



HAL
open science

The linguist and workplace norms

Anne Condamines, Stéphanie Lopez, Maxime Warnier

► **To cite this version:**

Anne Condamines, Stéphanie Lopez, Maxime Warnier. The linguist and workplace norms. Cahiers de Lexicologie, 2017, 2017-1, pp.171-190. halshs-01591940

HAL Id: halshs-01591940

<https://shs.hal.science/halshs-01591940>

Submitted on 22 Sep 2017

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.

THE LINGUIST AND WORKPLACE NORMS

Abstract: Within the workplace, the notion of “linguistic norm” is often considered as equivalent to that of “Controlled Natural Language”, namely a set of recommendations intended to limit the difficulties inherent in the use of natural language. This point of view does not take into account another kind of norm that emerges spontaneously within groups of speakers, notably when they share a common goal. The present paper aims to demonstrate that these two kinds of norms, prescribed and spontaneous, must both be considered by linguists if they want to improve information transmission. The results obtained in two systematic studies are presented, one on an oral corpus in the Air Traffic Control domain and the other on a written corpus in the space domain. The implementation of prescribed linguistic norms in real situations is described. The authors propose to improve the design of Controlled Natural Languages by taking into account not only the experts’ knowledge but also linguistic knowledge, which has rarely been done.

Résumé : Dans le cadre professionnel, la notion de « norme langagière » est souvent considérée comme équivalente à celle de « langue contrôlée », c’est-à-dire un ensemble de recommandations censées limiter les difficultés liées à l’utilisation de la langue naturelle. Ce point de vue ne tient pas compte du fait qu’un autre type de norme s’instaure spontanément dans des groupes de locuteurs, en particulier quand ils ont un objectif en commun. L’article vise à montrer que ces deux types de normes, prescrite et spontanée, doivent être pris en compte par les linguistes s’ils espèrent contribuer à améliorer la transmission d’informations. Les résultats obtenus pour deux études systématiques, l’une, réalisée sur un corpus oral, dans le domaine du contrôle aérien, l’autre, réalisée sur un corpus écrit, dans celui du spatial sont présentés. Ils montrent comment se fait, en situation réelle, la mise en œuvre des normes langagières. Les auteurs émettent des propositions pour améliorer la constitution de langues contrôlées, qui tiennent compte non seulement des connaissances des experts du domaine mais aussi des connaissances sur le fonctionnement de la langue, ce qui est peu souvent le cas actuellement.

Keywords: Genre, Controlled Natural Language, Normalization, Norming

Mots-clés : Genre, Langue Contrôlée, Normalisation, Normaison

1. Introduction

While psychologists, ergonomists or sociologists are frequently called upon by firms, especially in the field of Human Factors, this is rarely the case for linguists. This is surprising since language is omnipresent within the workplace, where most

of the information is transmitted through spoken or written language. Nevertheless, most of the speakers within firms are not aware that using natural language is not always the best way to convey information. Language can indeed be polysemous, ambiguous and vague (Tuggy, 1993). What does not cause a problem in everyday conversation may become crucial in situations where human lives or large amounts of money are involved (possibility of delays, increased costs, litigation, or even accidents). In many firms, linguistic norms have been established to control and mitigate the difficulties linked to the use of natural language to convey information. These norms are materialized as Controlled Natural Languages (CNL) or written recommendations. Unfortunately, in some cases, such norms are not very efficient, because they are either too restrictive or not restrictive enough. This may be partly explained by the fact that other norms are at work among speakers within firms, namely the spontaneous linguistic regularities that emerge whenever human beings are involved in a common task. These regularities also constitute norms since all the speakers involved in a common task have to use them in order to be accepted as a member of the speech community (Hymes, 1972). But unlike Controlled Natural Languages, these norms are not consciously prescribed.

In this paper, we aim to show that these two kinds of norms – the prescribed one and the spontaneous one – must be considered within the workplace in order to improve the way information is conveyed. First, a definition of Controlled Natural Languages will be presented and their similarities with sublanguages will be underlined (2.1). In order to show that two linguistic norms are at stake within the workplace, our assessment of CNL is based on two terms used by the French Linguistic School of Rouen: *normalisation* and *normaison*, respectively translated as “normalization” and “norming” in this paper (2.2). Then, we present the results of a study carried out for the French Civil Aviation University (École Nationale de l’Aviation Civile) in the domain of Air Traffic Control (ATC). It consists of a comparative analysis between two corpora: one representing the prescribed norm of the domain and the other corresponding to real pilot-controller communications (3). Finally, a study conducted for the French Space Agency (Centre National d’Études Spatiales) is presented: it introduces a method for building a CNL based on real linguistic productions (4).

2. Linguistic Norms within the Workplace

Very few linguists work in professional settings even though language is omnipresent in the workplace and is used as a tool to convey both written and oral information. In this kind of context, the term “norm” is generally used to refer to Controlled Natural Languages or to writing guidelines developed to limit specific problems associated with the use of language. Admittedly, such norms are necessary but they would be more effective if linguists – as experts of the language and its difficulties – took part in their development, along with psychologists and domain experts. A prescribed linguistic norm can only be truly efficient when it takes into consideration the fact that a norm can also be a spontaneous form emerging in a speech community (Hymes, 1972).

