



HAL
open science

A Case Study on Evaluating the Relevance of Some Rules for Writing Requirements through an Online Survey

Maxime Warnier, Anne Condamines

► **To cite this version:**

Maxime Warnier, Anne Condamines. A Case Study on Evaluating the Relevance of Some Rules for Writing Requirements through an Online Survey. 25th IEEE International Requirements Engineering Conference, Sep 2017, Lisbonne, Portugal. hal-01543005

HAL Id: hal-01543005

<https://hal.science/hal-01543005>

Submitted on 20 Jun 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.

A Case Study on Evaluating the Relevance of Some Rules for Writing Requirements through an Online Survey

Maxime Warnier

CLLE-ERSS (CNRS & Université Toulouse – Jean Jaurès)
Centre National d'Études Spatiales
Toulouse, France
maxime.warnier@univ-tlse2.fr

Anne Condamines

CLLE-ERSS (CNRS & Université Toulouse – Jean Jaurès)
Toulouse, France
anne.condamines@univ-tlse2.fr

Abstract—As part of a research project that aims at proposing a new methodology for defining a series of rules for writing good requirements – often referred to as a Controlled Natural Language (CNL) – for the French Space Agency (CNES, Centre National d'Études Spatiales), we asked both experienced engineers and non-experts to fill in an online questionnaire in order to gather their perception about requirements written according to recommendations commonly found in CNLs, and to compare them with seemingly more natural and less restrictive formulations. The examples we used for this case study were adapted from genuine requirements in French, extracted from several specifications of a recent space project. Our main goal is to evaluate whether (and to what extent) the writing rules we considered may be relevant for the engineers at CNES. In particular, we try to identify cases where the experts' opinions differ from the recommended use and where these rules could thus probably be adapted.

Index Terms—requirements, technical writing, controlled language, evaluation, survey.

I. INTRODUCTION AND CONTEXT

The importance of Requirements Engineering (RE) has been progressively recognized by most major companies and institutions involved in the development of software or systems, so that it is now generally considered a necessary condition for the success of large projects [1] – as is the case, for example, with space projects –, since it helps avoid problems later in the development process. However, some of the good practices and principles that are recommended at each step of the requirements' lifecycle (from elicitation to validation) may be time-consuming. As a consequence, stakeholders who do not yet master them all are forced to set their priorities; and in practice, *writing* good requirements has not always been one of them – but fortunately, thanks to more and more research in this field ([2-4], to cite but a few examples), this is tending to quickly change.

This is the case at CNES (Centre National d'Études Spatiales), the French Space Agency, where RE methods and tools are now applied more and more systematically, in particular to ensure the traceability of the thousands of requirements that are written for projects such as satellites and space probes (and to

facilitate their management), but where no rules are imposed for the writing of these requirements – which means that each engineer is free to write them the way (s)he thinks is best. However, aware of the many problems that may arise from poorly written requirements (e.g. requirements that are not clear, ambiguous, and so on [5]), the Quality Assurance Sub-Directorate of CNES asked us to propose a solution to improve their quality.

Different types of solutions already exist and they can be more or less efficient and easy to apply. Some of them consist in formal languages [6], while some others preserve most of the naturalness of natural languages; some are rules that must be learned beforehand by the engineers, and some are tools that can semi-automatically check that these rules were followed, and sometimes provide hints on how to reformulate the problematic phrases or sentences [7].

In the case of CNES, a more “naturalist” [8] approach is preferred, because it would not be too disruptive with regard to the way engineers currently write (and read) specifications and requirements (as previously mentioned, they are written in unconstrained natural language – French or English, depending on the project). In more concrete terms, we wish to design a Controlled Natural Language (CNL), that 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” [9], in order to avoid problems such as ambiguity and vagueness. Since no CNL is currently imposed at CNES, we are perfectly free to either adapt an existing guideline or to create our own set of linguistic rules. Nevertheless, there is a long tradition of CNLs (dating back to the 1930s [10] and renewed in recent decades [11]), including CNLs for requirements specification [12], which can at least partly inspire our work (as we do not intend to create a new one from scratch) and should serve as a basis for our reflection. At the same time, it is of utmost importance for us to design a CNL that would be perfectly adapted to the specification process at CNES (i.e. to what CNES engineers are familiar with). In other words, it is not our intention to propose a generic solution that could be used in all companies, because we assume each of them has its own way of writing specifica-

tions (due to its specific context or history), and it would be too demanding (or even counterproductive) to impose on their technical writers a totally different style of writing requirements. Therefore, we would like to take their specificities (if they exist) into account – which also explains why we do not think an existing CNL could be used as is.

The structure of this paper is as follows: in section II, we specify the aim of this case study and introduce our two research questions; in section III, we describe the design of the experiment we conducted to answer these questions and the reasons that motivated our choice of phenomena to investigate; in section IV, we analyze and comment on some of the results that were obtained; in section V, we point out several limitations regarding these results and finally, we briefly conclude in section VI.

II. AIM AND RESEARCH QUESTIONS

Although the CNL we aim to create will be (to a certain extent) unique, the method we used to design it is reproducible and may thus be of interest to the scientific community. It is based on the analysis of corpora of genuine requirements to find linguistic regularities (see [13] for more details).

In our opinion, an important step in the design of a CNL – albeit often overlooked – is the validation of its rules. Indeed, irrelevant restrictions are at best useless, at worst noxious, and should in any case be avoided. Unfortunately, many CNLs (or similar guidelines) do not specify how the relevance and efficiency of the rules they suggest were evaluated, or even whether it was actually tested before publication. For instance, there were several criticisms of ASD-STE [14] (known at that time as AECMA SE) – one of the most famous CNLs still in use nowadays –, which were the starting point of a series of empirical studies in the 1990s – with mixed results [15-17]. Interesting empirical investigation has also been conducted on corpora of requirements [18]; however, we are not aware of other studies where the judgment of requirements writers on rules was systematically asked for.

