United International College, China
E-mail: malilaprado@uic.edu.cn

performing tasks within or across teams. For example, engineers may need to talk to pilots about problems that occurred on a previous flight, write reports, refer to manuals to carry out aircraft maintenance, crosscheck items with cargo staff and flight dispatchers, and make decisions regarding the airworthiness of the aircraft. While this may take place in the local language in exchanges involving national carriers, multinational teams are commonly involved in such tasks (Ma, Drury, and Marin 2009). Additionally, as most documentation comes from the aircraft manufacturers, all materials are produced in English to facilitate the mobility of aircraft for transportation of passengers and cargo, relocation, and trade. This ensures that an aircraft can be repaired in another country, the aircraft's documentation can be inspected by international authorities, or the aircraft can be sold to a company based in a different country.

For this reason, some countries, including Brazil, opt to teach professionals how to handle documentation in English (Almeida and Prado 2011; Terenzi 2021). Maintenance technicians, pilots, and flight attendants (among others) are required to take English proficiency exams assessing their reading and translation skills when they apply for jobs with airlines to demonstrate their ability to read manuals and documentation (Zhang et al. 2020). While some airlines offer their employees English for Specific Purposes (ESP) courses focusing on reading aviation documents (Almeida and Prado 2011), it is common practice to include ESP modules in aviation-related courses (Terenzi 2021). These practices are especially feasible in Brazil because there are only three major airlines in the country, which ensures a concentration of standards. Moreover, written Brazilian Portuguese is in many ways similar to English, and the many cognates along with research-based strategies facilitate the teaching of reading for decades (Bocorny 2011). Finally, code-meshing is frequently adopted, particularly as regards technical terms (Friginal, Mathews, and Roberts 2020).

China, on the other hand, follows a different scenario, including the number of airlines, a multiplicity of practices, and a greater variety of aircraft in the fleet. In addition, the Chinese writing system is significantly different from that of English, thus increasing the learning challenge. In response, translation of documentation becomes the adopted practice (Liu 2020). While Ma et al. (2009) confirm the practice, they argue that there remains a large amount of texts in English, thus compelling maintenance personnel in China to use this documentation.

Misinterpretation of aircraft manuals can lead to serious consequences, and aircraft maintenance engineers and pilots must receive appropriate training and education to understand the content of the manuals in addition to having access to up-to-date and accurate information. For this reason, a corpus of aviation documentation is of utmost importance as it can enable investigations of patterns affecting policies such as proficiency exams and language standardization such as the Simplified English initiative, pedagogical materials design for in-service (Almeida and Prado 2011) and future professionals (Terenzi 2021), and translation practices found in airlines (Liu 2020).

This paper describes the first phase of collaborative work by two higher education institutions that joined forces to build a corpus with a view to investigating the language of documentation for teaching purposes (following Bhatia 2022) as well as for lexicography purposes (following Peixoto 2020a, 2020b). The corpus will be used in (among others) designing a visual dictionary with collocates (Prado and Terenzi 2022), preparing pedagogical materials for aircraft maintenance personnel (Terenzi 2021), pilots, and air traffic controllers (Prado 2019; Tosqui-Lucks and Prado 2021), and constructing trilingual glossaries. This paper presents initial investigations from the corpus under construction.

The paper is structured as follows: it opens from an overview of language-related studies in the aircraft maintenance field, drawing attention to this specific niche and its demands. It presents a comparable corpus consisting of three languages, including English, with Chinese and Portuguese still under construction. It discusses trials conducted in the English and Chinese sub-corpora, challenges faced, and the importance of task-sharing in multidisciplinary teams. It concludes by proposing future studies the corpus might promote along with the importance of combining linguistics and professional expertise not only to develop materials for the aviation industry but also to assist the development of trainees who will experience a professional setting during their education.

2. The Language of Aircraft Maintenance

The language used in aircraft maintenance ranges from the documentation prepared by aircraft manufacturers to ground personnel involved with the loading of aircraft, mechanics responsible for the airworthiness of the aircraft, and flight and cabin crew. However, similarly to Sarmento (2008) and Bocorny (2008), our main focus is on maintenance personnel and pilots and their interpretation of reading materials.

2.1. Aircraft Maintenance Communication-related Problems

The high-stakes environment of aviation alone justifies rigorous research of all its facets, including human factors, technical advances and error protection as well as mistakes that can lead to fatal disasters. Drury and Ma (2003) have investigated misinterpretation of aircraft maintenance manuals as a contributing factor in aircraft accidents, categorizing language errors in verbal and written communication. Among verbal errors are inadequacy of the message, different accents, and the poor quality of the public address (PA) system, while written errors include difficulty understanding documents in English and poor translation.

It is important to note that accidents are typically the result of a complex combination of factors, and the cause of an accident can be difficult to determine. However, as Mathews (2020) points out, when air crashes happen, investigations often focus on technical aspects and overlook communication problems, which "ha[ve] not been given the same systematic, consistent, comprehensive review that other human elements of human performance have received" (Friginal, Mathews, and Roberts 2020, p. 62). Indeed, Mathews (2020) considers such contributions as potentially greater than officially reported.