In order to distinguish and outline these two kinds of norms (the prescribed one and the spontaneous one), we will use the French terms *normalisation* and *normaison*, commonly used by the French Linguistic School of Rouen:

L'analyse tirerait profit à opposer deux procès normatifs : la normaison, relevant de l'activité spontanée à l'œuvre dans tout échange, et la normalisation, domaine des interventions conscientes et planifiées (Gaudin, 1993: 173). [The analysis would gain by contrasting two normative processes: norming, which is an intrinsic feature of spontaneous language activity, and normalization, which is a conscious and planned activity].

As shown by the morphological similarity of the two terms, in both cases, it is the norm that is at issue. In the first case, however, the norm is considered from a descriptive and non-conscious point of view, whereas in the second case, it is considered from a prescriptive and conscious point of view. As linguists, we thus think it necessary to take into account the following two realities within the workplace: the need for controlled languages to guide the way in which natural language should be used, and the fact that, within restricted communities, language is always subject to variation.

2.1. Controlled Natural Languages and Sublanguages

A detailed analysis of the definition of the notions of normalization and norming shows that they are quite close to those of Controlled Natural Language and sublanguage. Moreover, when examining the way CNLs and sublanguages are described, one observes that they differ in only one respect: the former is prescribed while the latter is not.

The notion of sublanguage was first introduced by Harris (1968) from a mathematical point of view, in order to explore the possibility of reaching the core meaning of a corpus by using linguistic transformations. This led Harris to consider sublanguages as subsets of natural languages. Generally, sublanguages are described as functioning on a deviant or more restricted mode than general language.

What is clear is that whether or not it is considered as a subset of all the sentences of a language (and described with notions such as “deviant” or “restrictive”), a sublanguage has its own rules (its own norms) which are developed spontaneously by the speech community:

Grammaticality in a subject-matter sublanguage is determined by whatever officially prescribed or implicit norms of usage exist among the specialists in the subject-matter field (Lehrberger, 1986: 27).

From a communicative point of view, sublanguages are not far from the notion of textual genre, defined by Bhatia (1993: 13) as “a recognizable communicative event characterized by a set of communicative purpose(s) identified and mutually understood by the members of the professional or academic community in which it regularly occurs”. As a matter of fact, the terms *sublanguage* and *genre* are often conflated. For example, Somers (1998: 131) defines *sublanguage* as “an identifiable genre or text-type in a given subject field, with a relatively or even absolutely closed set of syntactic structures and vocabulary”.

In this paper, we do not aim to clarify the differences between the terms *sublanguage*, *genre* or even *specialized language* or *Language for Specific Purposes*, but rather to stress that recurrent and specific situations lead to specific linguistic regularities. As a result, since the communicative situations change, linguistic regularities themselves are likely to change and, as it is impossible to determine all the changes in a given situation, the use of language, even if

controlled, may always generate risks of misunderstanding (Condamines, 2008; 2010). Nevertheless, we believe that CNLs used in professional settings should be created on the basis of real spontaneous regularities, because such CNLs, being more natural, are more likely to be used.

In Kuhn's (2014) definition of CNL, the only difference with a sublanguage is that the former is a constructed language. Nevertheless, a CNL, like a sublanguage, maintains a strong relationship with natural language:

A **controlled natural language** is a constructed language that is based on a certain natural language, being more restrictive concerning lexicon, syntax, and/or semantics while preserving most of its natural properties (Kuhn, 2014: 123).

Kuhn (*ibid.*) notes that a CNL "is explicitly and consciously defined, and *is not* the product of an implicit and natural process (even though it is based on a natural language that *is* the product of an implicit and natural process)". We should add that CNLs are generally built by domain experts (speakers of the language) who do not have a conscious knowledge of the speech regularities of the sublanguage they are using. Considering the passive but real knowledge on which they are built, it is interesting to study the rules contained in CNLs, even if they are not built according to a systematic approach.

Another aspect of CNLs to bear in mind is that they are often related to and based on one another. Early CNLs and current ones still share similarities. The first CNLs were based on studies carried out nearly a century ago on the readability of texts (DuBay, 2004; Klare, 1976). The current tests defined for evaluating readability, based on human appreciation, have not really changed except for the recent use of eye-tracking (O'Brien, 2010).

2.2. Normalization and Norming: Two Moments in the Creation of a CNL

It seems obvious that sublanguages (corresponding to the norming process) and CNLs (corresponding to the normalization process) are linked and that they are related to the same natural language. Sublanguages appear spontaneously in identified communicative situations while CNLs are consciously built for the same situations by authorized bodies. Consequently, when developing a CNL, two pitfalls should be avoided: it should be neither too constraining nor too broad; in other words it should be not too different from the existing sublanguage in a specific situation.

Our goal is to combine crucial requirements to help official bodies define efficient linguistic norms based on their domain's spontaneous regularities while taking into account the fact that language constantly changes in response to the extra-linguistic context, as illustrated in Figure 1 (based on Lopez (2013: 31)).

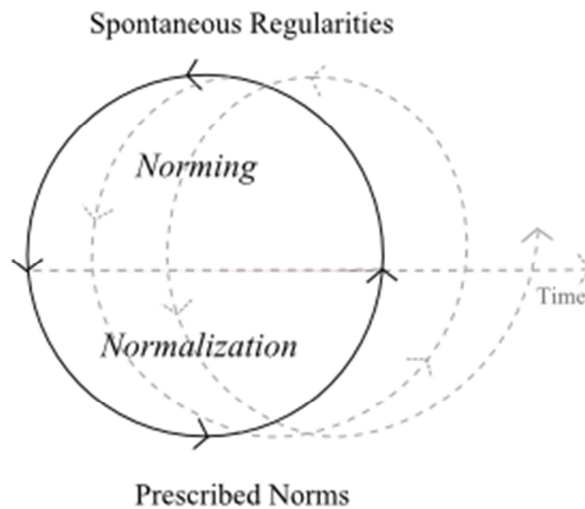


Figure 1: Expected influence of spontaneous regularities on prescribed norms and vice versa

In order to illustrate our perspective, the following sections present two research projects conducted in domains where misunderstandings may have major consequences: Air Traffic Control and the specification of space projects. In the former case, we show that the spoken CNL in use (named *phraseology*) is not always strictly followed and we try to explain in which cases and why this occurs. In the latter case, we describe a method to build a written CNL based on the spontaneous regularities of a specific sublanguage.