Generally speaking, most CNLs often justify why the rules should be followed (e.g. to avoid ambiguities, to make the texts shorter), but they rarely specify how they were elaborated and they do not provide concrete evidence in favor of their effectiveness (is a shorter text really easier to understand? should all potential ambiguities really be avoided at all costs, even though some of them are unlikely to lead to misinterpretations?). Of course, that such evidence is not provided does not necessarily mean it does not exist – but it cannot be taken for granted. As we wish to avoid this pitfall, we believe that every rule of the CNL should be submitted for approval to (a sample of) the users it is intended for, before it is actually proposed or imposed in a real use context. Naturally, this will not prove that it improves comprehension (which would require a much more complex experiment), nor that it is the best possible rule (which is probably impossible to prove), but, at the very least, this will allow us to know more about the users' opinion about it and to avoid absurd recommendations that could be counterproductive.

Unfortunately, we are convinced that it is not sufficient to give the writers a theoretical rule and then to simply ask them if they think that this rule is good or not. Indeed, we know from our experience and discussions with them that their perception of the way they write requirements is quite different from the way they actually write them: they are sometimes positive about the fact that a certain word or structure must be avoided, and still that word or structure is very often found in the texts they (or their colleagues) wrote. (The same phenomenon applies to every one of us about the way we think we speak our own language.) Moreover, it is understandable that it would be hard for them to consider all the different cases where this rule could apply: it could be perfectly fine to follow it in some cases, and irrelevant or simply impossible in other cases.

Therefore, instead of showing them the rules themselves, we preferred to ask them to rate concrete examples of requirements written according to these rules (the rules are not given, so that they are not influenced), in comparison with examples written in a different manner. We believe that if the former are rated significantly higher than the latter, they should indeed probably be preferred, and the rule may be useful. If not, the underlying rule may need to be adapted, or its relevance might be questioned.

These considerations led us to ask the following research questions: (RQ1) for each writing rule we consider, do users think that requirements written according to the recommendation are better? The answer should provide us with valuable information about their perception of the rules. If it is positive, then it is a concrete element in favor of the rule; if it is negative, we will try to determine why. Secondly, we would like to know (RQ2) if experienced writers and laymen share the same opinion. If not, this could indicate that requirements differ from general language.

III. EXPERIMENT

For this exploratory case study (which aims mostly at determining the feasibility of our research procedure [19]), we selected a few recurring writing rules that can often be found in various forms in well-known CNLs; of course, they could be changed or adapted if necessary (in the future, we intend to test our own rules). In order to evaluate them more objectively, we asked French speakers (regardless of their expertise in technical writing and their knowledge of space systems) to rate several possible formulations of the same requirement (some take the recommendations into account, others do not). In doing so, we are collecting subjective data [20].

As we explained in section II, our main goal is to determine whether the requirements that are written in accordance with the recommendations of the CNLs are systematically and significantly preferred (i.e. considered better and clearer) to requirements that express the same need, but are formulated differently (thus disregarding the principles advocated by the controlled languages).

We are also interested in exploring whether the opinion of the experienced technical writers differs significantly from that of speakers who have no (or very little) experience of reading and writing specifications. Indeed, we assume that the experts

(whose opinion matters more to us, since they are the final users of the rules) may be influenced: either indirectly by specifications they previously read, or directly by rules they became aware of elsewhere. Other distinctions could be meaningful (e.g. does the expertise of space projects have some influence?), but will not be considered here.

A. Survey

Our survey was conducted through an online questionnaire made with Qualtrics¹. It is composed of several successive pages and the user has to (anonymously) complete it entirely for the results to be exploitable. After the user has given his/her consent, (s)he is asked to read a short text that gives a basic explanation of what a *requirement* is (it stresses that requirements are mandatory and that they must be correctly interpreted by all stakeholders; we kept it as simple as possible and did not give examples in order to avoid biases); it is obviously intended for non-experts, as experts do not need such explanations.

Then the task is described: there are twenty questions (one per page) and, for each of them, up to six different formulations of the same requirement are proposed. The user must rate them: 5 for the “most appropriate” proposition(s) – i.e. the clearest/least ambiguous one(s) – and 1 for the “worst” proposition(s). To help him/her make his/her decision, (s)he should imagine that (s)he is in charge of building the system in compliance with that requirement. In addition, we asked participants to give the highest score (5) to *at least one* of the propositions on the page (even though (s)he considers they are not perfect), and then to rate the others by comparison, because we wanted to know which ones were preferred; without that precision, we were afraid that in some cases, all propositions would be rated 1 for reasons unrelated to the phenomenon we were investigating. (Experts in particular tend to be very strict on what they consider to be a requirement.)

This rating system has two main advantages over simply ranking the propositions: it allows for two propositions to be given the same score if the user considers that they are roughly equivalent, and we can measure the distance between two scores (4 is less than 5, but still much better than 2 or 1).

After this introduction, the user can start rating the different formulations for each question (see Fig. 1 for an example).

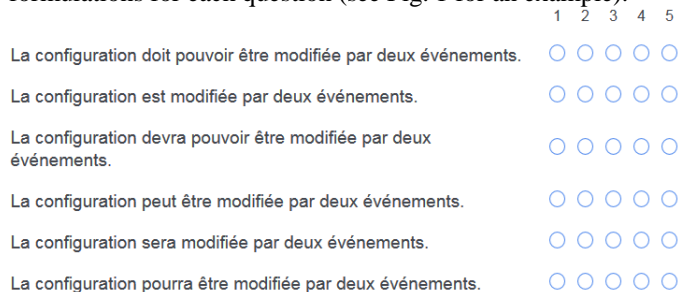


Fig. 1. Example of question

¹ Available at https://toulousepsychology.eu.qualtrics.com/SE/?SID=SV_3wpmNkQbMDpIm9

Each question is shown on a new page (there is a brief reminder of the task on each page). There are no clues about the differences between the propositions. Each proposition on the page must receive a score before the user can go on to the next page; additionally, (s)he may enter some comments in a free-text field (for instance, (s)he can justify his/her choice).