Regarding aviation maintenance, training and accident analysis are largely based on a toolkit known as "the Dirty Dozen" (Dupont 1997), which consists of twelve factors that contribute to human errors, placing *lack of communication* in first position. Chatzi et al. (2019) address the importance of communication both in the exchange of information within and between teams and in access to aircraft maintenance documentation. In their study, the authors link communication and trust because "[a]t the organizational level, when organizational culture supports open and free communication among all levels of employees, it is expected from them to enhance their trust levels towards each other and their organization" (p. 11). To ensure that communication and trust are equally valued in aviation maintenance, the authors suggest ongoing training focused on a framework of human errors "especially in the written forms of communication, e.g., documentation, manuals, work cards, etc." (p. 11).

To mitigate errors, initiatives around the globe have proposed simplifying the language of manuals. One example is ASD-STE100 standards developed by a working group within the AeroSpace and Defense Industries Association of Europe (2021). ASD-STE100 focuses on a register intended for technical writers that is as precise and objective as possible. It presents suggestions that range from lexical to structural choices. Examples of attempts to reduce the risk of ambiguities include: (1) Lexical choices: The verb *fall* has the meaning of "*to move down by the force of gravity*" rather than "*decrease*" (p. ii); (2) Structural choices: "Use only the active voice in procedural writing. Use the active voice as much as possible in descriptive writing" (p. 1-3-1). Indeed, Sarmento (2008) confirmed the low use of the passive voice, at least when combined with modal verbs, showing that it accounts for only 16.32% of the content of aircraft maintenance manuals. Moreover, these standards are often updated with the latest research on technical writing.

2.2. Studies of Aircraft Maintenance Language

Studies of the language used in aircraft maintenance are based on the specific needs of aircraft maintenance engineers (Terenzi 2014; Terenzi and Augusto-Navarro 2018; Niamsuwan 2017; Embryany and Ratmanida 2020), hence the high frequency of technical terms and specialized vocabulary. Given that most documentation is produced in English, this highly specialized language can pose a challenge for the 80% of all aircraft technicians around the world who are non-native speakers of English (Friginal, Mathews and Roberts 2020), however mitigated by strategies such as "use of Simplified English, full translation, use of an English-speaking coach, and provision of a local language glossary" (Drury and Ma 2003, p. 49).

In the two countries under study in this paper (China and Brazil), the strategies listed in Drury and Ma (2003) and designed to mitigate communication errors are applied differently. In Brazil, most documentation, including manuals, checklists, work cards, and reports are preserved in the English original. Thus most work on language is conducted through glossaries (Terenzi 2020) and the teaching of reading and writing (White, 2018), along with a surge of corpus-informed materials in the last decade resulting from academic research (Gabrielatos and Sarmento 2006; Bocorny 2008; Zuppardo 2013; Terenzi 2021).

Using a corpus of aviation manuals, Sarmento (2008) and Bocorny (2008) investigated modal verbs and noun phrases. Sarmento (2008) analyzed how modal verbs signal problems as modal verbs frequently occur in sections related to caution and warning. She illustrates this observation with the statement: "Warning: Do not get hydraulic fluid on you. Hydraulic fluid BMS3-11 can cause injury to persons" (p. 2018). Similarly, Bocorny (2008) looked into noun phrases and how they can be taught to Brazilian Portuguese speakers given that the syntax of noun phrases in English and Portuguese differs. Terenzi (2021) suggests the use of corpus linguistics in teaching aviation English to future aircraft maintenance engineers and has supervised a number of papers examining the congruence between these areas. All authors justify the use of translation in ESP classes, as does Čarapić (2022).

Borowska (2017, p. 246) cautions that "[w]hen it comes to technical terminology, it may be sometimes useful to provide learners with L1 translation in order to reduce error occurrence, but it would not work so well as a communication strategy." Thus while translation may be a useful strategy in teaching English to aviation personnel, it can also serve as an "intervention" (Ma et al. 2009, p. 32) in order to promote better practices in professional settings.

Given that Chinese differs substantially from English in addition to the presence of a highly diversified aircraft fleet, the language needs of aircraft maintenance engineers are met

mostly through translation (Liu 2020). However, despite Liu's call for research on specialized language training for translators and aircraft maintenance engineers, including the translation of technical manuals, there is a dearth of studies in this field. Among these, Wen and Li (2011) analyzed ten popular English-Chinese translation methods based on examples from the *Chinese-English-Russian Dictionary of Aerospace.* The authors found ten translation strategies, including literal translation, free translation, and transliteration, all centered on individual English words and their equivalents in Chinese based on existing knowledge of the aviation and translation fields, not on a corpus. Yet research has shown that analyzing collocates is paramount to the understanding of subtleties that may accompany individual words forming part of particular textual features and rhetorics and therefore cultural (pre)suppositions (Trimble 1985). For example, Sarmento (2010) investigated *can*, the most common modal verb in aeronautical manuals, along with its collocates *be*, *cause*, and *result.* She observed that these collocates flag a negative consequence that may result from a poorly accomplished task or a malfunction in the aircraft. In particular, Sarmento highlights the importance of designing pedagogical materials for this particular target public by taking such linguistic behaviors present in manuals into account.

In a corpus of aviation English consisting of operation manuals, aviation law and regulations, aviation journal abstracts, and teaching materials totaling 350,000 words, Zhao (2014) analyzed the words *fail* (v.) and *failure* (n.), contrasting their use in the aviation corpus with that in BNC (Davies 2004), a general English corpus. This contrast leads her to claim that aviation English tends to adopt more verb nominalizations than general English because *failure* is used more frequently in the aviation English corpus than in the general English corpus. This alone would justify the need to employ specific language corpora in the teaching and translation of aviation English as these tools help draw attention to the phenomena present in that professional domain. Zhao also claims that enlarging the corpus will be necessary for further analysis.