3. Implementation of the Prescribed Norm of Air Traffic Control

In the domain of Air Traffic Control, where pilots and controllers internationally share the same controlled language to secure and optimize their most frequent communications, linguistic normalization is necessary to provide clear and unambiguous verbal communication between experts in this field. In this part, we present how this normalization is implemented in ATC communications.

3.1. Context of the Study

At the end of World War II, the rapid growth of civil aviation and radiotelephony entailed the development and adoption of prescribed linguistic norms and recommendations in the domain of ATC. Pilot-controller (or air-ground) communications have been performed ever since by means of a specialized language (considered as a CNL) usually based on English, called *phraseology*. It was designed by the International Civil Aviation Organization (ICAO) to try and ensure safety in the most common radiotelephony communications through simplified but strict lexical, syntactic and semantic rules (ICAO, 2010). Its main characteristics are as follows (Lopez, 2013; Mell, 1992; Philips, 1991)¹:

- the omnipresence of the imperative form in the controllers' messages;
- the rarity of interrogative and negative forms;
- the rarity of modals;
- the deletion of determiners, subject pronouns and prepositions;

¹See 3.3 for a few examples of what phraseology looks like.

- the deletion of the auxiliaries *be* and *have* in *be + past participle* forms, *be + -ing* forms and *have + past participle* forms;
- the nominalization of concepts;
- a highly specialized, univocal and finite lexicon;
- the use of an alphabet specific to the aeronautical domain;
- the specific spelling and pronunciation of numbers.

These particular characteristics of phraseology make it the essential communication tool² for transmitting the fundamental information required in routine air navigation situations. However, several problems relating to the implementation of phraseology have been identified by Aviation Experts and professional Aviation English teachers. These problems and difficulties can result from a variety of reasons such as non-compliance with the norm by pilots and controllers (Mell, 1992: 75), the fact that they have become accustomed to the phraseology and thus less vigilant (Estival and Molesworth, 2012: 353; Hawkins, 1987: 154) or the level of English of the speakers (Estival and Molesworth, 2012; ICAO, 2010). For the ICAO (2010: 1-1), the latter criterion is the common thread running through three of the most fatal accidents in aviation history. Therefore, in the last two decades, radiotelephony communication standards have been revised and a minimum level of English has been imposed on all pilots and controllers throughout the world (ICAO, 2010: vii).

In France, the École Nationale de l'Aviation Civile (ENAC) – the French Civil Aviation University – has to comply with the new language proficiency requirements of the ICAO. It has therefore specific needs in terms of English radiotelephony teaching. In order to get a better idea of the actual implementation of phraseology in air-ground communications, the ENAC initiated an observation of the use of English by French controllers and international pilots in routine ATC situations. The aim of this research project³ was to describe and categorize the different uses of radiotelephony English by means of a comparative study between two corpora: one representing the prescribed norm of the domain and the other consisting of real air-ground communications.

3.2. Corpora under Study

The two corpora on which our analysis is based were designed and compiled according to specific criteria⁴. To represent the linguistic norm of ATC, the reference corpus (henceforth referred to as *RefC*) comprised examples in English from two phraseology manuals corresponding to the only available official documents containing representative samples of standardized official phraseology. It is a small corpus as it only contains 7,154 word tokens of 671 word types. The second corpus, representing the real usage of the norm (henceforth referred to as *UseC*), consisted of the orthographic transcription of about twenty-two hours of recordings of real air-ground communications at three French control centers. It contains 24,465 tokens of 682 word types.

² For more information about the limitations of phraseology and the notion of “plain language”, the natural language to which pilots and controllers must resort to when facing situations not covered by phraseology, see Lopez (2013).

³ This study was funded by the ENAC in the context of a doctoral research project.

⁴ For more information about the methodology we used, see Lopez (2013).

The first step in the analysis was to classify the different words contained in both corpora according to their part-of-speech or grammatical categories. This task was handled manually as the particular syntactic structures of phraseology do not allow a correct automatic tagging of the corpora. Even though the resulting classification displays a balanced overall distribution of part-of-speech categories between RefC and UseC, several salient discrepancies can be observed within the different categories. The nouns, adjectives, interjections, determiners and pronouns are unevenly distributed in the two corpora, as illustrated in Figure 2.

