



**HAL**  
open science

## Linguistic Analysis of English Phraseology and Plain Language in Air-Ground Communication

Stéphanie Lopez, Anne Condamines, Amélie Josselin-Leray, Mike O'Donoghue, Rupert Salmon

► **To cite this version:**

Stéphanie Lopez, Anne Condamines, Amélie Josselin-Leray, Mike O'Donoghue, Rupert Salmon. Linguistic Analysis of English Phraseology and Plain Language in Air-Ground Communication. *Journal of Air Transport Studies*, 2013, 4 (1), pp.44-60. halshs-00924821

**HAL Id: halshs-00924821**

**<https://shs.hal.science/halshs-00924821>**

Submitted on 7 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Linguistic Analysis of English Phraseology and Plain Language in Air-Ground Communications**

Stéphanie Lopez<sup>(1,2)</sup>, Anne Condamines<sup>(1)</sup>, Amélie Josselin-Leray<sup>(1)</sup>, Mike O'Donoghue<sup>(2)</sup> & Rupert Salmon<sup>(2)</sup>

<sup>(1)</sup> CLLE-ERSS & University of Toulouse-le-Mirail  
Maison de la Recherche, University of Toulouse-le-Mirail  
5, allées Antonio Machado  
F-31058 Toulouse Cedex 9, France

<sup>(2)</sup> ENAC, École Nationale de l'Aviation Civile  
7, avenue Edouard Belin  
BP 54005  
F-31055 Toulouse Cedex 4, France

Corresponding author:  
Stéphanie Lopez  
ENAC  
7, avenue Edouard Belin, BP 54005  
F-31055 Toulouse Cedex 4, France  
Phone: +33(0)562 174 348  
E-mail: [Stephanie.lopez@enac.fr](mailto:Stephanie.lopez@enac.fr)

Stéphanie Lopez is a Ph.D. candidate in English Linguistics at the University of Toulouse-le-Mirail. Her doctoral research project consists of a linguistic analysis of pilot-controller (or air-ground) communications in English.

Pr. Condamines is a senior research linguist in the CLLE-ERSS Institute at the University of Toulouse-le-Mirail where she specialises in semantics, terminology and specialized corpora.

Dr. Josselin-Leray is an assistant professor in the English Department of the University of Toulouse-le-Mirail. Her fields of interest range from corpus linguistics to lexicography.

Mike O'Donoghue is the head of the Languages division at ENAC, the French Civil Aviation University. He is also a Board member of the International Civil Aviation English Association.

Rupert Salmon is an English teacher in the Languages division at ENAC where he is also in charge of the pedagogical materials.

# Linguistic Analysis of English Phraseology and Plain Language in Air-Ground Communications

## Abstract

As the ENAC<sup>1</sup> organises English training for France's air traffic controllers and must comply with ICAO<sup>2</sup> language proficiency requirements, it has specific needs in terms of English radiotelephony teaching. Consequently, it has initiated, in collaboration with CLLE-ERSS<sup>3</sup>, a study of the usage of English by French controllers with international pilots. The aim of this project is to describe the different uses of English phraseology and plain language within pilot-controller (or air-ground) communications via a comparative study between two corpora: one representing the prescribed norm and made up of examples of English from two phraseology manuals; the other consisting of the orthographic transcription of recordings of real air-ground communications. The comparative study presented here is conducted at a lexical level. It focuses on the discrepancies observed in the distribution of the corpora lexicon. In the long run, the results from the contrastive study can be used to improve English radiotelephony teaching.

**Keywords:** Air Traffic Control, Language for Specific Purposes (LSP), Corpora, Linguistic Comparative Analysis, Norm, Usage

## 1. Introduction

In some professional contexts, accomplishing a very specific task can entirely depend on verbal communication between experts of a given field: being able to communicate is for these experts a necessity for sharing and transferring the specialised knowledge required to fulfil their job. When these communication-dependent situations are recurrent enough, linguistic norms can be created by institutions and authorities, who then enforce them. The aim of these linguistic norms is usually to create less ambiguous communication thanks to simplified rules (at a syntactic, lexical and semantic level for instance). The linguistic normalisation also enables the various interlocutors to minimise their linguistic and cognitive efforts in carrying out the task at hand thanks to their shared knowledge (Falzon, 1986). The use of natural language, on the other hand, would not be efficient enough to express this common knowledge and could easily lead to rough estimations, misunderstandings and incomprehensions (Vergely, 2008).

The domain of air traffic control offers an instructive example of such an established linguistic norm: that of *phraseology*, the specialised language used by pilots and controllers to conduct what is intended to be unambiguous and effective radiotelephony communications. One should actually talk about *phraseologies* since civil aviation uses six official languages<sup>4</sup> in which phraseologies are employed. It is generally in English – used as a *lingua franca*

---

<sup>1</sup> “École Nationale de l’Aviation Civile” is the French Civil Aviation University : <http://www.enac.fr/en/>

<sup>2</sup> The International Civil Aviation Organisation is the United Nations’ institution in charge of developing the standards behind the normalisation of international civil air transport: <http://www2.icao.int/en/home/default.aspx>

<sup>3</sup> “Cognition, Langues, Langage, Ergonomie - Équipe de Recherche en Syntaxe et Sémantique” is a French linguistics research institute : <http://w3.erss.univ-tlse2.fr/>

<sup>4</sup> The six official languages of civil aviation are English, French, Spanish, Russian, Arabic and Chinese.

