

Designing the *Radiotelephony Plain English Corpus* (RTPEC): A specialized spoken English language corpus towards a description of aeronautical communications in non-routine situations

Malila Prado - Patricia Tosqui-Lucks
Universidade de São Paulo / Brazil

Abstract – Pilots and air traffic controllers need to undergo a specific English test in order to be granted a license for international operations. A language proficiency scale was developed to serve as a parameter to all aviation regulatory agencies throughout the world by targeting the language produced specifically by air traffic controllers and pilots in radio communications when non-routine situations (such as technical problems, bird strike, changes in weather, health problems on board, etc.) occur (ICAO 2010). However, there is a lack of empirical investigation which could shed light upon this particular register helping the users of the scale with its understanding. In an attempt to fill this gap, this paper outlines a compilation of the *Radiotelephony Plain English Corpus* (RPTEC), a spoken corpus of aeronautical communication consisting of transcriptions of exchanges between pilots and air traffic controllers in non-routine situations for research and pedagogical purposes. By presenting steps taken during the process, we intend to provide fellow researchers with data which may suit other purposes and yield further analyses, as well as enlighten similar investigations in the field of English for Specific Purposes.

Keywords – corpus design; spoken corpus; aviation language; English for Specific Purposes

1. INTRODUCTION¹

Pilots and air traffic controllers (ATCOs) around the world are required to take a proficiency test which evaluates their ability to communicate in English on the radio in order to operate internationally. Such is the decision of the International Civil Aviation Organization (ICAO) after a series of accidents in which language was a contributory cause. By means of a document which prescribes procedures for both training in and testing of aviation English, ICAO (2010) determines that English proficiency of these

¹ This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

professionals is to be assessed according to six different criteria: pronunciation, structure, vocabulary, fluency, comprehension and interaction. Test takers need a minimum level 4 in all criteria to be granted the English proficiency license to operate internationally.

The lack of clarity in the description of aviation English *per se* is claimed to have offered different views of understanding (Alderson 2010). According to ICAO (2010), the scope of teaching and assessment is both standard aeronautical phraseology –a set of phrases and words that are fully described in official documents (Philps 1991)– and plain English, which is the language used when phraseology is not sufficient (ICAO 2010: x). The latter is a more spontaneous portion of language needed mostly when pilots and air traffic controllers face abnormal situations, that is, when something unexpected happens, since phraseology covers all routine situations of a flight. Despite being spontaneous, the plain English used in radio communications does not resemble informal spoken English, as this scope of language follows certain rules governed by aeronautical phraseology, such as being structurally concise and lexically restricted (Bieswanger 2016). Although routine communications should occur based on the standard aeronautical phraseology, the professionals often refer to plain English even in normal situations, particularly with pragmatic speech acts such as greeting or thanking one another (Lopez 2013).

Standard aeronautical phraseology (SAP) in English is documented by ICAO in different manuals (2001, 2007) as mandatory for international flights,² and it is usually learned and trained ever since pilots and ATCOs are in the initial process of their career. This scripted language focuses on precise and predetermined lexicon and reduced grammatical items with a view to eliminating ambiguity (Philps 1991; Moder 2013; Bieswargen 2016; Estival *et al.* 2016). The large number of aircraft under the control of the same professionals at the same time urges the adoption of specific call signs for each aircraft, which are airline and flight numbers (such as *GLO 323*, *Fastair 345*, *Speedbird 981*), instead of using personal pronouns (*I*, *we*, *you*). Likewise, the ATCO must be identified by the ground control facility (like Miami Center, Brasilia Control, GRU Tower) to avoid confusion. Communications should take place as illustrated in extract (1).

² According to ICAO documents, the standard phraseology must be used, whenever possible, in all phases of flight: for clearance and taxi, take-off and departure, read-back, climb, cruise and descent, approach and landing.