To control order-effect bias [21], the questions are displayed in random order, and so are the propositions on each page. (With the exception of a trap question, where two identical formulations were proposed, to make sure the users give them the same score. Although sixteen participants gave two different scores to the same proposition (a difference of one point for thirteen of them, and two points for the remaining three), the two means were very close: 2.83 and 2.84/5. This suggests that, thanks to the quite high number of participants, the scores are fairly reliable.) The time spent on each page was also recorded, but was not limited; the average time spent on the whole survey was 32 minutes 22 seconds.

Lastly, we asked the users several questions to determine their profile: age - gender - do you work/have you worked for CNES? - do you work/have you worked on space projects? - did you know what a requirement was before you opened this survey? - are you used to *reading* requirements? - are you used to *writing* requirements? - have you been trained to write requirements? - have you worked on Pléiades requirements? [as those requirements served as models for our questions, see below] - have you been trained in linguistics? - is French your first language?

In most cases, if the answer is “yes”, the user may also add some brief comments (for how long, in what context, and so on). (S)he may also add some final remarks and optionally leave an email address.

The link to the survey was sent by email to 135 people at CNES and was also shared on social networks. This allowed us to reach both experts and non-experts.

B. Choice of questions

The most critical part is to decide which questions should be included in the survey, and thus which rules will be tested. As mentioned before, we focused on a few phenomena that are not too highly language-dependent and are often addressed by CNL and/or requirements engineering rules: modalisation (should a modal verb such as “must” always be used? should the requirement be expressed with the present or future tense?), number of sentences (is it better to make two independent sentences, instead of coordinated clauses?), anaphora (should pronouns always be avoided?) and voice (should the active voice always be used?). This is for instance the case in the *Guide for Writing Requirements* by INCOSE (the International Council on Systems Engineering), a CNL for requirements writing that aims “to draw together advice from a variety of existing standards into a single, comprehensive set of rules and objectives” [22] and that served as a basis for this research; but similar rules can be found in many other CNLs.

Each of the twenty questions is used to test one of the four above-mentioned phenomena – but naturally, this is not explained to the users, although it may be quite easy to guess.

Because it was important to us to use examples close to what the engineers of CNES actually write and read (as explained in section I), we carefully selected sentences from a corpus of authentic requirements written in French for two satellites called Pléiades; this means that these examples were used in a real project and are credible for CNES engineers. However, for the sake of simplicity (and also to make sure that we are testing only one phenomenon at a time), we had to shorten and/or to simplify most of them. For example, the requirement “[...] Le plan TC devra respecter les contraintes décrites dans [DR20]” became “Le plan de télécommande devra respecter les contraintes décrites dans le document DR20” (“The remote control plan will have to respect the constraints described in document DR20”). For each of these requirements, we then produced variants to test the phenomenon we are interested in: in this case (modal verb in the future: “devra respecter” / “will have to respect”), we created a proposition with a modal verb in the present (“doit respecter” / “must respect”), another one without a modal verb, and so on. These semi-authentic requirements were then approved by CNES experts. They are all grammatically correct.

1) Modalisation

Thanks to this series of six questions, we hoped to find out whether the users prefer requirements written with or without modal verbs, and whether they prefer requirements in the present tense or in the future tense.

The *Guide for Writing Requirements* by INCOSE (from now on, INCOSE GWR) does not impose the use of a modal verb, but recommends “us[ing] an agreed convention for distinguishing mandatory requirements from non-mandatory”, citing for example the “MoSCoW” convention (Must/Should/Could/Would). Other guidelines recommend using the modal “shall” (or sometimes “must”) to signify that a requirement is mandatory.

All requirements written at CNES are considered mandatory (although metadata can be added to indicate their priority level), so our question is slightly different: should they all be written with a modal verb (typically, the French verb “devoir”) to underline that they are mandatory or, on the contrary, is the modal verb redundant – because we know from the context that they are always mandatory? From our point of view, both possibilities would be acceptable and, in fact, both would be pragmatically equivalent – that is, they have the same meaning and they should have the same effect on the reader. But in practice, do readers prefer requirements with a modal verb to emphasize their necessity, or do they favor shorter ones without a modal verb?

Moreover, should the requirements be written in the present tense (which may make them sound more urgent) or in the future tense (because the system does not yet exist at the time the requirements are written)?

All these options (with vs. without modal, present vs. future) were found in our corpus. We do not think that one of them is intrinsically better than the others, but (like INCOSE GWR) we do think that writers should be consistent when drafting specifications.

For four questions in this category, we proposed four possibilities:

- (a) main verb in the present tense,
- (b) “doit” (± “must”) + infinitive,
- (c) main verb in the future tense,
- (d) “devra” (± “will have to”) + infinitive.

For two other questions, six possibilities had to be rated:

- (a) main verb in the present tense,
- (b) “peut” (± “may”) + infinitive,
- (c) “doit pouvoir” (± “must be able to”) + infinitive,
- (d) main verb in the future tense,
- (e) “pourra” (± “will be able to”) + infinitive,
- (f) “devra pouvoir” (± “will have to be able to”) + infinitive.

These last two questions are a bit different, because they imply that a choice can be made or that a possibility must exist.

2) Number of sentences

Here, we are interested in complex requirements containing multiple clauses. Should such a requirement be written as one long sentence with coordinated clauses or is it better to have multiple sentences (one sentence per clause, often without coordination)?

For the four questions in this category, we proposed a binary choice:

- (a) one complex sentence,
- (b) two sentences.

CNLs in general reject long and complex sentences, and INCOSE GWR is no exception, since it tells the reader to “avoid combinators” – “combinators” being defined as “words that join clauses together, such as ‘and’, ‘or’, ‘then’, ‘unless’”. INCOSE GWR’s authors clearly prefer shorter and simpler (but more numerous) sentences. Although we understand their position, we fear that if these so-called combinators are to be avoided, the logical relation between the two clauses might seem less obvious to the reader.

Note that for option (b), it is up to the reader to decide whether the two sentences should belong to the same requirement or to two distinct requirements, as advised by the principle of atomicity; to keep the experiment simple, we did not take this difference into account (the two sentences are simply visually separated by a line break).