(Crystal, 2003; Seidlhofer, 2005) – that international flights are dealt with: it allows dialogue between a controller and a pilot who do not necessarily share the same first language. For instance, an aircraft flying in French controlled airspace can receive control services in French or in English, depending on the pilot's first language. The ICAO's Annex 10 volume 2 (2001) explicitly confirms the function of English as the common language of aeronautical aviation:

Air-ground radiotelephony communications shall be conducted in the language normally used by the station on the ground or in the English language (5.2.1.2.1).

The English language shall be available, on request from any aircraft station, at all stations on the ground serving designated airports and routes used by international air services (5.2.1.2.2).

English phraseology and the different uses made of it are at the core of our study, conducted within Lopez's doctoral research project. This project has been initiated by the French Civil Aviation University (ENAC), in collaboration with the linguistics institute CLLE-ERSS, in order to try and meet some of the ENAC's specific needs in terms of English radiotelephony teaching<sup>5</sup>. The aim of this research project is to draw up a panorama of the different types of usages made of the English language by French controllers and pilots from all over the world in radiotelephony communications and bring their differences and similarities to light. The method of analysis consists of a comparative study between two corpora (see section 4): one representing the prescribed norm and the other representing the real usages made of it.

In this paper, we aim at presenting to what extent some usages of English by pilots and controllers in real air-ground communications can differ from the prescribed norm by the presence of markers of a subjective individual speaker. To do so, we first introduce the specialised languages used in radiotelephony (sections 2 & 3). We then present the two corpora under study (section 4). Finally, we introduce various comparisons between these two corpora as well as some preliminary results (section 5).

## 2. English Phraseology

In air traffic control, air-ground communication is mainly performed using a specialised or operative<sup>6</sup> language known as *phraseology*. It was created and has been continually updated by the International Civil Aviation Organisation to cover the most common and ordinary situations encountered in air navigation in order to optimise and ensure safety in radiotelephony: "the purpose of phraseologies is to provide clear, concise, unambiguous language to communicate messages of a routine nature" (ICAO, 2010: 1.1.3). Phraseology and the messages that employ it are therefore subject to simplified but strict syntactic, lexical, semantic and phonetic rules. The following examples, extracted from our reference corpus (see section 4), give an idea of what phraseology looks like:

(a) *P: golf charlie delta, request **Right turn** when airborne.*<sup>7</sup>

---

<sup>5</sup> The ENAC is in charge of the English training for France's air traffic controllers and pilots and has therefore to comply with ICAO language proficiency requirements.

<sup>6</sup> We use the same term as Falzon (1986) prefers it to "specialised language" to refer to languages shaped by the type of knowledge peculiar to a specific activity, i.e. by "operative knowledge".

<sup>7</sup> Messages beginning with "P:" correspond to pilots' messages while those introduced by "C:" correspond to controllers' messages.

*C: golf charlie delta, Right turn approved, runway 0 6 cleared for take-off.*

*P: runway 0 6 cleared for take-off, Right turn, golf charlie delta.*

- (b) *C: Citron Air 3 2 4 5, multidirectional departure runway 2 8, at 800 feet turn Right heading 3 1 0, climb 3000 feet QNH.*

*P: multidirectional departure runway 2 8, at 800 feet turning Right heading 3 1 0, climb 3000 feet QNH, Citron Air 3 2 4 5.*

- (c) *P: Blagnac Tower, good morning, foxtrot bravo x-ray.*

*C: foxtrot bravo x-ray, good morning, pass your message.*

*P: foxtrot bravo golf bravo x-ray, PA28, VFR from Albi to Blagnac for touch-and-go, Agen next, 1500 feet, echo time 1 0 0 5, with information India. Requesting joining instructions.*

*C: foxtrot bravo x-ray, roger, report echo.*

*P: will report echo, foxtrot bravo x-ray.*

Phraseology's specific and very particular characteristics – which make it obscure for everyone but experts – have been previously described as (DGAC, 2007; Mell, 1992; Philps, 1989, 1991; Rubenbauer, 2009):