Referring again to ASD-STE100, one of its injunctions is: "Do not use technical verbs as nouns" (p. 1-1-2). In ASD-STE100, verbs can be selected through a list of related technical words along with the segment of speech in which they should be used, as in the example below:

> Non-STE: Test the system for leaks.
> STE: Do the leak test of the system.
> Or
> STE: Do a test for leaks in the system. (p. 1-1-2)

The word *test* is approved as a noun in ASD-STE100. Although both the words *fail* and *failure* are also approved, common replacements for the verb are suggested in the manual:

> Non-STE: If the instrument fails to respond, do a test.
> ("Respond" is an unapproved word related to different contexts)
> STE: If the instrument does not operate correctly, do a test. (p. 1-1-14)

The initiative of simplifying the English used in manuals was shown to be particularly helpful to non-native speakers of English living in the United States (Chervak and Drury 2003) but not as effective in a subsequent study with personnel living in other countries, particularly Chinese and Spanish speakers (Ma et al. 2009). In the later study, Ma et al. mentioned that in such cases, translation was the most efficient strategy among maintenance personnel. The example of *fail* (v.) and *failure* (n.) given in Zhao (2014)

helps address the difficulty involved in translating phrasal structures found in manuals into Chinese. Yet the translations need to follow the same high-level standards of safety that underlie the writing of manuals. Thus there needs to be a multidisciplinary team working on translation choices as these do not imply stylistic choices but instead aim for objectivity, unambiguity, and clarity for the benefit of an international audience.

2.3. Translating Aircraft Maintenance Documentation

The lack of research may also hide problems that can arise when translating aircraft maintenance manuals from English into other languages because technical terminology and the use of jargon can pose a significant challenge for translators. In fact, translations of aircraft maintenance manuals may contain errors that can compromise the integrity of the aircraft and that of the personnel involved. Moreover, we found no studies on the impact of machine translation on the quality of translations of aircraft maintenance manuals, particularly in documents translated from English and then translated back to English.

Bai (2023) conducted a data analysis of 90 papers indexed by the Chinese scholar database CNKI (https://en.cnki.com.cn). The author claims that although there is a general agreement that work on aviation terminology is essential, it does not account for the complexities that exist in the air industry. Bai also suggests that in the process of translation, the translator determines not only whether a word is a civil aviation term but also its specific professional domain. Moreover, there is a discrepancy between the market, which produces a considerable number of translations, and research, which is mostly conducted by teachers and students in universities. Thus it is likely that relevant research does not converse with practice or vice-versa.

As regards research, Borowska (2017, p. 49) points out that although "national aviation languages are in use in every country [,...] scientifically nothing has been done in this field apart from some lexicographic work or presentations of aviation phraseology," which mostly focuses on radio communications between pilots and controllers. While Borowka lists the few university theses written in local languages, scientific publications need to be in a language accessible to a global audience. In fact, we could only reach the number of publications listed in the present review because we work as a multilingual team. However, we acknowledge that important publications could not be covered because we could not understand them.

Nevertheless, the studies we address here suggest that there is a need for specialized language training, including in the translation of technical manuals for aircraft maintenance engineers and students of technical translation. Regarding Bai's mention of the problematic lack of collaboration between the market (such as airlines and translation agencies) and researchers, further research is needed if we are to fully understand the best practices for teaching the translation of aircraft maintenance manuals as well as what kinds of pedagogical materials and activities are most effective for developing these skills.

3. The Corpus

Corpus linguistics (CL) employs machine-readable banks of texts of specific genres for linguistic research (McEnery and Wilson 2012). While the benefits of CL outstrip the scope of this paper, we concentrate here on linguistic investigation favoring glossary building (Tagnin 2015) and pedagogical materials design (Friginal et al. 2020; Friginal and Roberts 2022).

Tagnin (2015, p. 361) addresses the importance of considering the "translator's terminological needs," which go beyond equivalent terms. According to the author, the translator needs to have easy access not only to terms but also to their behavioral profile, that is, their collocates and phraseology, in order to produce a natural text in the target language. The solution is to equip translators of specific languages with corpora made up of texts originally produced in the technical domain. These texts need to be authentic, which means that they need to be produced in and for the technical area in question, in both the source and the target languages. Known as comparable corpora, this type of corpus contains original texts (i.e., not translated) in at least two languages.

In a similar vein, (Peixoto 2020b) employs CL tools to propose entries and terminological definitions for an aeronautical meteorology glossary that should be descriptive rather than prescriptive. She also recommends following the norms of the target public, which may be multiple and operating in a continuum from more to less specialized. We follow her approach while observing this continuum because we are dealing with different professionals with highly specific tasks that interweave into a common goal: flight safety. For example, when checking aircraft limitations caused by certain minor failures, pilots and maintenance technicians need to discuss together the requirements of the flight plan and the aircraft status to comply with these requirements. Both teams thus need to observe details of their own specialized area so as to make a decision over whether or not it is safe to fly.