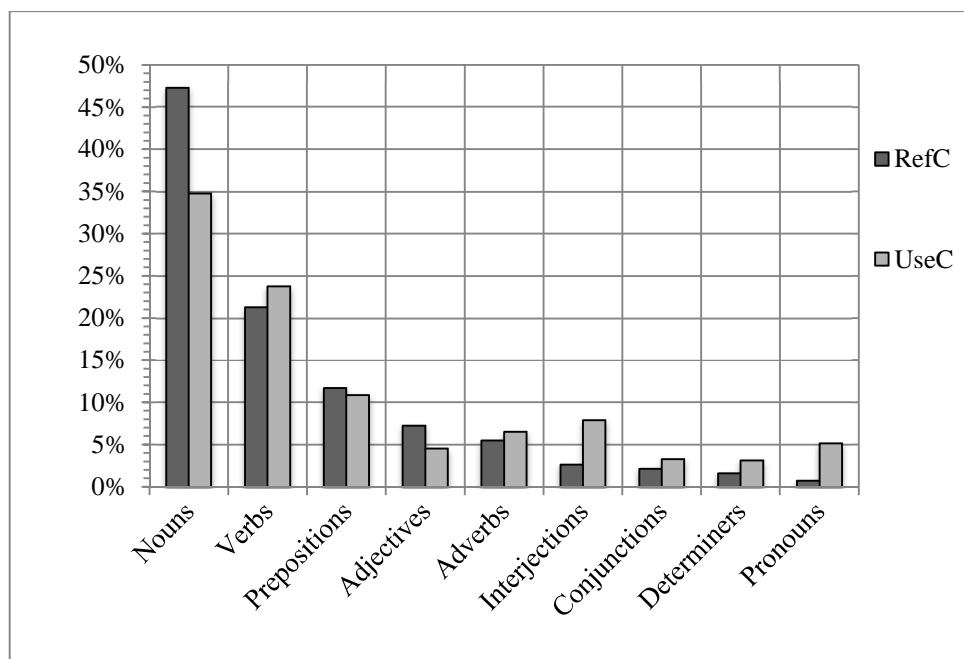


Figure 2: Classification of the words contained in RefC and UseC by grammatical category

A detailed analysis of the categories exhibiting the most discrepancies revealed different types of variations within pilots' and controllers' messages. Depending on the situation, their messages presented variations from standardized phraseology from a lexical, semantic or syntactic and discursive point of view. While some variations reflect the influence of natural language on the speakers, others can be considered as a strategy to "humanize" radiotelephony communications, as presented in 3.3.

3.3. Results

In her study of ATC communications, Wyss-Bühlmann (2005: 55-65) determined three different types of variations within normalized messages: the replacement of a standard phrase, the addition of an element to a standard phrase, and the omission of a standard phrase.

These three variation processes can also be observed in UseC. In accordance with the characterization suggested by Wyss-Bühlmann (2005: 106-107), the variations in UseC can be classified according to their orientations. Some variations seem to result from individual linguistic creativity, thus illustrating the influence of natural language on real air-ground communications, whereas others can be

considered as strategic variations since they try to cover specific communication needs such as the clarification of a message or the request for a particular action.

In UseC, these variations, whether free or strategic, belong to three different categories: they can be lexical, semantic and/or syntactic. The potential boundaries between those categories remain unclear however since one message can contain several types of variations at once: a lexical variation can induce a semantic one and occur within a message presenting a syntactic change from the prescribed norm. Due to lack of space, only a few examples of the different types of variations observed in UseC will be briefly introduced here⁵.

One recurrent example of free lexical variation is the omission by controllers and pilots of the term *decimal* when transmitting a radio frequency while it is mandatory in phraseology (ICAO, 2001: 5-5), as illustrated in the following example extracted from UseC:

C: Speedbird 81, climb flight level 3 4 0, contact [Centre] 1 2 4 0 8 0, goodbye (instead of “1 2 4 decimal 0 8 0”).

This kind of omission seems quite problematic as most of the pilots' requests for controllers to repeat or confirm a frequency primarily occur when *decimal* has been omitted. The omission of *decimal* and the need for several confirmations of a given frequency thus result in the unnecessary occupation of the communication channel, potentially blocking the transmission of important and urgent messages.

An example of a strategic syntactic variation in UseC is the use of natural language instead of phraseology in situations for which it exists. For instance, some controllers give authorization clearances by means of declarative sentences containing modals rather than the imperative structure prescribed in phraseology:

C: Olympic 2 0 7, you may reduce 1 8 0 knots at your convenience (instead of “reduce 1 8 0 knots”).

Likewise, to have a better chance of obtaining a positive answer to a specific personal request, some pilots prefer the interrogative form rather than the prescribed imperative or *-ing* form, thus mitigating the strictness of phraseology and sounding less bossy or demanding:

P: -huh- can you please coordinate -huh- routing via [Town]? (instead of “coordinate routing via [Town]”)

This subtle kind of strategy can be seen as an attempt by the speakers to obtain what they want: both pilots and controllers resort to strategic syntactic variations in routine situations in order to “humanize” some of their messages. These variations seem to result from the speakers' decision to conduct ATC communications with mutual respect and in a spirit of partnership. The high frequency of politeness and greeting markers (even though forbidden) found in UseC (1,495 tokens – 78% of the total number of interjections in UseC) corroborates this hypothesis.

Another example of strategic syntactic variation found in UseC is the addition of an element to a standard phraseology phrase for reasons of clarity and comprehension. For instance, some controllers choose to add “(on) reaching” (31 tokens) at the end of the clearance “maintain flight level...” thus emphasizing the importance of not breaching unauthorized spaces in order to keep a safe minimum distance between planes. Other controllers avoid the potential ambiguity of the verb

⁵For more details, see Lopez (2013) and Lopez, Condamines and Josselin-Leray (2014).

report when followed by a noun – which, according to the context, can mean 1/“*call me back later once you have accomplished what I ask*” or 2/“*tell me now what is the flight parameter I need to know*” – by adding an element which conveys a third unequivocal meaning, as illustrated in the following examples:

1/ C: [...], **report Apron** (*Report later, when you are at the Apron*)

2/ C: [...], **report present heading** (*Tell me/What is your heading now?*)

3/ C: [...], **report your new heading to [Center] 1 3 3 6 9 0** (*Report your heading to the next control center*).