- The omnipresence of the imperative form in the controller's messages (due to his role as an administrator who provides pilots with manoeuvre instructions and authorisations):  
e.g. “*turn Right*” and “*climb 3000 feet*” in example (b) above, “*report echo*” in (c), etc. – rather than “*we would like to turn*”, “*you should climb*” or “*could you report*”, etc.
- The rarity of the interrogative and negative forms.
- The almost complete absence of modals.
- The deletion of determiners:  
e.g. “*request Ø Right turn*” in (a), “*Ø heading 3 1 0*” in (b), etc. – rather than “*I request a Right turn*” or “*the/your heading is 310*”.
- The deletion of subject pronouns:  
e.g. “*Ø request Right turn*” in (a), “*Ø turning Right*” in (b), “*Ø will report*” in (c), etc. – rather than “*I request*”, “*we are turning*” or “*we/I will report*”, etc.
- The deletion of prepositions:  
e.g. “*departure Ø runway 2 8*” and “*climb Ø 3000 feet*” in (b), etc. – rather than, “*departure from runway 28*” or “*climb to 3000 feet*”, etc.
- The deletion of auxiliaries *be* and *have* in [*be + past participle*] forms, [*be + -ing*] forms and [*have + past participle*] forms:  
e.g. “*Right turn Ø approved*” and “*Ø cleared for take-off*” in (a), “*Ø turning Right*” in (b), etc. – rather than “*Right turn is approved*”, “*you are cleared for take-off*” or “*we are turning Right*”, etc.
- The nominalisation of concepts:  
e.g. “*Right turn*” in (a), “*multidirectional departure*” in (b), etc. – rather than “*you should turn Right*” or “*you will follow the multidirectional route*”, etc.
- A highly specialised, univocal and finite lexicon (less than 1000 different words):  
e.g. “*QNH*” in (b), “*VFR*” and “*touch-and-go*” in (c), etc.
- An alphabet proper to the aeronautical domain:  
e.g. “*golf charlie delta*” in (a), “*foxtrot bravo x-ray*” and “*information India*” in (c), etc. – rather than “*GCD*”, “*FBX*” or “*information F*”.
- The specific spelling and pronunciation of numbers:  
e.g. “*runway 0 6*” in (a), “*Citron Air 3 2 4 5*” (with “3” pronounced as “tree”) in (b), etc. – rather than “*runway 6*” (without “0”) or “*Citron Air 3245*” (with “3” pronounced as “3”).

Phraseology's syntactic, lexical and semantic characteristics make it the essential communication tool for the transmission of the fundamental information required for providing optimal and safe guidance of air traffic. However, since it has been created to cover only a limited number of air navigation situations, phraseology is a limited tool:

While ICAO standardized phraseology has been developed to cover many circumstances, it cannot address all pilot and controller communication needs. It is widely acknowledged by operational and linguistic experts that no set of standardized phraseologies can fully describe all possible circumstances and responses (ICAO, 2010: 1.2.3).

Thus, when facing situations for which phraseology does not exist, pilots and controllers must resort to a more natural language known as “plain language”.

### 3. Plain Language

Pilots and controllers' communication needs in situations for which phraseology is not enough requires the usage of natural language – though constrained by phraseology's rules of clarity, preciseness and concision (Mell, 1992: 73). This form of natural language is referred to by the ICAO as “plain language” and is prescribed as a last resort when phraseology has reached its limits:

ICAO standardized phraseology shall be used in all situations for which it has been specified. Only when standardized phraseology cannot serve an intended transmission, plain language shall be used (2001: 5.1.1.1).

ICAO standardized phraseology should always be used in the first instance (2010: 4.3.3).

The transition from an operative language, such as phraseology, to natural language in unusual situations is accounted for by Falzon (1986: 37) by the absence of procedure patterns in such situations which leads operators to use a more powerful but not specialised representation tool, i.e. natural language. Unlike natural language, prescribed linguistic norms leave indeed no room for creativity. According to the ICAO, natural language – and the creativity that it implies, particularly when dealing with an unexpected turn of events – is the best instrument for human interaction:

Linguistic research now makes it clear that there is no form of speech more suitable for human communication than **natural language**. [...] Human language is characterized, in part, by its ability to create new meanings and to use words in novel contexts. This creative function of language is especially useful in accommodating the complex and unpredictable nature of human interaction, including in the context of aviation communications. There is simply no more suitable form of speech for human interactions than **natural languages** (2010: 1.3.2).

Nonetheless, the terminology chosen by the ICAO to refer to the language used when phraseology does not exist is “plain language”, not “natural language”. One could then assume that plain language and natural language are not alike: plain language should not be considered as natural language since it is supposed to comply with phraseology's standards. It has indeed been recently officially defined as such by the ICAO:

Plain language in aeronautical radiotelephony communications means the spontaneous, creative and non-coded use of a given natural language, although constrained by the functions and topics (aviation and non-aviation) that are required by aeronautical radiotelephony communications, as well as by specific safety-critical requirements for intelligibility, directness, appropriacy, non-ambiguity and concision (2010: 3.3.14).

Plain Language can thus be considered as the spontaneous, creative and non-coded use of a given natural language within the context of the very specific domain of air traffic control. Yet, professional context is not enough to avoid the presence of linguistic difficulties, such as polysemy or impreciseness, which, while harmless in every day communications, could lead to serious consequences in professional contexts due to a lack of correctly transferred information (Condamines, 2008). In this context, can plain language really be considered as sharing phraseology's characteristics of clarity, preciseness and concision? Furthermore, the linguistic difficulties related to the use of plain language are acknowledged by the ICAO:

The features of plain language, [...], can be far from plain and present a challenge to listening skills. They include the use of a wider vocabulary referring (often with less precision) to domains and topics outside the aviation area (medicine, military organizations, etc.), references to complex notions such as hypothesis (we may divert), indirectness (we would like a request) and, under stressful conditions, much longer and less organized sentences (2010: 3.3.16).