3.1. The texts

In view of the critical genre analysis proposed by (Bhatia 2008; 2017; 2022) and the need for professional language to be used as a model rather than that of idealized native speakers (Bhatia 2022), we combined forces from different technical and higher education institutions to gather documentation employed in aviation, more especially by airlines. Our team is collecting documents such as Minimum Equipment Lists (MEL), Flight Operation Handbooks, Aircraft Maintenance Manuals (AMM), checklists (such as Tasks or Quick Reference Handbooks), logbooks, and reports. These documents are handled by a wide array of professionals, including maintenance engineers, ground staff, cargo loaders, flight dispatchers, and pilots working for airlines, maintenance centers, and airports.

As described in Terenzi (2021), part of this corpus has been used in diverse studies of aircraft courses held in English in Brazil. Terenzi (2020) describes some of these investigations, including two corpus-based studies focusing on the most common verbs used in aircraft maintenance manuals, one study of similarities and differences in the use of the words *aircraft, airplane, plane, jet, jetliner, airship, airplane*, and *airliner* while drawing attention to the contexts and genres in which they appear and to the development of technical glossaries containing names (clusters) of fasteners (*nuts, rivets, screws*, and *bolts*).

Another corpus that has been cleaned up and reorganized for the current project is described in Almeida and Prado (2011). The corpus informed the syllabus design of an online course in reading for aircraft maintenance personnel in a major airline in Brazil. Currently, this corpus is being used for guidance in the compilation of an English-Chinese comparable corpus as the texts also deal with business jets (such as the Cessna Citation X) and discontinued aircraft (such as the MD-11), even though some of the documents are outdated. We are not concerned with the age of the fleet per se because maintenance personnel may refer to those documents to repair and maintain older but still operating aircraft. Rather, it is the need for regular textual updates resulting from reports

of misreading, incidents, or recalls that deserve our attention. For this reason, we categorized the texts according to their language, genre, and finally aircraft type. This allows us to upload the folders to the software according to the task at hand.

The current composition of the corpus is as follows:

Table 1 Corpus of Aircraft Documentation

Sub-corpora	Text Type	Aircraft Type	# of texts	# of words	T/TR (%)
English	Aircraft Maintenance Manuals Aviation reports	B737 AMM	43	33,947,536	0.17
		B737 AOM	37		
		B737 SRM	8		
		B767 AMM	42		
		B767 AOM	70		
		B777 SRM	3		
		A319 AMM	69		
		A320 AMM	8		
		A320 SRM	1,194		
		A320 TSM	35		
		A330 AMM	37		
		E190 AMM	40		
		E190 AOM	472		
		MD11 AOM	47		
		MD11 Task cards	36		
		Cessna single engine AMM	1		
		Cessna jets AOM	26		
		Helicopter manuals	48	525,711	2.12
		Airworthiness Directives	61	183,297	2.42
		Safety reports	30	992,113	1.32
		Magazine articles	56	1,761,104	2.24
		Accident reports	352	2,260,878	1.17
		Handbooks (ab initio training/other kinds of aircraft)	9	1,095,723	1.81
Chinese	Aircraft Maintenance Manuals Flight operations	B787 MEL	1	7,656,680	1.98
		B737MAX MEL	1		
		B737 AMM	86		
		B737-6 Line maintenance manual	1		
		B737-8 Line maintenance manual	1		
		B737 AOM	167		
		A320 AOM	1		
		B737-NG Max FCTM	1		
		B737-8 QRH	1		
		B747 FCOM	1		
		B757 FCOM	1		
		B767 FCOM	1		
		B787 FCOM	1		
		A320 SOP	1		
		A330 FCTM	1		

Brazilian Portuguese	Accident reports Aviation Agency Regulatory documents	CENIPA Relatórios finais	111	221,865	4.10
		ANAC Diretriz de aeronavegabilidade	458	347,369	2.48
	Portuguese-English documents	ANAC Diretriz de aeronavegabilidade / Airworthiness Directives	414 381	342,665 274,798	2.15 2.01
		Relatorios de acidentes / Accident reports	8 8	68,231 58,595	7.96 7.22
Total			4370	49,736,565	

This list shows a comparable corpus defined as "a corpus with original texts in both languages" (Tagnin 2015, p. 361), with the proviso that the texts should belong to the same genre. These do not consist of translated texts but of originally written texts in the languages under study and with the same communicative function (Bowker and Pearson 2002). Comparable corpora are useful tools in glossary making due to the better understanding of the terms in the source texts they permit along with helping to develop a coherent and consistent translation for these into the target language. The use of comparable corpora is seen as highly beneficial in domain-specific translation processes by authors such as Bowker and Pearson (2002) and Tagnin (2015), among others

The only texts that might not fit the rule above are those listed in the bottom row, which are translated texts. However, they are not paralleled and the subcorpus does not always have corresponding texts. They represent the importance placed on communication that must come through, in the local language for domestic needs and in English for international exchanges. It can also be observed that this subcorpus holds the highest T/TR; however, the reason might be the small number of texts it consists of.

Regarding the texts, their sizes can be either a paragraph long (in the case of some airworthiness directives), or a thousand-page long (in the case of maintenance manuals). Consequently, we do not rely on number of texts to consider the representativeness of the corpus.