3) Anaphora

Anaphora is a possible source of ambiguity in natural language, since it is not always clear to which antecedent the anaphoric word refers. It is not surprising, then, that many CNLs wish to prevent it (even though it is not always described in proper linguistic terms). One of the rules of INCOSE GWR relates to pronouns: “repeat nouns in full instead of using pronouns to refer to nouns in other requirement statements”. (They also give an example that suggests that anaphoric terms should be avoided even when the antecedent can be found within the same requirement.)

To investigate the effects of this writing rule, there are four questions; for each of them, the user is asked to rate four possibilities:

- (a) definite article + noun (only the head in two cases, the full noun phrase in the other two cases),
- (b) demonstrative + noun (head only),
- (c) demonstrative + noun (full noun phrase),
- (d) personal pronoun.

For another question, there are only three possibilities (because (b) and (c) are identical).

The answers should tell us if it is indeed better to always repeat the noun, or if the use of pronouns (which seems much more natural) may be preferable in some cases.

Of course, we proposed only unambiguous requirements (i.e. if a pronoun is used, there is only one possible antecedent within the requirement), as it is obvious that truly ambiguous requirements must be avoided and would very likely have been rated very low.

4) Voice

This last category also addresses a classical rule of CNLs: the compulsory use of the active voice (or, in other words, the prohibition of the passive voice). One of the rules proposed by INCOSE GWR gives us a clue about the reason why the passive voice is often dispreferred: “use the active voice with the actor clearly identified”. Indeed, while the subject is mandatory in a sentence written in the active voice (e.g. “The user does the action”), the agent is only optional in a sentence written in the passive voice (e.g. “The action is done [by the user]”). (Besides, it is generally thought that the passive voice is harder to process and understand.)

Two questions leave the choice between four possibilities:

- (a) active voice,
- (b) passive voice with agent,
- (c) passive voice without agent,
- (d) the French pronoun “on” (that can be roughly translated in English by the indefinite pronoun *one* or the generic *you*).

Of course, we do not expect the last two options to get good scores, since the agent is not clearly identified (and hence they are less complete than the first two). We were nevertheless interested in the pronoun “on” because it appears to be very common in our corpus, and we hypothesize that it is seen by the writers as a convenient way to use the active voice without specifying the agent.

We also selected in the corpus three requirements written in the passive voice without agent. Only two possibilities remain:

- (a) passive voice without agent,
- (b) the French pronoun “on”.

But in both cases, they can be considered bad requirements; for this reason, the results for these last three questions will not be presented in the next section.

IV. RESULTS

A. Overview

87 people answered our online survey between May and August 2016. The answers they gave to the questions presented in section III.A are shown in Table 1.

TABLE I. PROFILES OF THE USERS WHO ANSWERED THE SURVEY

Question	Answer	
	yes	no
Do you work/have you worked for CNES?	41 (47.13%)	46 (52.87%)
Do you work/have you worked on space projects?	36 (41.38%)	51 (58.62%)
Did you know what a requirement was before you opened this survey?	60 (68.97%)	27 (31.03%)
Are you used to <i>reading</i> requirements?	48 (55.17%)	39 (44.83%)
Are you used to <i>writing</i> requirements?	38 (43.68%)	49 (56.32%)
Have you been trained to write requirements?	18 (20.69%)	69 (79.31%)
Have you worked on Pléiades requirements?	8 (9.20%)	79 (90.80%)
Have you been trained in linguistics?	20 (22.99%)	67 (77.01%)
Is French your first language?	82 (94.25%)	5 (5.75%)

As can be seen from the table, almost half of the answers (41) were given by CNES employees (or former employees), but only 36 people declared that they had previously worked on space projects. 8 people had even worked on Pléiades requirements and might therefore be more familiar with the examples (which were adapted from requirements found in specifications of Pléiades – see section III.B). More than two thirds of the users who filled out the survey declared that they already knew what a requirement was (we were not expecting such a high proportion!); 48 of them (more than half of the respondents) said they were used to reading requirements and 38 said they were used to writing requirements, but only 18 had been specifically trained in requirements writing. Almost all respondents were native French speakers. Interestingly, there were more people who declared they were used to reading requirements than CNES employees, which means that some of them have read specifications written in other companies, and maybe even for other domains (software engineering, for instance).

We decided that the 48 people who answered “yes” to the question “are you used to *reading* requirements?” would be considered “experts” for the analysis of the results and that the remaining 39 people would be considered “non-experts”. Although it could be argued that this is not a fine-grained distinction and that we have to trust these users’ answers (we did not actually check their expertise in any way), it allows us to divide the total population into two comparable groups of similar sizes. We assume that if they have frequently read technical specifications, they probably must have some (passive) skills in interpreting requirements. At the very least, it implies that these users are familiar with this particular kind of texts (which is not the case of the others) and therefore that they were in a better position to complete the pseudo-task that was presented in this survey (as they were asked to imagine that they were in charge of reading and applying the requirements that were presented to them).

We think that this distinction between experts and non-experts will be interesting for the analysis of our results. Indeed, a previous study [23] showed that the experts’ opinion about the technical documents they wrote and the feedback from the users for whom these documents were intended could be very different. In our case (communication from experts to experts), we also expected to get different opinions, due to the specific context of technical writing: because of their experi-

ence and of what they know about requirements writing (and RE in general), experts may have a more informed opinion on the best way to convey information through natural language when writing a specification. Non-experts (a sort of control group composed of participants with various backgrounds), on the other hand, will probably refer to their knowledge of general language and ignore the technical terms unknown to them (since they lack expertise in requirements reading, but also, most likely, in the domain of space projects). This explains why we are interested in both groups, although, of course, the rules are intended for experts only.

B. Analysis

As explained in section III.A, for each question, each respondent had to give each proposition a score between 1 (the worst) and 5 (the best). We could then easily compute the average score of a proposition and compare it with the score of the other propositions for the same question (the higher the score, the more the proposition was preferred to the others).