These controllers resort to a linguistic and cognitive shortcut by contracting two different phraseology phrases (“report heading” and “contact + frequency”). By doing so, they avoid the possible ambiguity of the verb *report*.

To conclude this section, the comparison of RefC and UseC reveals a high number of recurrent variations between the prescribed linguistic norm in the domain of ATC and the real use made of it, even though the norm was implemented for security and optimization reasons. While non-strategic or free variations could be curtailed by insisting on the importance of respecting phraseology whenever it is available, strategic variations seem quite difficult to control as they try to meet pilots’ and controllers’ communication needs and specific requests. It is likely that strategic variations (addition of an element to a standard phrase, replacement of a standard phrase, change in syntactic structure, use of politeness and greeting markers, etc.) will remain as long as air-ground communications are performed by human pilots and controllers:

It is a fact that, in ATC speech, these strategic changes to standard phraseology will convey the thoughts and emotions of the speakers involved and will have effects on the interlocutors as long as voice communication plays a part in pilot-controller interaction (Wyss-Bühlmann, 2005: 186).

As some of these variations correspond to spontaneous regularities found in the domain, we suggest that they should be taken into account by the competent authorities when updating or improving the established prescribed norm. Their inclusion would guarantee a better implementation of the norm by pilots and controllers.

4. Prescribed and Descriptive Norms in the Specifications of Space Projects

This second study⁶ originated from a request by the Quality Assurance Sub-Directorate of CNES (Centre National d’Études Spatiales), the French Space Agency: how could the writing of space project requirements be improved? *Requirements* are defined by IEEE (1990) as “conditions or capabilities that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed document”⁷. Most of them are written and organized in documents called *requirement specifications*, i.e. sets of requirements (which may also contain other information, such as comments or justifications). It is now widely acknowledged that requirements are critical, because they greatly

⁶ This research project is carried out as part of a PhD thesis granted by CNES and the Regional Council Midi-Pyrénées.

⁷ An example of a single requirement (expressing such a capability) is the following sentence: “For each measure session, ORAMIC must precisely return the attitude of the satellite Microscope.”

influence the success or failure of the projects they are written for – especially in the case of large projects, involving many people over several years. Consequently, requirements have been studied in the field of requirements engineering and several good practices have been proposed. Besides, requirements have a contractual value for the stakeholders (in this case, CNES and its clients or suppliers) who are obliged to respect and verify them.

4.1. Context and Aim of the Study

Unlike other workplaces where a vast range of (more or less constraining) preventive solutions exist, none of them are currently used at CNES, meaning that the engineers are free to write requirements the way they like and feel is the best. Despite its drawbacks, natural language also has interesting characteristics, such as simplicity and expressivity, that are worth preserving when trying to restrict it. In any case, natural language remains unavoidable in some of the stages of the specification process.

Our aim is therefore to design a Controlled Natural Language (i.e. a set of rules) for requirements writing in French at CNES that would prevent most of the potential problems related to language, while remaining as close as possible to the way engineers are already used to writing (i.e., close to what we can find in existing specifications) (Warnier, 2015), since, if we want our rules to be actually applied (and not only to be seen as an external constraint), they have to be tailored to real practice. If they are not, they will at best be ignored, at worst be counterproductive. To do so, the first step is to detect linguistic regularities in the specifications. However, it could be argued that if the CNES engineers are not asked to follow any writing rules, then there may not be any regularities (because each writer has his or her idiosyncratic style). Nevertheless, we assume that spontaneous regularities are likely to arise as these engineers form a professional community and because requirement writing is a recurring, well-identified task – leading exactly to what has been called *genre*, or to what other authors have called *sublanguage* (see section 2.1). Furthermore, writers may also be influenced, for instance, by the feedback they receive when the specifications are revised by the project leader, or simply by their own intuitions. Therefore, our aim is to build the grammar of this genre or, in Somers' words, to identify the “set of syntactic structures and vocabulary” (Somers, 1998: 131) of this sublanguage, and to use it to define the rules of our CNL.

4.2. Corpus under Study

The corpus on which our analyses are based is composed of requirements in French, automatically extracted from the specifications of two different projects, namely “Pleiades” (approximately 2,500 requirements, totaling 120,000 words) and “Microscope” (approximately 1,000 requirements, totaling 44,000 words)⁸. The specifications of these two projects (15 for each) share common points: they represent the different levels of the so-called “product tree” (system, segment, interface) and their writers have similar profiles. The projects also have notable differences: a difference of scale (Pleiades being a much bigger project than

⁸ Pleiades consists of two very-high-resolution Earth-observation satellites (launched in 2011 and 2012), while Microscope concerns a smaller satellite that has been launched in 2016 to test the weak equivalence principle.

Microscope), and consequently a difference in the number of requirements per specification and a difference in goals (and thus a difference of domain). Although they represent two rather different projects conducted recently at CNES, the existence of a common “corporate culture” can be assumed. Albeit limited (as is often the case in specialized domains), the size of the corpus should allow us to perform some automatic analyses, followed by manual revision.