(1) **ATCO:** Fastair 345 when passing flight level 80 contact Alexander control 129.1

Pilot: When passing flight level 80 129.1 Fastair 345

The example above starts with an instruction given by the ATCO and is read back by the aircraft, whose call sign is *Fastair 345*. The use of call signs also prevents the ambiguity of deixis (see Garcia 2016). To further illustrate the difference between SAP and plain aviation English (see Bieswargen 2016), the latter being the object of this piece of research, example (2) presents the transcription of a real event.

(2) **Pilot:** Control / Aircraft 1010 / report //

ATCO: Go ahead //

Pilot: Aircraft 1010 / we have an indication of engine fire / procedures have been applied / we have no further uh confirming indications of engine fire / we are now recovering as a precautionary measure / I'll keep you posted // and uh souls on board uh seventy-nine passengers plus four crew / correction plus five crew //

ATCO: seventy-nine passengers plus five crew / roger / thank you // Aircraft 1010 / at this moment your position is uh eight miles on the right downwind // you are the number one for the approach now expect vectors for the ILS //

Pilot: We'll keep this heading for a while / and uh we will can perform a normal circuit for runway zero three / Aircraft 1010 //

ATCO: Aircraft 1010 // roger / can you confirm the engine with the problem? //

Pilot: It's engine number two / number two //

ATCO: Number two / roger //

Extract (2) illustrates the use of commands such as *report*, *go ahead*, *expect vectors for the ILS* and concise expressions such as *souls on board*, *correction* and *roger*, occurring alongside some elements, such as *I'll keep you posted*, *We'll keep this heading for a while*, *thank you*, as well as modal verbs, which can be easily found in a more spontaneous talk. These are examples of plain aviation English. The purpose of our study is not to describe how speakers use SAP, but the manner in which their communication happens when it exceeds the limits of this prescribed and documented language. ICAO's documents state that even in communications that are held when the unexpected happens, pilots and ATCOs should adhere to the principles of the SAP, which are objectivity, brevity and clarity.

There is considerable debate as to whether plain aviation English and SAP can be separated (Lopez 2013: 118). Such debate is based on the fact that both of them are interconnected, and it is difficult to establish the point at which SAP ends and plain

aviation English begins. The present research adopts the view that “standardized phraseology and plain aviation English can be categorized as two distinct specialized registers” (Bieswanger 2016: 74). By considering them as two concurrent registers, we selected parts of communications between pilots and controllers undergoing abnormal or emergency situations, in which SAP may be present, but most of the excerpt represents plain aviation English.

In order to understand how plain aviation English is used in this specific context, we have undertaken the compilation of a corpus of radio communications between pilots and ATCOs in non-routine situations, namely the *Radiotelephony Plain English Corpus* (RTPEC), which is presented in the following sections. The corpus was designed to meet both research and pedagogical purposes and, for this reason, it contemplates certain characteristics required for both areas.

This paper is structured as follows: firstly, Section 2 introduces the theoretical basis of corpus linguistics as an approach to the description of language. Then, Section 3 discusses some of the criteria involved in the corpus design, considering categories, speakers, texts, the transcription model and the current amount of data. The paper concludes with some suggestions for future research.

2. CORPUS LINGUISTICS

Corpus linguistics is a research method which employs corpora for data extraction. A linguistic corpus is a bank of texts stored in computers, which allow for a (semi-) automatic extraction of data by using statistical analysis. Spoken texts must be transcribed as the computer software commonly used in such investigations is fed by the written word.³

Studies carried out with this methodology very often highlight evidence usually not perceived by the naked eye (Sinclair 2004). For this reason, a number of features have been brought out by such studies and have particularly favored fields related to dictionary-making, translation studies and, to a lesser extent, coursebooks, to mention but a few. A perspective underlying studies based on corpus linguistics is that language is stored

³ We are currently not aware of any tool that recognizes the spoken word without its corresponding written form.

cognitively and retrieved in chunks, rather than in single words, and built through social use, therefore conventionalized (Wray and Perkins 2000; Wood 2006).

