Error Typology and Remediation Strategies for Requirements Written in English by Non-Native Speakers

Marie Garnier[1], Patrick Saint-Dizier[2]
[1]CAS, Université Toulouse 2 Jean Jaurès, 5 allées A. Machado, 31058 Toulouse France
[2]IRIT-CNRS, 118 Route de Narbonne 31062 Toulouse France
E-mail: mgarnier@univ-tlse2.fr, stdizier@irit.fr

Abstract

In most international industries, English is the main language of communication for technical documents. These documents are designed to be as unambiguous as possible for their users. For international industries based in non-English speaking countries, the professionals in charge of writing requirements are often non-native speakers of English, who rarely receive adequate training in the use of English for this task. As a result, requirements can contain a relatively large diversity of lexical and grammatical errors, which are not eliminated by the use of guidelines from controlled languages. This article investigates the distribution of errors in a corpus of requirements written in English by native speakers of French. Errors are defined on the basis of grammaticality and acceptability principles, and classified using comparable categories. Results show a high proportion of errors in the Noun Phrase, notably through modifier stacking, and errors consistent with simplification strategies. Comparisons with similar corpora in other genres reveal the specificity of the distribution of errors in requirements. This research also introduces possible applied uses, in the form of strategies for the automatic detection of errors, and in-person training provided by certification boards in requirements authoring.

Keywords: requirements authoring, error analysis, error correction

1. Motivation

In most international industries, English is the main language of communication for technical documents. Product specifications, business rules or requirements form a specific genre in technical documentation (all instances of this genre will hereafter be called 'requirements' for the sake of simplicity). These documents are designed to be easy to read and as efficient and unambiguous as possible for their users and readers. They must leave no space for personal interpretation.

In the case of international industries based in non-English speaking countries, the professionals in charge of writing requirements are very often non-native users of English, and these professionals rarely receive adequate training in the use of English for this task. As a result, requirements can contain a relatively large diversity of lexical and grammatical errors. One of the standard ways of simplifying the task of writing requirements and harmonizing results is the use of a chosen controlled natural language for any particular business or industry. However, resorting to the use of controlled languages only offers a limited solution to the linguistic problems encountered by non-native English users: having been designed for native or near-native users of the language, controlled languages do not address the proper use of English grammar and lexicon, nor do they issue warnings about the loss of intelligibility that may occur as a consequence of lexical and grammatical errors (even though they do issue such warnings on other aspects linked to intelligibility). As a result of this situation, requirements may include errors that decrease readability, increase the risk of ambiguity and misinterpretation, and possibly lead to problems of misconception and lack of productivity and efficiency.

Recent years have seen a rise in concern for good practices in requirements writing, and the corresponding development of training programs and certifications that go beyond the application of controlled languages (e.g. the *International Requirements Engineering board*, with the *Certified Professional for Requirements Engineering* certification).

We feel it is now crucial to take into account specific authoring problems encountered by non-native users of English when they produce requirements, in order to guarantee a satisfactory level of language quality. In this project, we propose to document and analyze linguistic errors found in a representative corpus of requirements written in English by native speakers of French, and present the different ways in which the results from this research can be applied in order to improve language use in requirements.

2. Controlled Languages and the Language of Requirements

Various sectors of industry and business have seen the creation of guidelines for writing documents following controlled languages issued by user consortium and companies, resulting in a large diversity of specifications (e.g. *ASD-Simplified Technical English*, *SLANG*, *Attempto simplified English* (Kuhn et al., 2013), see also (Wyner et al., 2010) for some examples of implementation). Norms tend to emerge, such as *Semantics of Business Vocabulary and Business Rules* (*SBVR*) *Structured English* for business rules and *International Council on Systems Engineering-Object Management Group* for requirements (Hull et al. 2011). The reader can consult a detailed synthesis and classification of CNL principles and projects in (Kuhn, 2014), which also investigates grammars for CNL. Various companies in the sectors of space and aeronautics carry out further specialized analysis for critical systems. Additionally, *RUBRIC, a Flexible Tool for Automated Checking of Conformance to Requirement Boilerplates*, is dedicated to requirement control.

The authoring principles and constraints developed for requirements and similar genres generally include dictionaries as well as rules limiting the types of grammatical structures that can be used. For example, passive forms and forms expressing the future are prohibited in most norms dealing with requirement authoring (e.g. INCOSE, IEEE830), and the use of modal auxiliaries is highly constrained, with some modals being prohibited while others being only used with a specific meaning (e.g. *shall* and *must*). At the lexical level, the number of verbs allowed for one domain is fairly limited, and the included verbs must be unambiguous and precise. The use of deverbal nouns is also not recommended. Complex formulations are to be rephrased, following precise recommendations (Grady, 2006), (Saint-Dizier, 2014), (Fuchs, 2012). Stylistic constraints also abound, whether they are of a general nature or imposed by a given company, domain or genre. Given the complexity and range of these recommendations, it is not surprising on the one hand that native speakers have difficulties expressing requirements that comply with the recommendations, and on the other hand that non-native users struggle even more given the added difficulty of writing in a second language. Requiring the use of controlled and simplified languages does not guarantee that the documents will be less ambiguous and better written where non-native users are concerned (see Grady, 2006). To overcome some of these difficulties, predefined patterns (also called *boilerplates*) have been defined, but from authors' experiences, it seems that they are not flexible enough to adequately capture the level of complexity necessary in some requirements.