First of all, we performed a series of statistical tests on the numerical results we obtained. For each question, we used the Friedman non-parametric test to determine whether some propositions were significantly preferred to the others (i.e. if the answers were not randomly distributed). For each proposition, we used the Mann-Whitney non-parametric test to see if the two groups gave significantly different results.

We also grouped the questions by category (see section III.B) to see if more general conclusions could be drawn for certain phenomena (e.g. use of pronouns vs repetition of nouns). Friedman and Mann-Whitney tests were used again, and we then used the Wilcoxon signed-rank test to see if the highest score was significantly higher than the scores of other propositions. The significance level was set at 0.05 (and we applied the Bonferroni correction for the Wilcoxon tests).

Thanks to the results of these tests, we were able to answer our main research question (see section II): for each writing rule we consider, do users think that requirements written according to the recommendation are better? We were also able to answer the secondary research question: do experienced writers and laymen share the same opinion?

In summary, three points of view were compared:

- (a) the recommendations proposed by the CNLs (= the normative point of view),
- (b) the opinion of the experts who answered this survey,
- (c) the opinion of the non-experts who answered this survey.

Besides, these three points of view can be confronted to what is actually found in the corpus of genuine requirements; this allows us, for instance, to see if the recommendations are followed, or if the experts write in accordance with their opinion.

We will now try to see if these points of view converge or not, and we will try to understand why; for this purpose, we were sometimes helped by the comments made by the respondents. An English translation is provided for each question, after the French text.

1) Modalisation

“L’ensemble des TC est / doit être / sera / devra être saisi en base de données.”

“All the remote controls are / **must be** / will be* / will have to be entered in the database.”

- CNLs would recommend (= in bold): “doit être”.
- Experts prefer (= underlined): “doit être” (4.38).
- Non-experts prefer (= in italics): “doit être” (4.15).
- Found in corpus (= *): “sera”.

Although the original requirement was written in the future tense (without a modal), both experts and non-experts prefer the variant with the modal verb “devoir”, which also complies with the recommendations of some CNLs. This is a remarkable case of convergence.

“Le DUPC combine / doit combiner / combinerà / devra combiner en une seule séquence de lecture l’ensemble des ordres de programmation des trois canaux.”

“The DUPC *combines* / **must combine*** / will combine / will have to combine all the programming commands of the three channels in a single reading sequence.”

- CNLs would recommend: “doit combiner”.
- Experts prefer: “doit combiner” (4.40*²).
- Non-experts prefer: “combine” (4.08*).
- Found in corpus: “doit combiner”.

This time, non-experts prefer the present tense without a modal. The original requirement was written with the modal verb “devoir”, that is, according to the recommendations and to the experts’ declared preference.

“Sur ordre du système, le LVC dérouté / doit dérouter / dérouterà / devra dérouter l’exécution du MDP principal vers le MDP secondaire.”

“By system command, the LVC *reroutes** / **must reroute** / will reroute / will have to reroute the execution of the main MDP towards the secondary MDP.”

- CNLs would recommend: “doit dérouter”.
- Experts prefer: “doit dérouter” (4.33*).
- Non-experts prefer: “déroute” (3.85*).
- Found in corpus: “déroute”.

Results are very similar to those of the previous requirement, excepted that the present tense was used in the original requirement.

“Le plan de télécommande respecte / doit respecter / respectera / devra respecter les contraintes décrites dans le document DR20.”

“The remote control plan respects / **must respect** / will respect / will have to respect* the constraints described in document DR20.”

- CNLs would recommend: “doit respecter”.
- Experts prefer: “doit respecter” (4.63*).
- Non-experts prefer: “doit respecter” (4.23*).
- Found in corpus: “devra respecter”.

² * indicates that there is a significant difference between the scores given by the two groups for this option.

The results are very similar to those obtained for the first requirement: experts, non-experts and CNLs all recommend the use of the modal verb “devoir”.

TABLE II. SUMMARY

	<i>General Mean</i>	<i>Mean Experts</i>	<i>Mean Non-Experts</i>	<i>Significant difference E/NE?</i>
doit	4.19	4.43	3.88	yes
devra	3.62	3.70	3.53	no
present	3.30	2.97	3.70	yes
future	3.17	2.96	3.42	yes

There is a significant difference between the first option (“doit” + infinitive) and all the others.

RQ1: yes, experts prefer the requirements written according to the recommendations (modal verb to stress that it is mandatory). Note however that all variants were found in the corpus (although they were written by experts).

RQ2: yes, both experts and non-experts prefer to use the modal verb “devoir” (although non-experts may sometimes omit it).

“L’opérateur a le choix du mode de saisie : il rentre / peut rentrer / doit pouvoir rentrer / rentrera / pourra rentrer / devra pouvoir rentrer l’adresse directement ou indirectement en précisant le nom d’un label.”

“The operator has a choice of input mode: he enters / *can enter* / **must be able to enter** / will enter / will be able to enter* / will have to be able to enter the address directly or indirectly by specifying a label name.”

- CNLs would recommend: “doit pouvoir rentrer”.
- Experts prefer: “doit pouvoir rentrer” (4.17*).
- Non-experts prefer: “peut rentrer” (3.92*).
- Found in corpus: “pourra rentrer”.

Experts still prefer requirements written with “devoir”, even though it is longer. Non-experts, however, prefer to omit it in this case.

“La configuration est / peut être / doit pouvoir être / sera / pourra être / devra pouvoir être modifiée par deux événements.”

“The configuration is / *may be** / **must be able to be** / will be / will be able to be / will have to be able to be / modified by two events.”

- CNLs would recommend: “doit pouvoir être”.
- Experts prefer: “doit pouvoir être” (4.17*).
- Non-experts prefer: “peut être” (3.79*).
- Found in corpus: “peut être”.

The results are very similar: experts prefer to use “devoir”, even though there are four different verbs because of the passive voice (“doit pouvoir être modifiée”); non-experts prefer to omit “devoir”.