Conventionalized language is revealed by the high frequencies with which certain elements occur in a given linguistic community. The more often they occur, the more conventionalized they are. Such frequencies can only be identified through corpora, and their relevance –particularly when it comes to corpus design– relies on the context from which the language was taken. Therefore, the origin of the texts along with their context of production is of utmost importance. The texts must represent the use of the language in the context of production being investigated. Some writers advocate for as faithful a representation of the spoken language as possible, even though it is questionable whether the written form can actually cover all the non-verbal aspects present in real interaction (Haberland 2010).

The corpus should encompass the linguistic community from which the texts are extracted. In recent years corpora have become larger as computer capabilities have increased (the 560-million-word *American Corpus of Contemporary English*⁴ is an example). However, some authors defend the use of smaller corpora for specific context investigations, such as studies in the context of Language for Specific Purposes (Gavioli 2005) or Pragmatics (Vaughan and Clancy 2013), as few occurrences may be enough to represent the language characteristics of these groups of speakers. That is the case of aviation English, due to the fact that the language is concise, the vocabulary is restricted and the context is very limited. Aviation English itself has witnessed studies in corpora mainly with the objective of describing language use; some examples can be found in Sarmento (2008), Bocorny (2008), Lopez (2013) and Tosqui-Lucks (2018).⁵

3. THE COMPILATION PROCESS

It is commonly stated that spoken corpora are challenging for many different reasons: the compilation itself, copyright issues, the choice for spontaneous or guided language, the quality of video and audio equipment and the transcriptions (Adolphs and Knight 2010).

⁴ <https://corpus.byu.edu/coca/>

⁵ Tosqui-Lucks (2018) described 16 studies, from different parts of the world, from the 1990s to the present date, all dealing with aviation English and Corpus Linguistics.

These issues were faced in this piece of research as well. In this section, we elaborate on each of the choices made throughout the process.

3.1. *The criteria*

According to Bieswanger (2016: 74), “[w]hile standardised phraseology is concerned with the fairly restricted aspects of routine air traffic control issues, plain Aviation English covers a broader range of topics in non-routine situations, such as emergencies as well as other unusual or unexpected contexts.” Considering this, the first decision regarding our corpus was to select audios in which operational problems occur, these being of human, weather or mechanical nature, insofar as the oral communication happens between ATCOs and pilots. Our first endeavor was the compilation of transcripts of accident reports; however, we soon realized that these transcripts consider mostly the communication happening inside the flight deck, thus not corresponding to the scope of language being studied. Moreover, as prosody might have relevance in the analysis, all transcriptions should necessarily have their audio files. Finally, the selection of the events should be focused on solvable problems whenever possible, as the corpus texts are intended for students of aviation English.

To this end, we resorted to a well-known website in the area of aviation English called Live Air Traffic,⁶ which stores communications held in different parts of the world, especially the United States. Events representing non-routine situations are rare, and the overwhelming amount of data being stored 24 hours a day at different airports worldwide hindered our search. One of the sections of this website solves the issue of finding communications with technical problems by providing a bank of audios of different kinds. Our goal was then communications of non-routine matters, but they still needed to be carefully selected. After collecting the audios, we verified their source in different websites: newspapers, TV news, accident/incident databases and *The Aviation Herald*.⁷

The time period of the communications selected was also taken into account. As there was no control over the identification of the speakers, we chose to collect files of events which happened from 2008 onwards, the year when the ICAO proficiency

⁶ <https://www.liveatc.net>

⁷ A web 2.0 model site built with the participation of users and members (<https://www.avHerald.com>).

requirements referred to in Section 1 of this paper should be implemented by all state members.