Challenges faced are mostly related to cleaning up the texts for trade secrets as documents of this kind constitute intellectual property. Thus we removed watermarks and any identification of aviation agencies or airline names. Some of the texts, particularly the corpus that informed the course described in Almeida and Prado (2011), were corrupted and needed repairing. Additionally, a major difficulty is the machine readability of manuals in Chinese. Since corpus linguistics software processes words by considering spaces before and after words as frontiers – something that does not occur in Chinese – we needed to use other software to segment the texts. The software we adopted is SegmentAnt (Anthony 2017). This is a user-friendly suite to which we simply upload the .txt file (or copy and paste the text in the input text box) and click on START. When the software completes the task, it creates a folder in the same location of the source text, adding the prefix "seg" (segmented) to the name of the file. The file retains a .txt extension, thus facilitating its use in most corpus linguistics tools available, including Antconc, a software suite also provided by Anthony (2022).

The input text, before segmentation, can be seen on the left side of Table 02; the output text, segmented, is on the right side of the same table:

Table 2 Samples of original and segmented texts

Unsegmented text	Segmented text
用途 空调系统控制飞机内部环境， 提供给机组、旅客和设备。	用途 空调 系统控制 飞机 内部 环境， 提供 给 机组、 旅客 和 设备。

This process is important for the detection and extraction of words. As can be seen from Table xx, since Chinese characters are juxtaposed, the software cannot identify how many words are in the line. However, automatic segmentation allows us to use the software and extract words, collocates, and keywords and determine the type/token ratio (T/TR), a statistical measure dividing the number of discrete words in the text by the number of types of words (*the*, for example, is counted as one type, while the total frequency in which it appears in the corpus is counted as tokens). This helps us detect lexical variation in the corpus; that is, "the closer the result is to 1 (or 100 if expressed as a percentage), the greater the vocabulary variation" and vice-versa (McEnery and Hardie 2012). Following Prado and Tosqui-Lucks (2019), we used this measure to guide the design of the corpus because the lower the result, the more repetitive and standardized the language in the texts.

3.2. Software and Reference Corpora

To process the corpus, we employ different software depending on the research we carry out. For the project conducted in China, we started using Antconc (Anthony 2022), but we also used Sketchengine (Kilgarriff et al. 2014) when the size of the corpus widely surpassed the reference corpora we had available, and Wordsmith Tools (Scott 2016) to determine the T/TR of the corpus. Additionally, our choice was oriented by the language under investigation, as Antconc automatically recognizes Chinese characters without any other input. However, Wordsmith Tools needs to be set to Chinese in order to process the characters.

Corpus tools employed in the investigation of the corpus start from the extraction of data, including a wordlist with the most frequent words ranked at the top of the list, lists of two-, three-, four-, or more word clusters, concordance lines demonstrating the surroundings of the word or cluster under scrutiny, and keyword extraction with the use of a reference corpus. Based on these data extractions, the researchers may select and further analyze patterns. Given that we are in the corpus building phase, we need to run constant trials in the corpus to check the validity of the data as well as the number of documents we still need. We also verify the T/TR of the corpus to determine whether we need more texts of the same variety or a greater variety of documents. For example, we found that we had enough aircraft operational manuals because the T/TR of this genre was low.

We used Antconc for most trials as the tool offers wordlists, clusters, concordance lines extractions, and T/TR analysis. Problems with Antconc started when our corpus became too big and we needed a larger reference corpus, which is when we then decided to use SketchEngine for its catalog of corpora. We did this because we are interested not only in technical terms but also in the presence of patterns typical of the genre, such as

discourse markers (Zuppardo 2013), modal verbs (Gabrielatos and Sarmento 2006), and noun phrases (Bocorny 2008). As these words or clusters are more common than in the reference corpus, they can be extracted when both corpora are contrasted.

3.3. Extracting Keywords and Collocates

We now address the importance of collaboration within and between teams. One of the teams is made up of a linguist and several English language undergraduate students learning how to use CL tools for lexicography and pedagogical purposes at a Sino-Foreign institution based in China. They have the ability to ask appropriate questions related to language and to identify linguistic patterns that answer those questions. Another team consists of a linguist and several undergraduate aircraft maintenance students at a federal institution based in Brazil. These students are learning to become mechanics, and because they are already exposed to the terminology of the area, they can validate the terms selected by the first team. The third team, still under formation, will consist of a linguist who teaches translation of aviation English at an aeronautical university in China and will be responsible for validating the English-Chinese translations proposed by the first team. We also have one linguist specialized in aviation English and CL, working in the aviation industry, who oversees the collaborations.

The purpose of extracting keywords is merely to facilitate the lexicographic work. When extracted automatically, wordlists reveal the most frequent words in the corpus. However, this may hide technical words that are significant to the area. As most words used in any text are function words, these are salient in any wordlist, pushing many content words down the list. Consequently, we use a reference corpus to provide a contrast with the corpus under study and then filter the words that are typical of the latter. Still, the software will simply calculate the statistical properties of those words that are more common in one corpus than in the other. We used this list to manually select those that we found to be more significant to our study, examining them in concordance lines and in the co-text (a widened-out view of the word or phrase in the source text). This selected list thus becomes a candidate list and is further assessed by another team.

After being examined by the other team, the list goes back to the team of language students who will now identify the behavioral profile of the term; that is, they will investigate clusters and collocates of the term in question. The search for collocates also helps pinpoint terms in the source language that do not demonstrate clear equivalence (Tagnin, 2015). Examples are the words *screw* and *bolt*, *smoke* and *fumes*, or *airplane* and *airliner*, which have equal or similar equivalents in Portuguese and often confuse Brazilian learners of English (Terenzi and Pizzi 2020). Li, Zhu, and Zhou (2018) also point out problems affecting aviation English words when they have a different meaning depending on which word they collocate with, such as *gear* in *landing gear* and *gearbox* or *generator* in *oxygen generator* and *AC generator*.