TABLE III. SUMMARY

	<i>General Mean</i>	<i>Mean Experts</i>	<i>Mean Non-Experts</i>	<i>Significant difference E/NE?</i>
doit pouvoir	3.69	4.17	3.10	yes
peut	3.32	2.87	3.86	yes
pourra	3.10	2.68	3.62	yes
devra pouvoir	2.96	3.25	2.60	yes
present	2.91	2.62	3.27	yes
future	2.71	2.47	3.00	yes

There is a significant difference between the first option (“doit pouvoir” + infinitive) and all the others but the second one (“peut” + infinitive).

RQ1: yes, experts prefer the requirements written according to the recommendations (modal verb to stress that it is mandatory), even though this increases the length of the verb group.

RQ2: no, this time non-experts prefer to avoid the modal verb “devoir”, probably because the structure is too long and not elegant.

2) Number of sentences

“L’automate de la fonction IOS passe à ON et l’ensemble des traitements sur les équipements senseur solaire, magnétomètres, magnéto-coupleurs, roue à réaction est séquencé. / L’automate de la fonction IOS passe à ON.
L’ensemble des traitements sur les équipements senseur solaire, magnétomètres, magnéto-coupleurs, roue à réaction est séquencé.”

“The automaton of the IOS function goes to ON and all the processing on the solar sensor, magnetometer, magneto-coupler, and reaction wheel equipment is sequenced. / The automaton of the IOS function goes to ON.
All the processing on the solar sensor, magnetometer, magneto-coupler, and reaction wheel equipment is sequenced.”

- CNLs would recommend: two sentences.
- Experts prefer: one sentence (3.88).
- Non-experts have no preference (4.18).
- Found in corpus: one sentence.

The results of the statistical test were not significant for this question.

“Il est possible d’importer et d’exporter toutes les règles de transfert déclarées. / Il est possible d’importer toutes les règles de transfert déclarées.
Il est possible d’exporter toutes les règles de transfert déclarées.”

“It is possible to import and export all the declared transfer rules. / It is possible to import all the declared transfer rules.
It is possible to export all the declared transfer rules.”

- CNLs would recommend: two sentences.
- Experts prefer: two sentences (4.13*).
- Non-experts prefer: one sentence (5.00*).
- Found in corpus: one sentence.

Experts prefer to write to distinct sentences (but the difference is negligible: 0.01), while non-experts unanimously choose the requirement written with a single sentence.

“Les champs SM_ID et FM_ID seront extraits à partir de la BDS. / Le champ SM_ID sera extrait à partir de la BDS.
Le champ FM_ID sera extrait à partir de la BDS.”

“The SM_ID and FM_ID fields will be extracted from the BDS. / The SM_ID field will be extracted from the BDS.
The FM_ID field will be extracted from the BDS.”

- CNLs would recommend: two sentences.
- Experts prefer: one sentence (4.33*).
- Non-experts prefer: one sentence (4.97*).
- Found in corpus: one sentence.

Here, experts and non-experts both prefer the shorter solution, as it seems that the two operations may reasonably be expressed within the same requirement.

“Si le transfert d'un fichier de télémétrie échoue après le nombre de tentatives fixées, le fichier sur le destinataire est effacé et l'information est remontée aux IHM. / Si le transfert d'un fichier de télémétrie échoue après le nombre de tentatives fixées, le fichier sur le destinataire est effacé.
L'information est remontée aux IHM.”

“If the transfer of a telemetry file fails after the set number of attempts, the addressee file is deleted and the information is reported to the HCI. / If the transfer of a telemetry file fails after the set number of attempts, the addressee file is deleted
The information is reported to the HCI.”

- CNLs would recommend: two sentences.
- Experts prefer: one sentence (4.10).
- Non-experts prefer: one sentence (4.62).
- Found in corpus: one sentence.

The complex sentence with two coordinated clauses is preferred by both the experts and the non-experts.

TABLE IV. SUMMARY

	<i>General Mean</i>	<i>Mean Experts</i>	<i>Mean Non-Experts</i>	<i>Significant difference E/NE?</i>
one sentence	4.37	4.11	4.69	yes
two sentences	3.72	3.92	3.48	yes

There is a significant difference between the first option (one sentence) and the second one (two sentences).

RQ1: **no**, although CNLs and guidelines suggest that it is better to write multiple requirements instead of coordinating clauses, a single sentence seems sometimes preferable. However, it should be noted that the examples proposed here represent quite different cases. We would not dare to draw general conclusions at this point.

RQ2: yes, both experts and non-experts prefer to write a single sentence, even though it is longer and more complex.

3) Anaphora

“Le paquet cyclique ne sera généré que si le paquet / ce paquet / **ce paquet cyclique** / *il** est activé par le LVC.”

“The cyclic packet will only be generated if the packet / this packet / this cyclic packet / it is activated by the LVC.”

- CNLs would recommend: “ce paquet cyclique”.
- Experts prefer: “il” (4.58*).
- Non-experts prefer: “il” (4.85*).
- Found in corpus: “il”.

Both experts and non-experts prefer to use the unambiguous pronoun “il” instead of repeating the noun (as recommended by CNLs).

“La configuration du système est stockée à bord. La configuration / Cette configuration / **Cette configuration du système** / *Elle** ne peut être modifiée que par deux événements : événement 1 et événement 2.”

“The configuration of the system is stored onboard. The configuration / This configuration / This system configuration / It can be modified by two events only: event 1 and event 2.”

- CNLs would recommend: “cette configuration du système”.
- Experts prefer: “elle” (4.12).
- Non-experts prefer: “elle” (4.38).
- Found in corpus: “elle”.

Again, all users prefer the more natural use of a personal pronoun, instead of a repetition.

TABLE V. SUMMARY

	<i>General Mean</i>	<i>Mean Experts</i>	<i>Mean Non-Experts</i>	<i>Significant difference E/NE?</i>
pronoun	4.47	4.35	4.62	no
demonstrative + head	3.72	3.83	3.59	no
demonstrative + full NP	3.18	3.64	2.63	yes
definite	2.60	2.54	2.68	no

There is a significant difference between the first option (pronoun) and all the others.