During the process of beta-testing of the data compiled, we came across some problems. The first problem was related to the two most frequent content words in the corpus: *runway* and *engine*. The word *runway* corroborated aviation investigations attesting that most accidents happen on or near the runway.⁸ However, *engine* could mean an excess of audio files related only to engine problems, which was verified by searching for the main abnormal situation represented in each of the audios. We then noticed that there were only events with engine malfunctions, which allowed for few potential findings related to other terms. Thus, in order to increase the representativeness of the corpus, we tried to widen the variety of problems by choosing sources of events in different data: annual accident and incident reports, manufacturers' statistics and governments' reports; however, occurrences vary depending on the countries (some hold better safety records than others), manufacturers (as their aircraft systems are different) and time (aviation has become safer over time).

Any of these choices would be random. We thus decided to use a document entitled *Taxonomy of Occurrences* (ICAO 2006), composed of a script of categories which standardize accident and incident reports internationally as a means of diffusing information. There were other possible choices, but adhering to an official text which needs to be used when abnormal situations occur seemed to be a better way of approaching our selection standards. This document enumerates 33 different categories of occurrences in aviation, from minor to major incidents or accidents,⁹ such as bird strike, fire/smoke, fuel related, collision with obstacles, system/component failure and malfunction, among others. Based on this taxonomy, the initial idea was to compile at least three audios for each category.

By assuring that the audios met the criteria listed in the taxonomy, we started to obtain better results in terms of lexical range. The corpus was finalized when it reached 12 hours of audio material transcribed into 110,737 words, in a total of 130 texts about 31 different occurrence categories. The transcriptions are monolingual and non-annotated, as we intended to start from a corpus-driven study (Tognini-Bonelli 2001).

⁸ See <https://www.boeing.com/news/techissues/pdf/statsum.pdf> (20 December, 2018).

⁹ Broadly speaking, incidents are minor occurrences whose consequences are avoided, and accidents refer to damage to aircraft or injuries to people or even fatalities.

3.2. Corpus design

In our approach we argue in favor of aviation English as a lingua franca (Estival *et al.* 2016), as English is the language of aviation, and we assume that native and non-native speakers in aviation use the language in a similar manner. Previous studies have also shown that even native speakers of the language need to adhere to the aviation English standards, as aviation English is not a register acquired in daily life events (Bieswanger 2016). The objective is to compile a corpus of operational users of a vocational language, regardless of these users' mother tongue.

In a latter phase of the research, we suspected that the corpus was biased towards the North-American culture rather than representing patterns of an international community, since most audios collected were originally held in the United States. There was a need to implement the corpus with a more international variety of speakers. Nevertheless, with the continuous growth of international flights, it has become easier to find exchanges among different nationalities. Still, it is impossible to determine if the speakers are native speakers of English, bilinguals or even non-native speakers. Some airline companies are widely known for the high presence of international pilots in their payroll, especially those in the Middle East and Asia. Besides, the English proficiency of the transactions compiled has clearly improved in the late years (see Prado and Tosqui-Lucks 2017), which might be a result of the implementation of the proficiency requirements. This enhancement seems to point out that aviation English is becoming more widely used by an international community.

Seeking a broader representation of the professional community, we opted to add at least one more audio file to each of the categories suggested by the *Taxonomy of Occurrences* (ICAO 2006), inasmuch as it contained exchanges between international traffic, that is, a foreign aircraft in a particular airport or airspace. Whenever an airline company is considered foreign to a certain airport or airspace, radio communications are held in English even if the pilots and ATCOs share the same mother tongue. Besides, native speakers of English must allegedly accommodate their English to an international speech community (Estival *et al.* 2016) provided they are not in their local environment.

3.3. The transcription

Restricted to the scope of problems, the transcription starts when the problem is first mentioned, and finishes when the problem is either solved or handed out to another professional (i.e. a fire fighter or maintenance technician). For this reason, the size of the texts differs (some are less than a minute long and others are longer than 15 minutes).

By using the *Taxonomy of Occurrences* as a basis, we included the audio file names, time of each audio and the final count of words to organize the corpus. Each occurrence has a number of four to six audios to embrace as large a number of lexical items as possible. We also inform about the duration of each file, followed by the number of words transcribed and the final count of words for each category.