The notion of plain language, as defined and presented by the ICAO, is far from clear for civil aviation professionals in charge of English radiotelephony teaching. Consequently, in order to determine with greater clarity what constitutes plain language in air-ground communications, an observation of the different usages of English by French controllers and pilots from around the world by means of a comparative study between two corpora was initiated.

#### **4.Presentation of the Two Corpora Under Study**

A comparative study between a reference corpus (henceforth referred to as *RefC*), representing the prescribed norm, and a corpus representing the real usages made of it (henceforth referred to as *UseC*) is essential to the identification, description and categorisation of the different real usages made of radiotelephony English. Two such corpora had thus to be compiled. But before introducing the two corpora under study, we should first define the notion of *corpus* in linguistics: it is a large collection of texts or utterances gathered in electronic form according to a specific organisation and set of criteria in order to serve as a data-base for linguistic descriptions and analyses (Bowker & Pearson, 2002; Sinclair, 1991).

The first step in the compilation of *RefC* was to select official texts from which representative samples of standard phraseology could be extracted. This type of texts being quite rare, the examples in English from two phraseology manuals – one edited by the ICAO (2007) and the other by the French government (DGAC, 2007) – have been selected to constitute this corpus. By choosing those two phraseology manuals, we aim at representing the norm from an international as well as national point of view.

The second corpus consists of the orthographic transcription<sup>8</sup> of about twenty-two hours of recording of real air-ground communications from two French En-route control centres and

---

<sup>8</sup> A specific transcription protocol has been created and applied and the different transcriptions have been reviewed by air traffic control experts.

one French major airport<sup>9</sup>. This three centres have been chosen to ensure that the corpus is representative of the language used in every day radiotelephony<sup>10</sup>.

The first corpus, RefC, is constituted of a total of 11,844 word tokens and 805 word types<sup>11</sup> while the second corpus, UseC, contains 49,020 tokens and 1238 types, as illustrated in table 1:

	Reference Corpus (RefC)			Real Usages Corpus (UseC)			
	ICAO Manual	DGAC Manual	<b>Total</b>	Centre 1	Centre 2	Centre 3	<b>Total</b>
Tokens	5304	6540	<b>11,844</b>	15,498	11,127	22,395	<b>49,020</b>
Types	526	625	<b>805</b>	699	541	796	<b>1238</b>

Table 1: Number of word types and tokens in each sub-corpus

We should specify here that the total number of word types in each corpus – 805 for RefC and 1238 for UseC – does not correspond to the sum of the word types contained in each of their respective sub-corpora as the latter share some common word types. For instance, the word “will” is found in both manuals constituting RefC. One interesting thing to notice is that, in spite of the total occurrences in each corpus, the number of different word types they contain remains rather low. This results from the fact that the phraseology lexicon is finite, as mentioned earlier: the number of different word types used is limited.

The observation of the different uses of English by controllers and pilots initiated by the ENAC is conducted through a comparative analysis of these two corpora.

## 5. Comparing the Two Corpora

Phraseology’s specific features concern several linguistic levels: the lexical level, with a highly specialised lexicon; the semantic level, with univocal meanings; the syntactical level, with very specific sentence structures; and the phonetic level, with the standardised pronunciation of certain words. A detailed comparative analysis between our two corpora at each of these linguistic levels should be dealt with in Lopez’s thesis in order to point out the differences and similarities found between the prescribed norm and the real uses made of it. However, for lack of space, this paper only focuses on some of the lexical features of the two corpora. The various observations and comparisons of the data are made possible by the use of a processing tool known as a concordancer, which, among other things, allows one to know exactly how many times a word type is used and to have access to the contexts in which every occurrence of a word is used.

---

<sup>9</sup> To collect these communications, an official authorisation was needed beforehand as in France this type of data is not accessible to the general public. For reasons of anonymity, the names and locations of these three centres will not be revealed in this paper. They have been chosen for the concentration of English used on their frequencies as well as their interest for our research project.

<sup>10</sup> Different types of air traffic control (aerodrome, approach and en route), different control stations, time slots and interlocutors have been taken into account to constitute *UseC*.

<sup>11</sup> In a corpus, each different word is known as a “type” (or “word type”). For instance, “will” and “would” are two different word *types*. The number of time a given word type occurs in a corpus is known as “token” (or “word token”). For instance, 56 *tokens* of the type “will” are found in RefC. In other words, “will” occurs 56 times.