RQ1: **no**, CNLs recommend that pronouns should be avoided, but we see from these examples that if they are not ambiguous, experts think they are the best option.

RQ2: yes, both experts and non-experts prefer to use the pronoun.

“En cas de réception d'un fichier vide, le fichier vide / ce fichier / ce fichier vide / il n'est pas transmis.”

“If an empty file is received, the empty file / *this file* / **this empty file** / it is not forwarded.”

- CNLs would recommend: “ce fichier vide”.
- Experts prefer: “ce fichier” (4.10).
- Non-experts prefer: “ce fichier” (4.36).
- Found in corpus: “celui-ci” (not proposed in the survey).

Here, both experts and non-experts prefer to repeat the noun with a demonstrative, which is more in line with CNL recommendations.

“La liste des TCD est définie dans le document C12 ; la liste des TCD / cette liste / cette liste des TCD / elle est exhaustive.”

“The list of TCD is defined in document C12; the list of TCD / *this list* / **this list of TCD** / it* is exhaustive.”

- CNLs would recommend: “cette liste des TCD”.
- Experts prefer: “cette liste” (4.10).
- Non-experts prefer: “cette liste” (4.51).
- Found in corpus: “elle”.

The results are very similar: the head of the noun phrase is repeated with a demonstrative.

TABLE VI. SUMMARY

	<i>General Mean</i>	<i>Mean Experts</i>	<i>Mean Non-Experts</i>	<i>Significant difference E/NE?</i>
demonstrative + head	4.25	4.10	4.44	no
pronoun	3.63	3.60	3.65	no
demonstrative + full NP	3.50	3.81	3.12	yes
definite	2.96	3.21	2.65	yes

There is a significant difference between the first option (demonstrative + head) and all the others.

RQ1: yes, in this case, experts prefer to repeat the noun (but not in full – only the head of the noun phrase).

RQ2: yes, both experts and non-experts prefer to repeat the head. (It remains unclear for us why the results of these two questions are different from those we obtained for the first two questions, where pronouns were preferred.)

“L'utilisateur doit savoir à tout moment si l'utilisateur / si **cet utilisateur** / s'*il** est connecté sur le serveur GIDE nominal ou redondant.”

“The user must know at any moment if the user / this user / he is connected to the nominal or redundant GIDE server.”

- CNLs would recommend: “cet utilisateur”.
- Experts prefer: “il” (4.96).
- Non-experts prefer: “il” (4.95).
- Found in corpus: “il”.

Despite the recommendations of several CNLs, almost all respondents prefer the use of a pronoun in this case. Repeating the noun sounds unnatural.

4) Voice

“L'opération n'est nécessaire que si le LVC autorise la surveillance fonctionnelle / la surveillance fonctionnelle est autorisée par le LVC / la surveillance fonctionnelle est autorisée / on autorise la surveillance fonctionnelle.”

“The operation is only necessary if **the LVC authorizes functional surveillance** / *functional surveillance is authorized by the LVC** / functional surveillance is authorized / one authorizes functional surveillance.”

- CNLs would recommend: “le LVC autorise”.
- Experts prefer: “est autorisée par le LVC” (4.42).
- Non-experts prefer: “est autorisée par le LVC” (4.51).
- Found in corpus: “est autorisée par le LVC”.

Both experts and non-experts prefer the requirement written in the passive voice, even though most CNLs would not recommend it. A possible explanation for this may be that the passive voice allows for a focus on the phrase “surveillance fonctionnelle”, which is probably more important than the agent here.

“Le CCC contrôlera le vidage des tables. / Le vidage des tables sera contrôlé par le CCC. / Le vidage des tables sera contrôlé. / On contrôlera le vidage des tables.”

“**The CCC will monitor table dump**. / *Table dump will be monitored by the CCC**. / Table dump will be monitored. / One will monitor table dump.”

- CNLs would recommend: “le CCC contrôlera”.

- Experts prefer: “le CCC contrôlera” (4.56).
- Non-experts prefer: “sera contrôlé par le CCC” (4.62).
- Found in corpus: “sera contrôlé par le CCC”.

This time, experts do follow the CNLs, since they prefer the active voice (while the non-experts still prefer the passive voice, which was used to write the original requirement).

TABLE VII. SUMMARY

	<i>General Mean</i>	<i>Mean Experts</i>	<i>Mean Non-Experts</i>	<i>Significant difference E/NE?</i>
passive with agent	4.45	4.35	4.56	no
active	4.26	4.32	4.18	no
passive without agent	2.60	2.45	2.78	no
“on”	1.86	1.64	2.14	yes

RQ1: **no**, CNLs impose the active voice (or prohibit the passive voice), but from these two examples we can see that requirements written using the passive voice obtained a slightly higher score than their counterparts in the active voice. However, the difference is not significant; the active voice and the passive voice seem to be equally preferred by the experts.

RQ2: yes, the same results apply for the non-experts too.

V. THREATS TO VALIDITY

In this section, we would like to quickly summarize the main weak points of our experiment (some of them were already mentioned before).

First of all, this online survey allowed us to gather the opinion (i.e. preferences) of a limited number of users – but not to test their actual comprehension of the requirements. A comprehension test would be much more complex. Ideally, it should also take into account the context, which was not really the case here.

In addition, we did not test the expertise of the respondents: we categorized them as either “experts” or “non-experts” only on the basis of their anonymous answer to a question. This distinction was certainly convenient for us, but should not be considered as an absolute truth.

Lastly, through this experiment, we wanted to compare the scores obtained by different propositions inspired by recurring rules found in CNLs. But better rules – and thus better propositions – than the ones we tested here may well exist. In particular, it should be noted that we voluntarily chose to test only simple cases, and it is likely that more complex situations would require further examination.

Still, we believe that this small methodological contribution allowed us to objectively evaluate several rules often found in CNLs without concrete justification.

VI. CONCLUSION

In this paper, we presented the final step of our methodology for defining sound linguistic rules for requirements writing. Indeed, we think that many rules found in existing guidelines or controlled natural languages for technical writing (such as

the *Guide for Writing Requirements* by INCOSE) lack empirical evidence.