All the corpus texts were manually transcribed, as the sound quality, number of speakers and different accents hinder the use of any speech recognition software, which needs to be trained to a specific voice to enable the automatic transcription. The minimum number of participants in radio communications is two, but there is often a larger number when other aircraft are involved in the event. To assure control of features of the transcription which needed to be strictly connected to the context of production, the decision was to concentrate the transcription on one linguist researcher. The transcriptions were then reviewed by linguists specialized in aviation English and subject-matter experts (SMEs), namely pilots or ATCOs with international experience. Whenever possible, pilots and ATCOs from the region of a specific occurrence were requested to review the transcription, as they are supposedly more familiar with the waypoints, departure and arrival points, airport names and call signs present in the audio files.

Of great assistance during the transcription was the software *SoundScriber*,¹⁰ used by the University of Michigan in corpus compilation. It allows the transcriber to control the audio file without leaving the text processor. By means of shortcuts in the keyboard, the transcriber saves time by playing, pausing and rewinding the audio file; the loop, one of the most resourceful tools, replays a time segment continuously, moving ahead little by little, adjusted by the transcriber according to his/her typing speed (see Figure 1).

¹⁰ See <http://www-personal.umich.edu/~ebreck/code/sscriber/>

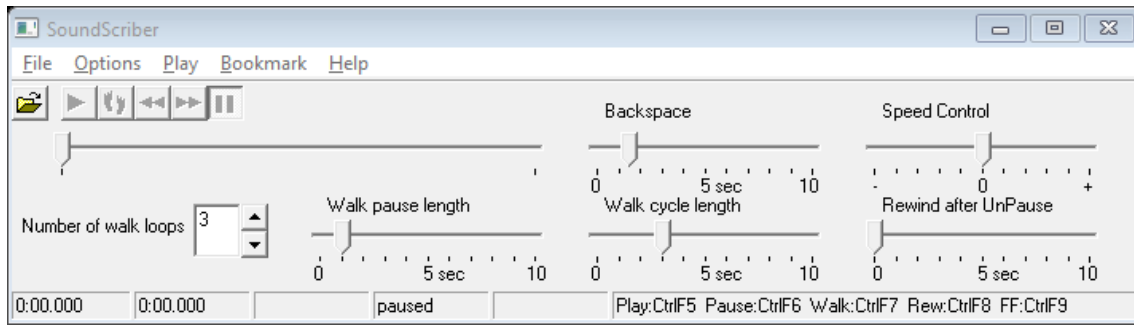


Figure 1: *SoundScriber* main screen

The spoken language is not governed by the same rules as the written language. Pauses and hesitations, for example, do not correspond to the syntactic rules of comma usage, and neither do breaks or false starts. The attempt to transform spoken language in a representation which can allow for linguistic investigations is a political issue (Haberland 2010: 62) for two main reasons: (i) we take the product from its source of production and it immediately becomes abstract; and (ii) a transcription is not the event itself, but a representation of the event. Still, the transcription needs to be linked to its source of production (such as speakers, event, place, time), as a means of preserving the analysis and readability of the texts.

During the first transcriptions, we faced problems as illustrated in extract (3).

- (3) **ATCO:** <aircraft call sign> while you're waiting uh just the reason for the uh wait is the tail strike may have been in the intersection just waiting for the inspection to ensure that the runway is serviceable.

The utterance in (3) is lengthy, and it is difficult to recognize where the pauses are. Adding commas would not represent the particularity of such an utterance, and it is also problematic to try to identify where punctuation marks may fit. In order to sort out this problem, we adopted a prosodic and pragmatic framework called the 'Language into Act Theory' (L-AcT) (Cresti 2000). Based on how the utterance was produced, the prosodic breaks are identified by adding slashes separating the tone units, as can be seen in extract (4).