In addition to the requirements corpus, we also used two other corpora (of the same size and in the same language) for comparison purposes. The first one was extracted from a technical handbook about the techniques and technologies used for building and operating spacecraft, written at CNES by experts for semi-experts, and the second one was composed of articles extracted from the French newspaper *Le Monde*. While the former is in theory close to the requirements corpus in several ways (it is also a specialized text concerning the space domain), the latter represents a genre (press articles) closer to “general language” (although it has, of course, its own specificities) and covers various domains. In the near future, we plan to build another comparison corpus, containing requirements from another field, as the ultimate goal is to find regularities in requirements, regardless of the domain.

4.3. Methodology

For each phenomenon that we examine before developing new rules (see below), two kinds of analyses will be applied: a quantitative one and a qualitative one. On the one hand, thanks to the quantitative analysis (indicating the frequency of the phenomenon in the different corpora), we hope to be able to show the existence of linguistic regularities specific to the requirements corpus, compared with the other corpora (which represent other genres). On the other hand, the qualitative analysis consists in manually reviewing a large sample of cases where the phenomenon occurs in the corpus, and trying to determine for each occurrence whether or not it is problematic. Naturally, we do not limit ourselves to a binary distinction (safe or dangerous), but we would like to propose a finer-grained classification showing in what condition(s) and for what reason(s) they represent potential sources of misunderstanding. This classification will then serve as a concrete and more reliable basis for the rules of the CNL (Condamines and Warnier, 2014).

In order to identify regularities in our requirements corpus, two theoretically distinct (but not incompatible) corpus linguistics approaches are suitable: the corpus-based approach and the corpus-driven approach (Tognini-Bonelli, 2001). According to Biber (2009: 276), *corpus-based* research “assumes the validity of linguistic forms and structures derived from linguistic theory”, whereas *corpus-driven* research “is more inductive, so that the linguistic constructs themselves emerge from analysis of a corpus.” To assist in finding recurring constructs with a corpus-driven perspective, natural language processing tools based on statistical algorithms can be used. In a corpus-based approach, however, it is the linguists’ role to make their own hypotheses to be tested on the corpus. To draw such hypotheses, we first reviewed several existing CNLs and tried to determine what linguistic phenomena were addressed (often implicitly) by the rules they impose.

We focused on two of these CNLs: the first one, called ASD Simplified Technical English is a well-known standard for the “preparation of maintenance

documentation” in the aerospace industry (AeroSpace and Defence Industries Association of Europe, 2007: i); the second one, called *Guide for Writing Requirements*, was proposed by the International Council on Systems Engineering (2011: 10) and aims at “draw[ing] together advice from a variety of existing standards into a single, comprehensive set of rules and objectives⁹”, and could thus be seen as the state of the art for requirements writing. Our main criticisms of the rules imposed by these two CNLs (and one of the reasons why we decided not to keep them unchanged) is that some of them are too constraining (and thus, not applied), others are insufficiently so (and thus, inefficient), and some are inconsistent (in particular with regard to the examples of the “good” or “bad” statements they provide). In other words, they are not always suitable for actual writing – not to mention the fact that their justifications, if present, are often evasive. However, these rules – whether they were actually followed or not – are an interesting starting point for our investigations, because they were probably the answers to previously identified problems.

Lastly, external reasons guided the choice of phenomena we selected for investigation: since we rely on automated tools to perform our analyses (and because we would like our rules to be implemented in the future, meaning that these tools have to be available for semi-automatic checking), we only selected phenomena for which such tools (preferably open-source, for confidentiality reasons) exist for French; this is obviously not the case for complex analyses dealing with semantic features¹⁰. Similarly, since the CNLs we considered are designed mostly for English while our corpus is in French, we limited ourselves to phenomena that are not too language-dependent and could be transposed from one language to the other.

4.4. Results

In this section, we will briefly comment on the results obtained by analyzing two phenomena based on two rules found in the *Guide for Writing Requirements*:

R19 /SINGULARITY/PROPOSITIONALS/

Avoid combinators.

Elaboration: Combinators are words that join clauses together, such as 'and', 'or', 'then', 'unless'. Their presence in a requirement usually indicates that multiple requirements should be written. (p. 39-40)

R25 /COMPLETENESS/PRONOUNS/

Repeat nouns in full instead of using pronouns to refer to nouns in other requirement statements.

Elaboration: Pronouns are words such as 'it', 'this', 'that', 'he', 'she', 'they', 'them'. When writing stories, they [...] are a useful device for avoiding the repetition of words; but when writing requirements, pronouns should be avoided, and the proper nouns repeated where necessary. (p. 43)

Concerning the first rule (R19), given the definition of “combinator”, we decided to look for all conjunctions (coordinating and subordinating). Concerning

⁹ Therefore, some of the rules imposed by Simplified Technical English (or similar ones) may also be present in the *Guide for Writing Requirements*.