Therefore, we decided to test several of them through an online survey, which was filled out by experienced writers but also by laymen. We asked them to rate variants of the same requirement, some written according to the recommendations and others not, because we wanted to verify that the former are really preferable to the latter. To make sure our examples were plausible, they were extracted and adapted from authentic requirements written and used at the French Space Agency.

We found that in some cases, the engineers agree with the recommendations: for instance, it seems relevant to impose a modal verb (such as “must” or “shall” in English, or “devoir” in French) in each mandatory requirement.

However, requirements written according to these rules are not always preferred. Some rules suggest that pronouns should be avoided, but it appears that requirements written using pronouns – as long as they are not ambiguous – are sometimes considered better, even by experts. Consequently, these rules could probably be refined to allow for more flexibility. Similarly, there are cases where the use of the passive voice with an agent seems justified (or at least innocuous) and where it would be counterproductive to forbid it.

In the future, we intend to perform a similar kind of experiment to test our own set of rules. Other analyses could also be performed on the results; for example, it would be interesting to see if experts used to reading and experts used to writing requirements share the same opinion.

ACKNOWLEDGMENT

We would like to gratefully acknowledge Daniel Galarreta, Jean-François Gory and Nicolas Deslandres (among other people at CNES) for their strong involvement in this doctoral research, which was granted by CNES and the Regional Council of Midi-Pyrénées (France). We also wish to thank Pierre-Vincent Paubel for his technical assistance, Aurélie Mouneyrac for her valuable advice on which statistical tests to perform, the reviewers of this paper for their valuable comments and all the people who kindly dedicated some of their time to answer our survey.

REFERENCES

- [1] I. Sommerville and P. Sawyer, *Requirements engineering: a good practice guide*. John Wiley & Sons, Inc., 1997.
- [2] A. Mavin, P. Wilkinson, A. Harwood, and M. Novak, “Easy Approach to Requirements Syntax (EARS),” in *2009 17th IEEE International Requirements Engineering Conference*, 2009, pp. 317–322.
- [3] I. Hooks, “Writing good requirements,” in *INCOSE International Symposium*, 1994, vol. 4, pp. 1247–1253.
- [4] I. F. Alexander and R. Stevens, *Writing better requirements*. Pearson Education, 2002.
- [5] F. Chantree, B. Nuseibeh, A. de Roeck, and A. Willis, “Identifying Noxious Ambiguities in Natural Language Requirements,” in *14th IEEE International Requirements Engineering Conference (RE’06)*, 2006, pp. 59–68.
- [6] B. Meyer, “On Formalism in Specifications,” *IEEE Softw.*, vol. 2, no. 1, pp. 6–26, Jan. 1985.
- [7] N. Carlson and P. Laplante, “The NASA automated requirements measurement tool: a reconstruction,” *Innovations in Systems and Software Engineering*, vol. 10, no. 2, pp. 77–91, Sep. 2013.
- [8] P. Clark, W. R. Murray, P. Harrison, and J. Thompson, “Naturalness vs. Predictability: A Key Debate in Controlled Languages,” in *Controlled Natural Language*, N. E. Fuchs, Ed. Springer Berlin Heidelberg, 2010, pp. 65–81.
- [9] T. Kuhn, “A Survey and Classification of Controlled Natural Languages,” *Computational Linguistics*, vol. 40, no. 1, pp. 121–170, 2014.
- [10] C. K. Ogden and G. Halász, *Basic English*. K. Paul, 1935.
- [11] C. Kamprath, E. Adolphson, T. Mitamura, and E. Nyberg, “Controlled language for multilingual document production: Experience with Caterpillar technical English,” in *Proceedings of the Second International Workshop on Controlled Language Applications*, 1998, vol. 146.
- [12] N. Fuchs and R. Schwitter, “Controlled English for Requirements Specifications,” *IEEE Computer Special Issue on Interactive Natural Language Processing*, 1996.
- [13] A. Condamines and M. Warnier, “Towards the creation of a CNL adapted to requirements writing by combining writing recommendations and spontaneous regularities: example in a space project,” *Language Resources and Evaluation*, pp. 1–27, 2016.
- [14] AeroSpace and Defence Industries Association of Europe, “Simplified Technical English. Specification ASD-STE100. International specification for the preparation of maintenance documentation in a controlled language. Issue 4.” 2007.
- [15] S. K. Shubert, J. H. Spyridakis, H. K. Holmback, and M. B. Coney, “The comprehensibility of simplified English in procedures,” *Journal of technical writing and communication*, vol. 25, no. 4, pp. 347–369, 1995.
- [16] S. Chervak, C. G. Drury, and J. P. Ouellette, “Simplified English for aircraft workcards,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1996, vol. 40, pp. 303–307.
- [17] K. M. Stewart, “Effect of AECMA simplified English on the comprehension of aircraft maintenance procedures by non-native English speakers,” *University of British Columbia*, 1998.
- [18] J. Krisch and F. Houdek, “The myth of bad passive voice and weak words an empirical investigation in the automotive industry,” in *Requirements Engineering Conference (RE), 2015 IEEE 23rd International*, 2015, pp. 344–351.
- [19] R. K. Yin, *Applications of Case Study Research*, 2nd ed. Thousand Oaks (CA): Sage Publications, 2003.
- [20] F. Ganier, “Évaluer l’efficacité des documents techniques procéduraux : un panorama des méthodes,” *Le travail humain*, vol. 65, no. 1, p. 1, 2002.
- [21] W. D. Perreault, “Controlling order-effect bias,” *The Public Opinion Quarterly*, vol. 39, no. 4, pp. 544–551, 1975.
- [22] International Council on Systems Engineering, “Guide for Writing Requirements.” INCOSE, 2011.
- [23] M. D. T. De Jong and Leo R. Lentz, “Expert Judgments Versus Reader Feedback: A Comparison of Text Evaluation Techniques,” *Journal of Technical Writing and Communication*, vol. 26, no. 4, pp. 1–1, Jan. 1996.