- (4) **ATCO:** <aircraft call sign> while you're waiting uh / just the reason for the uh wait is the tail strike may have been in the intersection // just waiting for the inspection to ensure that the runway is serviceable //

One slash (/) represents a break in the flow within the utterance; two slashes (//) indicate the end of the utterance, identified in the intonation. However, the end of the utterance does not necessarily mean the end of the turn, which might be held by the speaker. Because it is an oral production, this language is better approached as utterances rather

than sentences, which was facilitated by using the L-AcT. This choice would favor an investigation of the so-called conversation or spoken grammar (Rühlemann 2008), claimed to be different from the written form.

As the corpus was also built for pedagogic purposes to students of aviation English, the texts were preserved as clean of metalinguistic information as possible. Besides, the texts should be readable by research tools, allowing for a more clear-cut investigation. Separating tone units by lines, for example, would hamper the reading of the concordance lines, as the utterances would appear altogether. The slashes enable manual extraction in the sense that the units are visibly separated, even with segments corresponding to false starts or corrections, for example.

The first part of the text is the context of production, which contains the source of the audio file and any other database where a report of the event might be found, along with keywords to identify the problem and the professionals who have proof-read the transcription. This information goes between the tags <header> and </header>, identifying beginning and end of the setting (see Appendix 1 for a sample text).

Other tags used in the transcriptions to mark metalinguistic information include the following:

- <unreadable> refers to parts which none of the transcribers could understand, either because they were poorly pronounced or were waypoints specific of the location.
- <false start> identifies utterances that were initiated and then rephrased, and <break> identifies a break in the flow. They help the reader in the organization of certain disfluency features one might find in spoken language.
- <blocked transmission> means there was an overlap, which produces noise when two speakers hold the radio push-to-talk button at the same time.
- <pause> is used for silent pauses.
- <noreply> is introduced whenever one of the participants does not provide any feedback for the input previously given. This silence is part of the conversation, as it elicits the previous speaker to use strategies so as to finally obtain the response needed.

Some verbal aspects were also transcribed, even though they are not words *per se*, such as disfluencies: *uh* (for filled pauses), *uhh* (for longer filled pauses) and other noises. Numbers were transcribed in full for investigations of standard vs. non-standard phraseology, and words such as *okay* and *alright* were transcribed as such in order to avoid confusions with the letters *O* and *K* (for *okay*) and with the words *all* and *right*.

We opted to remove call signs of aircraft and ground stations when we printed out the transcriptions in paper publications or classroom handouts for the sake of preserving the identity.

To assess the representativeness of the corpus, there were constant measures of the type-token ratio (TTR) of the corpus, that is, the amount of running words divided by the number of types. Token refers to the total occurrences of a word, whereas type refers to the number of different words. This calculation reveals how dense or complex a corpus is; the higher the TTR, the richer the corpus. Corpus Linguistics tools such as *Wordsmith Tools* (Scott 2016) can quickly assist the researcher with calculating such complexity. The corpus of this study has shown to be of low lexical density, as its TTR oscillated between 6% and 3.68% in its first 20 transcriptions. Summing up, with the current 110,737 words, the final TTR has reached 2.9%; such a low density indicates that the corpus has highly repetitive patterns.

4. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The description presented in this paper aimed at providing a new contribution which brings together research projects with a shared interest.

The potential uses of the corpus are manifold, among others:

- (i) Investigations on the use of the language by specific local communities;
- (ii) a focus on the negotiation of meaning or other strategies that pilots and ATCOs use when trying to solve problems collaboratively;
- (iii) an examination of the use of language by native and non-native speakers;
- (iv) an analysis of aspects which are more intrinsic to language or metalinguistics such as a study of the linguistic areas prescribed in the ICAO scale, among others.