¹⁰ For example, a rule found in the *Guide for Writing Requirements* advocates that “a requirement statement should be expressed at a level of detail appropriate to its level of abstraction” (2011: 20).

the second rule (R25), we merely decided to look for all pronouns. The results are illustrated in Figure 3:

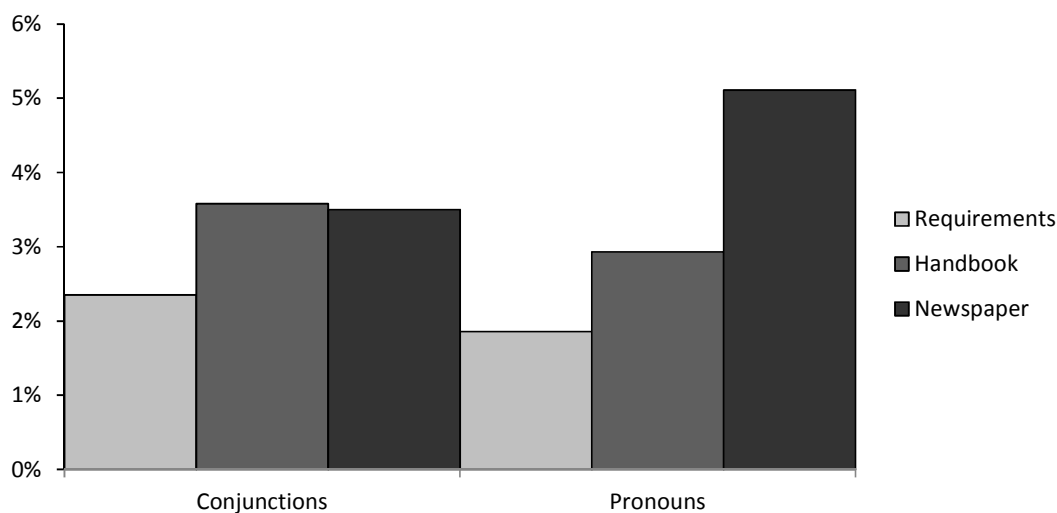


Figure 3: Proportion of conjunctions and pronouns in relation to the total number of words

As can clearly be seen, although they are present, conjunctions and pronouns are less frequent in the requirements corpus than in the other two corpora. This may indicate that requirements correspond indeed to a different textual genre.

When reviewing some examples of requirements containing conjunctions and pronouns, it appears that some of them are unavoidable:

“Le générateur de TCH vérifiera *que* la valeur du champ PHASE est comprise entre 0 *et* *FREQ_DIV-1*.” [“The generator of TCH will check *that* the value of the field PHASE is between 0 *and* *FREQ_DIV-1*.”]

Here, “*que*” (“that”) introduces the argument clause and is therefore mandatory, and “*et*” (“and”) is necessary to set the limits of the interval (announced by “between”).

Others, in contrast, could be avoided, but do not seem to be problematic:

“Le paquet ne sera généré *que* *s’il* est activé par le LVC.” [“The packet will be generated only if *it* is activated by the LVC.”]

Although “*il*” (“it”) could be replaced by “le paquet” (“the packet”) – resulting in a less natural sentence –, this does not seem to be necessary, because it does not have any other possible antecedent within the requirement, which is therefore not ambiguous.

Finally, some conjunctions and pronouns are indeed undesirable (and correspond probably to the issues that the rules seek to address):

“*Il* calculera aussi, à une fréquence paramétrable (ordre de grandeur 1 mois), la moyenne de mise en œuvre [...]” [“*It* will also calculate, at a frequency that can be parameterized (at monthly intervals), the average time for commissioning [...]”]

This requirement is no longer autonomous, because the subject pronoun “*il*” (“it”) refers to a noun defined in the previous requirement (which might be removed later).

“Les demandes sont saisies sur le FOS *et* le logiciel ARPE gère les conflits entre les demandes Spot, Hélios et Pléiades” [“The requests are to be entered on the FOS *and*

the ARPE software manages conflicts between the requests from Spot, Hélios and Pléiades”]

In the case of two independent clauses that have nothing in common, joined by “et” (“and”), as in the example above, it would certainly be better to write multiple requirements, as suggested by the Guide.

From these few examples, we can conclude that some words or phenomena are indeed sometimes problematic because they violate the principles of good requirements (e.g. atomicity or autonomy), but also that it is not the case for all their occurrences. Therefore, the rules found in CNLs should probably be refined to take the different contexts into account.

5. Conclusion

Within the workplace, the notion of norm requires a fine-grained analysis. It does not correspond only to the prescribed norm, appearing generally in the form of writing or oral recommendations. As in all speech communities, spontaneous regularities occur (generally referred to as sublanguages or textual genres) among speakers working together on a project or a task. These spontaneous norms do not disappear even when a prescribed norm exists. So, we claim that these two kinds of norms should be taken into account. In this paper, we briefly described the results of two analyses concerning the implementation of a prescribed norm in two professional contexts, one concerning oral Air Traffic Control communications and the other the written requirements of two projects in the space domain. Even though the situations are not exactly the same, the results show that the prescribed norm is not always followed because it is not always adapted to the situation, sometimes far from the spontaneous norms or simply too constraining.

More generally, the role of the linguist in the workplace is to support the need for Controlled Natural Languages in order to minimize the risks related to the use of natural language. Within the workplace, language use changes depending on the (emergency of the) situation, the relationships between the partners and many other factors. It is therefore necessary to take into account the language used *in vivo*, by analyzing real data and by building methods to describe uses and variations. From this point of view, the presence of linguists seems very useful within the workplace – and conversely, linguistic theories could certainly benefit from studies about the use of norms in actual professional contexts.

References

- AEROSPACE AND DEFENCE INDUSTRIES ASSOCIATION OF EUROPE (2007): *Simplified Technical English. Specification ASD-STE100*, International Specification for the Preparation of Maintenance Documentation in a Controlled Language, Issue 4.
- BHATIA Vijay K. (1993): *Analysing Genre: Language Use in Professional Settings*, London, Longman.
- BIBER Douglas (2009): “A Corpus-Driven Approach to Formulaic Language in English: Multi-Word Patterns in Speech and Writing”, *International Journal of Corpus Linguistics*, 14(3), p. 275-311.