## 5.1. Preliminary Methodology

The first preliminary step in comparing the vocabulary of the two corpora was to draw up a list of the different word forms they contain. Yet, from a lexical point of view, comparing a corpus made up of written data – and thus including no feature of verbal communication – with one made up of spoken data would not guarantee satisfactory results. Consequently, in order to obtain a well-balanced comparison of the lexicon found in the corpora, not all the different word types have been taken into account in our lexical analysis. The different categories of word types that have been excluded and the reasons for their removal are presented in the following table:

Excluded word types	Related Corpus	Examples	Reasons for Exclusion
Speech disfluencies <sup>12</sup>	UseC	<i>-huh-; we tr/ try;</i> etc	RefC does not contain any speech disfluency.
Politeness and greeting markers in languages other than English <sup>13</sup>	UseC	<i>arrivederci;</i> <i>merci beaucoup;</i> <i>konichiwa; hasta luego;</i> etc.	RefC is only constituted of examples in English.
Alphabet letters	Both corpora	<i>alpha; bravo;</i> <i>charlie,</i> etc.	The comparison of alphabet letters is not relevant for our study.
Proper Nouns <sup>14</sup>	Both corpora	<i>Air Citron; Albi;</i> <i>Airbus;</i> <i>Castelnaudary;</i> <i>Georgetown;</i> <i>Fastair;</i> etc.	Proper nouns cannot really be compared with one another as different proper nouns are found in the two corpora.
Hapaxes <sup>15</sup>	UseC	<i>actually; big;</i> <i>careful; east;</i> <i>reason;</i> <i>whatever;</i> etc.	Since they occur only once, these word forms cannot be considered as representative of the language used.

Table 2: Types of word forms excluded from our lexical comparison

By choosing not to take into account the word forms mentioned here, we aim at focusing on specific and recurrent air traffic control vocabulary as well as proceeding to a well-balanced comparison of the two corpora lexicon. The two corpora henceforth contain fewer word types and tokens: RefC is now constituted of 7185 tokens and 566 types while UseC contains 24,479 tokens and 408 types.

The second preliminary step in comparing the corpora lexicon was to classify the different word types left for the analysis according to their grammatical categories. Such a classification was made manually since the particular syntactic structures of phraseology do not allow a correct automatic tagging<sup>16</sup> of the corpora. The results of this classification show

<sup>12</sup> Speech disfluencies are typical features of spoken language. They include, among other things, cut-off words, repeated words or syllables and fillers such as *huh*.

<sup>13</sup> They depend on the interlocutors' creativity.

<sup>14</sup> They correspond to authentic or imaginary names of towns, airports, aircraft, beacons, etc. Only the proper nouns corresponding to different control stations on the ground have not been excluded from our analysis.

<sup>15</sup> Hapaxes are words which occur only once in a corpus.

<sup>16</sup> A tagged corpus contains word forms to which a grammatical tag has been applied.

that nouns are the part of speech most commonly found in both corpora (47.24% for RefC and 34.29% for UseC), followed by verbs (21.54% for RefC and 23.87% for UseC) and prepositions (11.59% for RefC and 10.76% for UseC). The other grammatical categories, i.e. adjectives, adverbs, conjunctions, determiners, interjections and pronouns, are present to a lesser extent (less than 8%). Some discrepancies have been observed in the distribution of several categories between one corpus and the other.

## 5.2. Discrepancies between the Two Corpora

The classification performed on the corpora lexicon reveals striking differences in the distribution of some grammatical categories between the two corpora: the nouns, adjectives, interjections, determiners and pronouns are unevenly distributed in RefC and UseC, as illustrated in Figure 1. This discrepancy in distribution can be seen as a reflection of the difference existing between the specific features of the prescribed norm (represented by RefC) and the uses made of it (represented by UseC). For some of the grammatical categories, we could go even further and consider them as preliminary clues to the potential differences in the characteristics of phraseology and plain language.