The present corpus has already rendered an investigation of the most frequent lexico-grammar patterns, aiming at a comparison with the structure and vocabulary areas listed in the ICAO scale. It has also been used for the analysis of fluency and interaction (other two linguistic areas prescribed by the ICAO scale). However, such studies are not exhaustive, and so far have only presented language to be prioritized in the aviation English classroom.

The corpus can be enhanced as an ongoing process to bring about investigations of different features. In our specific case, we seek a description that yields an empirical understanding of aviation English to support curriculum design and material development in order to provide tools for the field of teaching aviation English. Nevertheless, we do not exclude the benefits which this corpus can bring to the understanding and elaboration of tests, as it can provide test designers with materials to try to approach meaningful interaction with the test candidates.

Nonetheless, there are some caveats which need to be addressed in future research; one of them is related to the pronunciation features which are not yet linked to the written data. Despite the existence of software that enables the paralleling of the audio and written data, it is far time-consuming, and at present we lack the time or funding to do so. Still, a possible alignment considering utterances would generate broader understanding of pronunciation matters.

It should be emphasized that this corpus represents a very specific scope of an already specific field: aviation English. This corpus could be incorporated into a larger project containing other corpora collected from other parts of the world or even involving other aviation professionals.

The investigations with which corpora equip the researcher and material designer offer the opportunity to bring authentic language to the classroom, supplying the language professional with adequate tools to teach from meaningful input and, as such, promote better output from the students. We hope to contribute to the field of aviation English with this research.

REFERENCES

- Adolphs, Svenja and Dawn Knight. 2010. Building a spoken corpus: What are the basics? In Anne O’Keeffe and Michael McCarthy eds. *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 38–52.

- Alderson, Charles. 2010. A survey of aviation English tests. *Language Testing* 27/1: 51–72.
- Bieswanger, Markus. 2016. Aviation English: Two distinct specialised registers? In Christoph Schubert and Christina Sanchez-Stockhammer eds. *Variational Text Linguistics: Revisiting Register in English*. Berlin: Mouton de Gruyter, 67–85.
- Bocorny, Ana. 2008. *Descrição das Unidades Especializadas Poliléxicas Nominais no Âmbito da Aviação: Subsídios para o Ensino de Inglês para Fins Específicos (ESP)*. Porto Alegre: Universidade Federal do Rio Grande do Sul dissertation.
- Cresti, Emanuela. 2000. *Corpus di Italiano Parlato (Vol. I)*. Firenze: Accademia della Crusca.
- Estival, Dominique, Candace Farris and Brett Molesworth. 2016. *Aviation English: A Lingua Franca for Pilots and Air Traffic Controllers*. London: Routledge.
- Gavioli, Laura. 2005. *Exploring Corpora for ESP Learning*. Amsterdam: John Benjamins.
- Garcia, Angela. 2016. Air traffic communications in routine and emergency contexts: A case study of Flight 1549 ‘miracle on the Hudson’. *Journal of Pragmatics* 106: 57–71.
- Haberland, Hartmut. 2010. Pragmatics as a component vs. pragmatics as a perspective of linguistics. *Studies in Pragmatics* 12: 54–68.
- International Civil Aviation Organization (ICAO). 2001. *Annex 10 to the Convention on International Civil Aviation: Aeronautical Telecommunications*. Montreal: International Civil Aviation Organization.
- International Civil Aviation Organization (ICAO). 2006. *Aviation Occurrence Categories: Definitions and Usage Notes*. Montreal: International Civil Aviation Organization.
- International Civil Aviation Organization (ICAO). 2007. *Manual of Radiotelephony DOC 9432-AN/925*. Montreal: International Civil Aviation Organization.
- International Civil Aviation Organization (ICAO). 2010. *Manual of Implementation of the Language Proficiency Requirements (DOC9835-AN/453)* (second edition). Montreal: International Civil Aviation Organization.
- Lopez, Stephanie. 2013. *Norme(s) et Usage(s) Langagiers: Le Cas des Communications Pilote-contrôleur en Anglais*. Toulouse: Université Toulouse le Mirail dissertation.
- Moder, C. Lynn. 2013. Aviation English. In Brian Paltridge and Sue Starfield eds. *The Handbook of English for Specific Purposes*. West Sussex: Wiley-Blackwell, 227–242.
- Philps, Dennis. 1991. Linguistic security in the syntactic structures of air traffic control English. *English World-Wide* 12/1: 103–124.
- Prado, Malila C. A. and Patricia Tosqui-Lucks. 2017. Are the LPRs focusing on real life communications issues? *International Civil Aviation English Association*. Dubrovnik: Embry-Riddle Scholarly Commons, 1–20. <https://commons.erau.edu/cgi/viewcontent.cgi?article=1027&context=icaea-workshop> (22 October, 2019.)
- Rühlemann, Christoph. 2008. A register approach to teaching conversation: Farewell to standard English? *Applied Linguistics* 29/4: 672–693.
- Sarmento, Simone. 2008. *O Uso dos Verbos Modais em Manuais de Aviação em Inglês: Um Estudo Baseado em Corpus*. Porto Alegre: Universidade Federal do Rio Grande do Sul dissertation.
- Scott, Michael. 2016. *Wordsmith Tools version 7*. Liverpool: Lexical Analysis Software.
- Sinclair, John. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tosqui-Lucks, Patricia. 2018. Aplicações de corpora no ensino e na avaliação de inglês aeronáutico: Estado da arte, reflexões, direcionamentos. In Matilde Scaramucci, Patricia Tosqui-Lucks and Silvia M. Damiano eds. *Pesquisas sobre Inglês Aeronáutico no Brasil*. Campinas: Pontes, 89–114.
- Vaughan, Elaine Claire and Brian Clancy. 2013. Small corpora and pragmatics. In Jesús Romero-Trillo ed. *Yearbook of Corpus Linguistics and Pragmatics*. Dordrecht: Springer, 53–73.
- Wood, David. 2006. Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review* 63/1:13–33.
- Wray, Alison and Michael R. Perkins. 2000. The functions of formulaic language. *Language and Communication* 20: 1–28.