Ongoing analyses are carried out at every stage of the compilation to ensure the quality of the corpus. Given that the genres included in the texts may be repetitive, we need to make decisions regularly regarding documents to be added to the corpus as well as other aircraft types to orient the selection of manuals.

The next step is to use reference corpora in both languages to detect keywords in the corpora under study. As explained earlier, keywords will be selected as candidate terms and sent to different institutions. Once endorsed by the professional and technical teams,

these candidate terms will compose the visual dictionary (Prado and Terenzi 2022) and the trilingual dictionary.

While keyword extraction is not essential in the design of language learning activities specific to aircraft maintenance and operation, knowing what terms are key in the aviation industry is paramount to novice English teachers willing to work in the area. White (2018) refers to ASD-STE100 as a tool that equipped him for learning the terminology necessary to teach aircraft maintenance technicians to write reports. We strongly believe that combining ASD-STE100 with a corpus of aircraft documentation such as the one we are developing will afford the aviation community greater language awareness and thus improve safety standards.

4. Discussion and Conclusion

We have worked on corpus linguistics and aviation English individually since at least 2010. However, this is our first opportunity to be involved in a common project, combining forces and efforts to promote work collaboratively. Ideally, our final projects are a visual dictionary, a trilingual glossary, and pedagogical materials. However, it is the process described here that warrants most of our attention as researchers and teachers in higher education institutions. The process allows for collaboration between at least two countries, Brazil and China, in two different educational settings. On the one hand, enabling future language professionals to explore authentic resources with real needs fosters competences for future world markets, particularly in international settings. On the other hand, when teachers expand their pedagogical frontiers beyond the classroom, they foreground intercultural communicative competences and the use of English as an international language, while students' multilingual capabilities are enhanced and valued.

As to the corpus, the context of aircraft maintenance (be it engineers or language professionals such as translators) calls for teaching World Englishes for Specific Purposes (WESP, i.e. Bhatia 2022) and for the delineation of discourse community norms as well as differentiation in the performance of experts and novice rather than focusing on terms such as language learners and native speakers (Tarone 2005).

Because "[c]ommunication implies community and membership is mediated with the meaning of the text" (Widdowson 1998, p. 7), no language professional should stop at the word level and consider only decontextualized discrete units. While corpus linguistics studies take frequencies and therefore numbers as a starting point, they are not restrained by them. High frequencies indicate patterns prevalent in a given community, and linguists search for possible explanations of such patterns. A corpus of aircraft maintenance manuals, accident reports, and circulars (among other documents) may help visualize linguistic phenomena that would not be accessible to the naked eye. In turn, the phenomena under scrutiny can inform materials that can facilitate and improve the work of the professionals involved. Pedagogically, they can inform curriculum design, materials development, and assessment, while lexicography can enhance the quality of glossary and dictionary making.

Moreover, a corpus of this kind can boost future research in readability measures similar to Zhang et al (2020). With translated texts, we can also carry out analysis in line with Carvalho and Rebechi (2021), who checked the readability of both source and translated texts in the public health domain with a view to understanding whether these

texts succeeded in reaching a lay audience. However, in this study, we do not deal with a lay audience but with a variety of professionals from different settings, as suggested in Peixoto (2020a).

Furthermore, the corpus can also be employed in studies such as Drayton (2022), who combined two aviation genres documented standard phraseologies and real ATC communications to investigate how technical the vocabulary of radiotelephony is. In addition, this corpus can be contrasted with radio communications so as to verify the extent to which pilots borrow terms from aircraft manuals.

We have illustrated the first and perhaps most important phase of a collaboration between educational institutes in Brazil and China with the aim of not only training future professionals, but also informing better practices regarding aviation safety. We also call for future collaborations in educational and professional settings as a means of continuously developing projects that address real needs in the industry.

REFERENCES

Aerospace and Defence Industries Association of Europe. 2021. ASD-STE Simplified Technical English Specification STE-100. International specification for the preparation of technical documentation in a controlled language. Issue 8, April 2021. Brussels.

Almeida, Dilso and Malila Prado. 2011. “Desenvolvendo o conteúdo programático de um curso de inglês para mecânicos de aeronaves com base em um corpus DIY: um estudo de caso” [Developing the syllabus of an aircraft maintenance English course based on a DIY corpus: a case study]. *Aviation in Focus* 2 (2): 6–20.

Anthony, Lawrence. 2022. *Antconc* (Version 4.2.0) [computer software]. Tokyo: Waseda University. Available at: https://www.laurenceanthony.net/software/segmentant

白秀敏 [Bai, Xiumin]. 2023. “民航术语翻译研究的现状与展望:对象、内容与方法” [Present Status and Prospect of studies on civil aviation terminology: objects, contents and methods]. *China Terminology* 25(1): 74-87.

Bhatia, Vijay K. 2008. “Genre Analysis, ESP and Professional Practice.” *English for Specific Purposes* 27 (2): 161–74. https://doi.org/10.1016/j.esp.2007.07.005.

———. 2017. *Critical Genre Analysis: Investigating Interdiscursive Performance in Professional Practice*. New York: Routledge.

———. 2022. “World Englishes for Specific Purposes: A Multi-perspective View.” *World Englishes* 41 (4): 584–95. https://doi.org/10.1111/weng.12602.