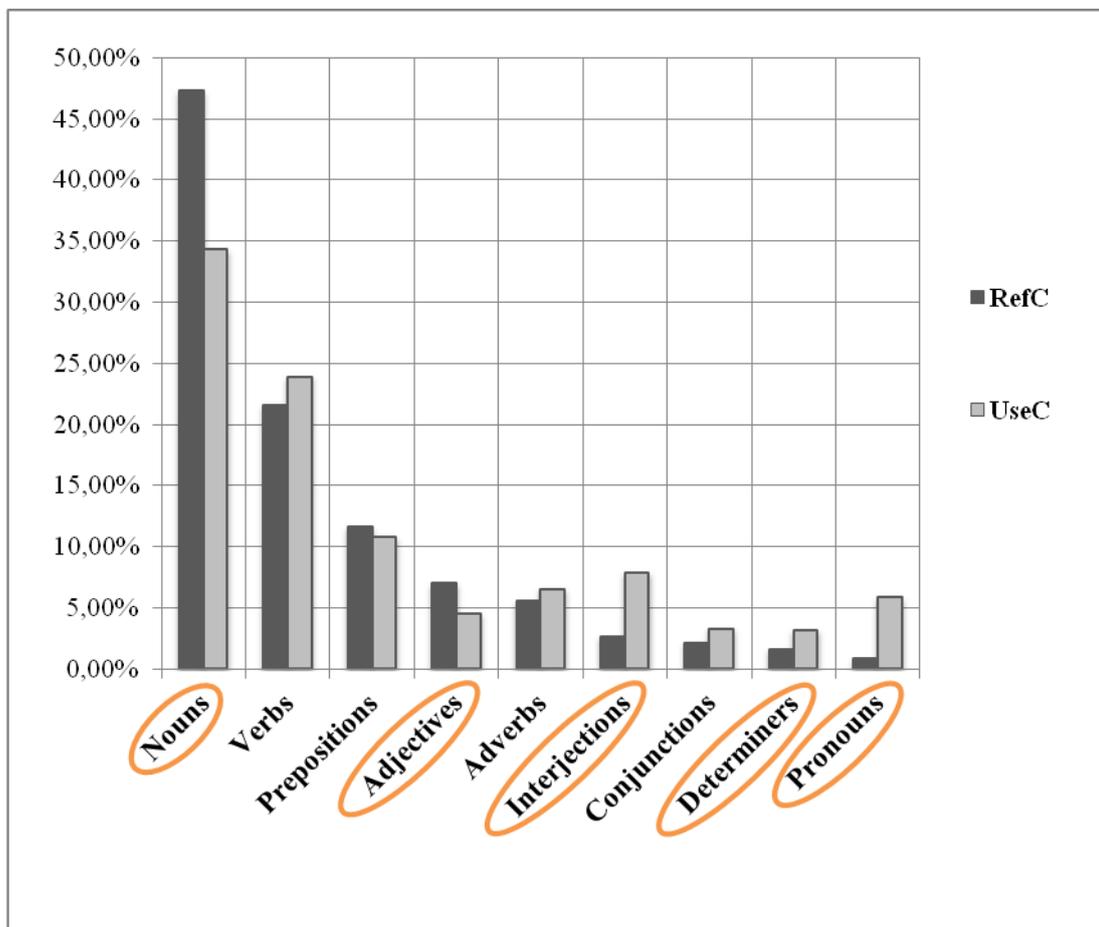


Figure 1: Distribution of the grammatical categories in the corpora

A detailed observation of the word forms contained in these unevenly distributed grammatical categories will help us to give a complete description of the lexical differences and similarities existing between the two corpora in the future. In this paper, we only discuss some of the word forms contained in the noun, interjection and pronoun categories.

### 5.3. Possible Comparisons between the Distribution of Some Word Forms

#### a. The Noun Category

The noun category is the most frequent category in both corpora: it accounts for 47.24% of all the tokens in RefC and for 34.29% of all the tokens in UseC. RefC and UseC contain respectively 301 and 147 noun word forms and have 95 noun forms in common, that is to say 26.84% of all noun forms. In other words, RefC contains 207 noun forms that are not present in UseC and UseC contains 52 noun forms are not present in RefC.

The three nouns used most often in RefC are “runway” (8.84% of all its noun tokens), “level” (7.6%) and “flight” (4.15%) while in UseC, the three most used are “level” (18.74% of all its noun tokens), “flight” (11.76%) and “heading” (6.04%). All the other nouns account respectively in RefC and UseC for less than 4% and less than 6% of all noun tokens. One interesting thing to mention is that the term “flight level” occurs only in one of the two manuals constituting RefC: no occurrence of “flight level” has been found in the French manual. Yet, if this manual took into account the extensive use of “flight level” by pilots and controllers in air-ground communications (61.54% of all “level” tokens in UseC), it would then reflect much better how phraseology and its standards are employed in real everyday radiotelephony.

Now, if we take a closer look at the noun forms that are specific to the real usage corpus (UseC), we can notice that all of them account for less than 1.4% of all its noun tokens, with only the three most frequent ones accounting for more than 1%. These three top noun forms are “sir”, “course”, and “Radar”<sup>17</sup>. Out of the 52 noun forms specific to UseC, up to 29 can be considered as not exclusively belonging to the air traffic domain. The word forms “sir”, “problem”, “madam”, “moment”, “afternoon”, “mountain(s)”, “question”, “best”, “help”, etc. indeed belong to a more general area. These noun word forms reflect a part of the lexicon needed by pilots and controllers to answer their communication needs that are not fulfilled by phraseology: they are everyday words used within radiotelephony communications.

The 207 noun forms specific to RefC account for less than 1.3% of all its noun tokens. 49 of them (16.50%) can be considered as specific to the domain of air traffic, such as, “helicopter”, “touch-and-go”, “transponder”, “airfield”, “aerodrome”, “airway”, “pilot”, “midrunway”, including 22 acronyms among which “ATIS” (Automatic Terminal Information Service), “CTOT” (Calculated Take-Off Time), “IFR” (Instrument Flight Rules), “NDB” (Non-Directional Beacons), “FIR” (Flight Information Region), “GNSS” (Global Navigation Satellite System), “RVSM” (Reduced Vertical Separation Minima), “VASIs” (Visual Approach Slope Indicators), “VMC” (Visual Meteorological Conditions) and “VFR” (Visual Flight Rules). These 207 noun word forms could undoubtedly be encountered in real air-ground communications: it is only by chance that they are not found in UseC: the specific air traffic situations in which these noun forms are generally used were not encountered while recording the communications constituting UseC.

---

<sup>17</sup> “Radar” refers here to a control station on the ground.

## **b. The Interjection Category**

According to the *Oxford Dictionary of English Grammar* (Chalker & Weiner, 1994) an interjection is a “minor word-class whose members are outside normal clause structure, having no syntactical connection with other words [...]”. We have thus decided to tag as “interjections” all the word forms which corresponded to this definition as well as those labelled as such in various English dictionaries.