Corresponding author

Malila Prado

Av. Prof. Luciano Gualberto, 403

3º andar – sala 14

Cidade Universitária- Butantã

05508-010 São Paulo - Brasil

e-mail: malilaprado@usp.br

received: August 2019

accepted: October 2019

Appendix 1: Sample text

<header>

Ukraine Airlines B737 x KLM F-70 at Brussels

Sep 11 2016

<http://avherald.com/h?article=49de7586&opt=0>

Venna

</header>

One two six six two five / Aircraft eight three kilo echo / good evening //

Aircraft International one four six / stop immediately / I say again / stop immediately //

Aircraft International one four six / hold position // Aircraft one seven two five / go around / I say again / go around / immediate right turn heading zero one zero //

Going around on heading zero one zero / Aircraft one seven two five //

Aircraft one seven two five / climb three thousand feet //

Say again / one seven two five? //

Aircraft one seven two five / climb three thousand feet //

Climbing three thousand / one seven two five //

Aircraft seven eight Quebec tango / line up two five right //

Line up two five right / Aircraft seven eight Quebec tango //

Aircraft eight echo x-ray / established ILS two five left //

Aircraft nine correction Aircraft eight echo x-ray / City Tower / hello / continue approach two five left / number 2 / wind one five zero degrees four knots //

Continue / eight echo x-ray //

Aircraft one seven two five / contact arrival again one one eight decimal two five //

One eight two five / Aircraft one seven two five/ speak to you in a minute or two //

Aircraft seven eight Quebec tango / wind one two zero degrees two knots / two five right / cleared for take-off //

Cleared for take-off two five right / Aircraft seven eight Quebec tango //

Aircraft International one four six / contact departure one two six decimal six two five / goodbye //

One two six six two five departure / thank you / Aircraft International one four six //