Bocorny, Ana Eliza Pereira. 2008. “Descrição das unidades especializadas poliléxicas nominais no âmbito da aviação: Subsídios para o ensino de inglês para fins específicos.” [Description of noun phrases in the aviation field: subsidies to the teaching of ESP]. Doctoral Dissertation. Porto Alegre: Universidade Federal do Rio Grande do Sul.

———. 2011. “Panorama dos estudos sobre a linguagem da aviação.” *Revista Brasileira de Linguística Aplicada* 11 (4): 963–86. https://doi.org/10.1590/S1984-63982011000400009

Borowska, Anna P. 2017. *Avialinguistics: The Study of Language for Aviation Purposes*. Warschauer Studien zur Germanistik und zur Angewandten Linguistik, vol. 29. Frankfurt am Main: Peter Lang.

Bowker, Lynne and Jennifer Pearson (Eds.). 2002. *Working with Specialized Language*. London: Routledge.

Čarapić, Dragana. 2022. "Importance of Translation in Teaching and Learning English for Electrical and Mechanical Engineering: A Case Study." *Journal of Teaching English for Specific and Academic Purposes*, October, 261–76. https://doi.org/10.22190/JTESAP2202261C

Carvalho, Yuli Souza and Rozane Rodrigues Rebechi. 2021. "Inteligibilidade e convencionalidade em textos de divulgação da área médica em português brasileiro [Readability and Conventionality in Expository Texts in Brazilian Portuguese]." *Revista de Estudos da Linguagem* 29 (2): 959. https://doi.org/10.17851/2237-2083.29.2.959-998

Chatzi, Anna V., Wayne Martin, Paul Bates, and Patrick Murray. 2019. "The Unexplored Link between Communication and Trust in Aviation Maintenance Practice." *Aerospace* 6 (6): 66. https://doi.org/10.3390/aerospace6060066

Chervak, Steven, and Collin, Drury. 2003. "Effects of job instruction on maintenance task performance." *Occupational Ergonomics*, 3(2): 121–132.

Davies, Mark. 2004. British National Corpus (from Oxford University Press). Available online at http://english-corpora.org/bnc/.

Drury, Colin G., and Jiao Ma. 2003. "Do Language Barriers Result in Aviation Maintenance Errors?" In *Proceedings of the 47th Human Factors and Ergonomics Society Annual Meeting*.

Dupont, Gordon. 1997. "The Dirty Dozen Errors in Maintenance." In *Proceedings of the 11th Symposium on Human Factors in Maintenance and Inspection: Human Error in Aviation Maintenance*. Washington, DC.

Embryany, Fairistya R. 2020. "A Needs Analysis of English Learning for Aircraft Maintenance Students. Proceedings of the 1st International Conference on Lifelong Learning and Education for Sustainability (ICLLES 2019), 35-39. Dordrecht: Atlantis Press.

Friginal, Eric, Elizabeth Mathews, and Jennifer Roberts. 2020. *English in Global Aviation: Context, Research, and Pedagogy*. London:: Bloomsbury Academic.

Friginal, Eric and Jennifer Roberts. 2022. "Corpora for Materials Design." In *The Routledge Handbook of Corpora and English Language Teaching and Learning*, edited by Reka R. Jablonkai and Eniko Csomay, 131–46. London: Routledge.Gabrielatos, Costas and Simone Sarmento. 2006. "Central Modals in an Aviation Corpus: Frequency and Distribution." *Letras de Hoje* 41 (2): 215–40.

International Civil Aviation Organization. 2010. *Manual on the Implementation of ICAO Language Proficiency Requirements*, (2nd ed.). Montreal: International Civil Aviation Organization.

Kilgarriff, Adam, Vít. Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. "The Sketch Engine: Ten Years On." *Lexicography* 1: 7-36.

Li, Zhizhuo, Min Zhu, and Yali Zhou. 2018. "The Translator's Subjectivity in Aviation English Translation." *International Journal of Applied Linguistics and Translation* 4 (3): 46-51.

Liu, Hui. 2020. "The Interplay between Terminology and Translation: Taking Terminology Translation in the Field of Civil Aviation from English into Chinese as an Example."

Terminologia e Traduzione, Publifarum 33: 1–5. http://www.farum.it/publifarum/ezine_articles.php?id=506

Ma, Jiao, Colin G. Drury, and Clara V. Marin. 2009. "Language Error in Aviation Maintenance: Quantifying the Issues and Interventions in Four World Regions." *International Journal of Aviation Psychology* 20 (1): 25–47. https://doi.org/10.1080/10508410903416136.

Mathews, Elizabeth. 2020. "How Incomplete Language Standards Threaten Aviation." *Aviation Week*, March 3, 2020. https://aviationweek.com/air-transport/opinion-how-incomplete-language-standards-threaten-aviation?fbclid=IwAR262pq8-6McH6YkH7_1PvpsXOAHI0cogGfANrsd7TbU0N4pufWxqeDPSZI

McEnery, Tony and Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Niamsuwan, P M. 2017. "A Study of Needs Analysis of English Language for Aircraft Heavy Maintenance Mechanics." Master's Thesis, Thammasat University, Phra Nakhon.

Peixoto, Rafaela Araújo Jordão Rigaud. 2020a. "Nas asas da tradução: Elaboração de glossário de meteorologia aeronáutica." *Revista CBTecle* 2 (1): 563–82.