RefC and UseC are thus respectively constituted of 2.63% and 7.84% of interjections. RefC contains 189 interjection tokens distributed in 10 different word types while UseC comprises 1918 interjection tokens for 26 different word types. The two corpora share 8 identical interjection forms. The main interjection forms in RefC are “Roger” (35.98% of all its interjection tokens), “wilco<sup>18</sup>” (14.29%) and “negative” (11.11%). These word forms are less frequent in UseC: “roger” accounts for 4.48% of all UseC interjection tokens while “wilco” accounts for 0.78% and “negative” for 1.15%. The three interjection word types used the most in UseC are the farewell and politeness markers “bye” (35.87%), “goodbye” (11.42%) and “thank you” (8.76%).

If we take a closer look at this type of marker, we can notice that they are not completely absent in RefC: “good morning”, “good day” and “thank you” are indeed part of this corpus. However, they are only to be found in the French manual and no greeting or politeness marker is used in the ICAO manual. Yet, the ICAO recommends, as part of the communicative functions of aeronautical radiotelephony communications, that users be able to understand and use markers referring to different attitudes such as politeness (2010: 3.4.9). According to Rubenbauer (2009: 72) expressions of courtesy can indeed “often be heard to facilitate the flow of information between participants in ATC or intra-cockpit communication”.

Greeting, farewell and politeness markers represent more than 61% of all UseC interjection tokens and involve up to 16 different word forms such as “hello”, “good morning”, “good afternoon”, “good evening”, “good day”, “bye”, “good bye”, “welcome”, “thank you”, “thanks” and “please”. The use of such markers is explained by Nübold and Turner (1983: 51; quoted in Rubenbauer, 2009: 27) by the fact that “the requirement to use English with the prescribed procedures is interfered with a constant, unremitting need which pulls the language into the opposite direction; by the human being’s desire to use language in its social and affective roles”. The quite extensive use of interjections and courtesy expressions in UseC could indeed be explained by the speakers’ prevailing need to customise and “humanise” air-ground communications and their perpetual repetitive tasks.

## **c. The Pronoun Category**

While pronouns are nearly absent from RefC (0.81% of all its tokens), they account for 5.82% of UseC. The 5 different pronoun forms found in RefC are “you” (65.52% of all its pronoun tokens), “I” (20.69%), “one” (8.62%), “me” (3.45%) and “what” (1.72%). On the

---

<sup>18</sup> The term “wilco” is used in radiotelephony as an abbreviation of “we will comply with”. We have chosen to consider it as an interjection since it is generally used outside normal clause structure and has no syntactical connection with other words.

contrary, UseC comprises 19 different pronoun forms of which the most used ones are “you” (44.28%), “we” (23.02%), “I” (9.19%), “it” (7.37%) and “that” (6.81%). All the other pronouns found in UseC account for less than 2.5% of all its pronoun tokens. The pronoun “we”, which is not at all present in RefC, is mainly used by pilots in UseC: 94.51% of all the 328 occurrences of “we” are in pilots’ messages. Controllers generally use the pronoun “I” rather than “we”. However, 16 occurrences of “we” in controllers’ messages can be found in UseC. It seems that some controllers tend to use the plural pronoun in situations for which they cannot provide pilots with what they want or need, as if trying to remind their interlocutors that the situation in which they are is not really up to the controller on frequency, and that a much more complex system is behind the provided control services. The pronoun “we” is also used by controllers to refer to themselves as a team as in France, *two* controllers deal with all the aircraft of a specific sector, even though only one of them is in contact with the pilots: they share the different air traffic control tasks the way two pilots share the tasks relating to the flight of an airplane. Some of the occurrences of the pronoun “we” in controllers’ messages are presented below:

- (d) *P: [...], any chance for higher level?*  
*C: [...], we call you back -huh- soon for climb if possible.*  
*P: thank you.*
- (e) *P: (right) so, we are flight level 3 4 0 on course to BOKNO, -huh- with the CBs<sup>19</sup> in sight, -huh- // we request a final 3 6 0 if possible.*  
*C: okay, we try to get higher for you, I call you back.*
- (f) *P: yes, [...] 5 0 5 8, requesting flight level 3 8 0, light turbulence.*  
*C: okay, 5 0 5 8, we tr/ we try // but -huh- it was impossible in the previous minutes, we try again.*  
*P: okay, that was 3 6, we're trying 3 8.*  
*C: yes sir, I know that but we try.*  
*C: [...] 5 0 5 8, I'm sorry but we tried again and it was impossible.*  
*P: okay merci, [...] 5 0 5 8.*
- (g) *P: -huh- [...] 8 1, would flight level 3 5 0 be available?*  
*C: -huh- okay, we check that and call you back sir.*  
*P: copied, [...] 8 1, thanks.*

The general use of pronouns in UseC can again be explained by the “human” character of the communications it comprises. We can indeed consider phraseology as an “objective” type of discourse which strives to reduce the presence of individual speakers to a minimum (Kerbrat-Orecchioni, 1999: 80): the main syntactic characteristics of phraseology (the deletion of subject pronouns, determiners and modals, for instance) illustrate the objectivity of this type of discourse. Therefore, air-ground communications containing subject pronouns, but also determiners, modals, or interrogative forms, can be considered a far more personal or subjective type of discourse. Pronouns can be seen as “subjectivity” markers which insist on the presence of individual speakers despite the norm that is imposed on them: a reminder that pilots and controllers are humans and not machines.