- CONDAMINES Anne (2008): “Peut-on prévenir le risque langagier dans la communication écrite?”, *Langage et société*, 125, p. 77-97.
- (2010): “Variations in Terminology. Application to the Management of Risks Related to Language Use in the Workplace”, *Terminology*, 16(1), p. 30-50.
- CONDAMINES Anne and WARNIER Maxime (2014): “Linguistic Analysis of Requirements of a Space Project and Their Conformity with the Recommendations Proposed by a Controlled Natural Language”, *Proceedings of 4th International Workshop on Controlled Natural Language (CNL 2014)*, Galway (Ireland), p. 33-43.
- DUBAY William H. (2004): *The Principles of Readability*, Costa Mesa, CA, Impact Information, [On line],
<http://www.impact-information.com/impactinfo/readability02.pdf>.
- ESTIVAL Dominique and MOLESWORTH Brett (2012): “Radio Miscommunication: EL2 Pilots in the Australian General Aviation Environment”, *Linguistics and the Human Sciences*, 5(3), p. 351-378.
- GAUDIN François (1993): *Pour une socioterminologie : des problèmes sémantiques aux pratiques institutionnelles*, Presses Universitaires de Rouen et du Havre.
- HARRIS Zellig (1968): *Mathematical structures of language*, New York, Interscience Publishers, John Wiley & Sons.
- HAWKINS Frank (1987), *Human Factors in Flight*, Aldershot, Ashgate.
- HYMES Dell (1972): “Models of the Interaction of Language and Social Life”, in J. J. Gumperz and D. Hymes (Eds), *Directions in Sociolinguistics: The Ethnography of Communication*, New York: Holt, Rinehart & Winston. pp. 35-71.
- ICAO (2001): *Aeronautical Telecommunications. Annex 10 to the Convention on International Civil Aviation*. Vol. 2, Communication Procedures Including those with PANS Status. 6th Edition.
- (2010): *Manual on the Implementation of ICAO Language Proficiency Requirements*, Doc 9835-AN/453, 2nd Edition.
- IEEE (1990): “Standard Glossary of Software Engineering Terminology”, *IEEE Std 610.12-1990*, p. 1-84, [On line], <http://doi.org/10.1109/IEEESTD.1990.101064>.
- INTERNATIONAL COUNCIL ON SYSTEMS ENGINEERING (2011): *Guide for Writing Requirements*, Version 1.
- KLARE George R. (1976): “A Second Look at the Validity of Readability Formulas”, *Journal of Reading Behavior*, (8)2, p. 129-152, [On line], <http://jlr.sagepub.com/content/8/2/129.full.pdf>.
- KUHN Tobias (2014): “A Survey and Classification of Controlled Natural Languages”, *Computational Linguistics*, 40(1), p. 121-170.

- LEHRBERGER John (1986): “Sublanguage Analysis”, in R. Grishman and R. Kittredge (Eds), *Analyzing Language in Restricted Domains*, Hillsdale, New Jersey, London, Lawrence Erlbaum Associates Publishers, p. 19-38.
- LOPEZ Stéphanie (2013): *Norme(s) et usage(s) langagiers : le cas des communications pilote-contrôleur en anglais*, Doctoral Thesis, University of Toulouse 2 – Le Mirail.
- LOPEZ Stéphanie, CONDAMINES Anne and JOSSELIN-LERAY Amélie (2014): “Analyse des communications pilote-contrôleur : entre norme et réalité des usages langagiers”, *Revue Française de Linguistique Appliquée*, 19(1), p. 87-101.
- MELL Jeremy (1992): *Étude des communications verbales entre pilote et contrôleur en situation standard et non-standard*, Doctoral Thesis, University of Toulouse 2 – Le Mirail.
- O’BRIEN Sharon (2010): “Controlled Language and Readability”, *ATA Scholarly Monograph Series*, 15, p. 143-168.
- PHILPS Denis (1991): “Linguistic Security in the Syntactic Structures of Air Traffic Control English”, *English World-Wide*, 12(1), Amsterdam, Honoré Champion, p. 103-124.
- SOMERS Harold (1998): “An Attempt to Use Weighted Cusums to Identify Sublanguages”, in D.M. W. Powers (ed.), *NeMLaP3/CoNLL98: New Methods in Language Processing and Computational Natural Language Learning*, ACL, p. 131-139.
- TOGNINI-BONELLI Elena (2001): *Corpus Linguistics at Work*, Amsterdam, John Benjamins Publishing.
- TUGGY David (1993): “Ambiguity, polysemy, and vagueness”, *Cognitive Linguistics*, 4(3), p.273-290.
- WARNIER Maxime (2015): “How Can Corpus Linguistics Help Improve Requirements Writing? Specifications of a Space Project as a Case Study”, *Proceedings of 23rd IEEE International Requirements Engineering Conference (RE’15)*, Ottawa, Canada, p. 388-392.
- WYSS-BÜHLMANN Eveline (2005): *Variation and Co-operative Communication Strategies in Air Traffic Control English*, Bern, Peter Lang.

Anne Condamines, CLLE (UMR 5263), Université de Toulouse
anne.condamines@univ-tlse2.fr

Stéphanie Lopez, CLLE (UMR 5263), Université de Toulouse
stephanie.lopez@univ-tlse2.fr

Maxime Warnier, CLLE (UMR 5263), Université de Toulouse et CNES
maxime.warnier@univ-tlse2.fr