Peixoto, Rafaela Araújo Jordão Rigaud. 2020b. "Aeronautical Meteorology Glossary: A Discussion on Term Definition in the ANACpedia Termbase." *The Especialist* 41 (3): 00-00. https://doi.org/10.23925/2318-7115.2020v41i3a10

Prado, Malila Carvalho de Almeida. 2019. "A relevância da pragmática no ensino do inglês aeronáutico: um estudo baseado em corpora [The importance of pragmatics in the teaching of aviation English: a corpus-based study]." Doctoral Dissertation. São Paulo: Universidade de São Paulo. https://doi.org/10.11606/T.8.2019.tde-16122019-181408.

Prado, Malila, and Daniela Terenzi. 2022. "Development of aircraft maintenance glossaries in higher education: exchanging knowledge in international encounters." *XIV Encontro Nacional de Tradutores e VIII Encontro Internacional de Tradutores - ENTRAD.* Porto Alegre: UFRGS.

Prado, Malila, and Patricia Tosqui-Lucks. 2019. "Designing the Radiotelephony Plain English Corpus (RTPEC): A specialized spoken English language corpus towards a description of aeronautical communications in non-routine situations." *Research in Corpus Linguistics*, 113-128. doi:https://doi.org/10.32714/ricl.07.06.

Scott, Mike. 2016. *Wordsmith Tools, Version 7.* Stroud: Lexical Analysis Software.

Tagnin, Stella Esther Ortweiler. 2015. "Corpus-Driven Glossaries in Translator Training Courses." *Oslo Studies in Language* 7 (1): 00-00. https://doi.org/10.5617/osla.1447

Tarone, Elaine. 2005. "English for Specific Purposes and Interlanguage Pragmatics." In *Interlanguage Pragmatics: Exploring Institutional Talk*, edited by Kathleen Bardovi-Harlig and Beverly Hartford, 157–74. Mahwah, NJ: Erlbaum.

Terenzi, Daniela. 2014. "Princípios norteadores para o planejamento de cursos de línguas para propósitos específicos em curso superior tecnológico (manutenção de aeronaves): Considerando visões de aprendizes, instituição formadora e empregadores English translation]." Doctoral Dissertation, Universidade Federal de São Carlos.

Terenzi, Daniela. and Augusto-Navarro, Eliane Hércules. (2018) "Planejamento de cursos de língua para propósitos específicos (instrumental) em cursos superiores tecnológicos: reflexões e estabelecimento de princípios norteadores com base em curso de tecnologia de manutenção de aeronaves [English translation]" In *Planejamento de cursos de línguas:*

Traçando rotas, explorando caminhos, edited by Rita de Cássia Barbirato and Vera Lúcia Teixeira da Silva , 17–48. Campinas: Pontes.

Terenzi, Daniela. 2020. “A linguística de corpus como meio de envolvimento de alunos de curso tecnológico com pesquisas na área de língua inglesa.” Simpósio Internacional de Ensino de Língua, Literatura e Interculturalidade (SIELLI) 2020, Goiás..

Terenzi, Daniela. 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 (10): 1122–34. https://doi.org/10.4236/ojapps.2021.1110084

Terenzi, Daniela and Maria Claudia Bontempi Pizzi. 2020. “The Relevance of Linguistic and Professional Knowledge as Contributing Factors for Technical Translation in the Area of Aircraft Maintenance.” *The Especialist* 41 (3): 00-00. https://doi.org/10.23925/2318-7115.2020v41i3a9

Tosqui-Lucks, Patricia and Malila Prado. 2021. "Corpora de inglês aeronáutico: desafios para o estudo da área e proposta de trabalho conjunto." *Tradterm* 37 (1): 148-174. https://doi.org/10.11606/issn.2317-9511.v37p148-174

Trimble, Louis. 1985. *English for Science and Technology: A Discourse Approach*. Cambridge: Cambridge University Press.

文军 [Wen Jun] and 李培甲 [Peijia Li]. 2011. “航空航天英语术语翻译研究 [A study on the translation of Aerospace English Terminology].” *广东外语外贸大学学报 [Journal of Guangdong University of Foreign Studies]* 03: 27-31+42.

White, Alan Robert. 2018. "A review of an aircraft maintenance related English writing course." *CreTech 2018 6th International Conference on Creative Technology 24th-26th July*. Bangkok: Rajamangala University of Technology Krungthep. 186-190.

Widdowson, Henry. 1998. “Communication and community: the pragmatics of ESP.” *English for Specific Purposes* 17 (1): 3-14.

Zhang, Genggeng, Eric Friginal, Elizabeth Mathews, and Jennifer Roberts. 2020. “The Language of Aviation: Corpus-Based Analysis of Aviation Discourse.” In *English in Global Aviation: Context, Research, and Pedagogy*, edited by Eric Friginal and Jennifer Roberts, 111–50. London: Bloomsbury Academic.

Zhao, Lizhu. 2014. “Application of Small Corpora in ESP Teaching: A Case Study of Aeronautical English Teaching.” *Journal of Beijing University of Aeronautics and Astronautics* (Social Sciences Edition) 2: 89-92.

Zuppardo, Maria Carolina. 2013. “A linguagem da aviação: um estudo de manuais aeronáuticos baseado na Análise Multidimensional.” *ReVEL* 11(21): 5-25. [www.revel.inf.br].