---

<sup>19</sup> A cumulonimbus (or CB) is a mass of thick cloud that usually involves rain and thunder and that cannot be crossed by any aircraft.

## 6. Conclusion and Prospects

The first results obtained by comparing the distribution of the corpora lexicon corroborate our idea of the relevance of a linguistic approach and, more specifically, of a comparative study between our two corpora of English radiotelephony. The preliminary results of our lexical analysis indicate a general pattern of similarities between the two corpora: both are constituted of a finite lexicon comprising less than 600 word types and being mainly composed of nouns, then verbs and prepositions. Yet, differences have also been observed and a description of the different markers introducing subjectivity in air-ground communication can help understand the use of a more natural language in radiotelephony. In addition, a more detailed comparison of the word forms distributed in the corpora in the various grammatical categories, as well as a comparison of the corpora at a syntactic, semantic and phonetic level will enable us to draw a panorama of the different types of usages made of the English language by pilots and controllers. Conducting the study at other linguistic levels will allow to observe, for instance, the word collocations, i.e., which words are generally used together, the syntactic structures employed by pilots and controllers, or the use of certain verbs with specific complements.

The various results obtained will be used by the ENAC for the English training it provides future controllers and pilots with. This training, based on real usages from different air traffic control centres in the world, tries to heighten future controllers and pilots' awareness about the various difficulties related to language uses. Original teaching materials could be founded on UseC and the results acquired could serve as the basis for various exercises. Such appropriate and up-to-date pedagogical materials could reflect both standard phraseology and the usages made of it in real air traffic control situations and thus, prepare controllers-and-pilots-to-be to face different types of language uses, as required by ICAO's language proficiency requirements.

## References

- Bowker, L. & Pearson, J. (2002). *Working with Specialised Language, a Practical Guide to using Corpora*. London, New York: Routledge.
- Chalker, S. & Weiner, E. (1994). *The Oxford Dictionary of English Grammar*. New York: Oxford University Press.
- Crystral, D. (2003). *English as a Global Language*. Cambridge: Cambridge University Press.
- Condamines, A. (2008). "Peut-on prévenir le risque langagier dans la communication écrite?" In P. Vergely (dir.) *Langage et Société, Communiquer au travail en situation de risque*, 2008/3(125), 77-97.
- DGAC (2007). *Phraséologie. Manuel de formation à la phraséologie à l'usage de la circulation aérienne générale*. Ministère de l'équipement, des transports et du logement, Direction Générale de l'Aviation Civile, Direction de la Navigation Aérienne, Service de l'Information Aéronautique, 5th ed.
- Falzon, P. (1986). *Langages opératifs et compréhension opérative*. Doctoral Thesis. University of Paris la Sorbonne, France.

- ICAO (2001). *Aeronautical Telecommunications. Annex 10 to the Convention on International Civil Aviation*, vol. 2, 6th ed.
- ICAO (2007). *Manual of Radiotelephony*. Doc 9432-AN/925, 4th ed.
- ICAO (2010). *Manual on the Implementation of ICAO Language Proficiency Requirements*. Doc 9835. AN/453, 2nd ed.
- Kerbrat-Orecchioni, C. (1999). *L'énonciation*. Paris : Armand Colin.
- Mell, J. (1992). *Étude des communications verbales entre pilote et contrôleur en situation standard et non-standard*. Doctoral Thesis. University of Toulouse le Mirail, France.
- Nübold, P. & Turner, J. (1983). "Linguistic Redundancy in English Aeronautical Radiotelephony. A Case Study". *Braunschweiger Anglistische Arbeiten*, 8. Braunschweig : Technische Universität Carolo-Wilhelmina.
- Philps, D. (1989). *L'anglais de la circulation aérienne*. Toulouse, France: ENAC.
- Philps, D. (1991). "Linguistic Security in the Syntactic Structures of Air Traffic Control English". *English World-Wide*, 12(1), 103-124.
- Rubenbauer, F. (2009). *Linguistics and Flight Safety: Aspects of Oral English Communication in Aviation*. Aachen, Germany: Shaker Verlag.
- Seidlhofer, B. (2005). "English as a Lingua Franca". *ELT Journal*, 59(4), 339-341.
- Sinclair, J.M. (1991). *Corpus, Concordance, Collocation*. Oxford, United Kingdom: Oxford University Press.
- Vergely, P. (2008). "Communications professionnelles et usage de la langue naturelle : une question de risque(s)?" *Les enjeux de l'information et de la communication*. [on line] Retrieved [December 2009] from: [http://w3.u-grenoble3.fr/les\\_enjeux/2008/Vergely/Vergely08.pdf](http://w3.u-grenoble3.fr/les_enjeux/2008/Vergely/Vergely08.pdf)