The need for assisting technical writers in their task (in particular those producing specifications) has motivated the development of the LELIE project (Barcellini et al., 2012), (Saint-Dizier, 2014). LELIE is a system that detects several types of error in technical documents (see Table 1 below), at any point in the authoring and revision stages. LELIE produces *alerts* that flag terms, expressions or constructions that need various forms of improvements. LELIE can also be tailored to the specific constraints of an industry or business, for example in the form of controls on style and the use of business terms.

LELIE and the experiments reported below were developed on the logic-based <TextCoop> platform (Saint-Dizier, 2012). LELIE is fully implemented in Prolog; its kernel is freely available for French and English. The output of LELIE is the original text with annotations.

Table 1 below shows a selection of the major errors found by LELIE. The corpus for this study includes about 35000 words of proofread technical documents from three companies (kept anonymous). Alert numbers relate the number of errors observed all over the corpus, doubles included, since the same error can be quite frequent and generates the same problems with every instance.

It is important to note that while following authoring norms is a necessity, this does not mean that texts become understandable. Norms are only one means to reach an authoring level that is simple and accessible to users.

In the chart below, the category of *Fuzzy lexical items* includes, for example, fuzzy determiners (*most, some*), fuzzy adverbs (*regularly, cautiously*), and fuzzy adjectives (*standard*). These terms may be more or less fuzzy depending on the context. The category of *Deverbal nouns* refers to the use of nouns derived from a verb instead of the verb itself, for example the use of the noun *installation* instead of the verb *install*. The use of negation is a delicate problem, since it is not allowed in most norms, but sometimes cannot be avoided without complex developments (e.g. *do not throw in the sewer*). The category of *Complex discourse structures* is related to sentence complexity metrics: requirements are supposed to be short and easy to read. It follows that the number of discourse structures in a requirement must be controlled. Passives are not necessarily errors, but, when possible, it may be recommended to use the direct voice.

Error type	**Nb. of alerts** **for 35000 words**
Fuzzy lexical items	182
Deverbal nouns	80
Pronouns with unclear reference	54
Negation	126
Complex discourse structures	127
Complex coordinations	57
Heavy N+N or noun complements	135
Passives	92
Sentences too complex	261
Incorrect references to sections or figures	57

Table 1. Selection of LELIE alerts in technical texts

These results show that there is an average of about one alert issued every 3 lines of text (about every 50 words), which is a very high rate. Alerts are issued even on documents that have been proofread by several trained writers. These results show that the use of controlled languages is not a catch-all solution to linguistic problems in requirement engineering. There appears to be space for tools that address linguistic problems from another angle, and in the case of our project by targeting errors produced by non-native users writing requirements in English, and more specifically native speakers of French.

3. Research Methodology and Background

3.1 Error Analysis

Our research relies on the manual analysis of corpora of requirements written in English by native speakers of French. Research in Second Language Acquisition has shown the influence of a speaker's first language on their use of a second language (e.g. Jarvis and Pavlenko, 2007), notably through the phenomenon of transfer, or cross-linguistic influence. Language transfer plays a major role in error production, and gives precious indications as to the author's intended meaning and possible remediation. Choosing to focus on requirements written by speakers of French means we can better tailor corrections and recommendations to this type of user. However, our method can be reproduced for other languages (e.g. Spanish or Thai speakers writing requirements in English).

We use the methodology of *error analysis*, a research method initially developed and used in the domain of Second Language Acquisition (Corder, 1981), and which is recognized as a valid way of gathering data about language errors (Cook, 1993). The analysis is conducted by a single trained linguist specialized in English grammar with a background in research on linguistics-based automatic grammar checking for English. For this project, the different steps are performed manually, due to the complexity of the task. Requirements are read and screened for errors, which are then categorized.

It is well-known that, when performing complex linguistic tasks with natural language, it is always preferable to have several researchers working in parallel in order to be able to compare findings through the *kappa* test. However, from a practical point of view, it is very difficult to find the resources to do so.

We are aware of the fact that there is already a large pool of research on L2 learners/users' errors, especially for English, as well as research on L2 corpora (e.g. Granger et. al., 2009). One of the objectives of our research will therefore be to find out whether the specificity of the task of writing requirements in English for industrial purposes warrants a specific treatment, or can be dealt with in the broader realm of second language acquisition research.

To find out whether errors found in requirements are only representative of errors produced by French speakers of a similar level in English in different types of writing, or highly specific to requirements writing, we compared our error corpus from requirements to a corpus of errors found in L1 French student essays in English, and to a corpus of errors found in scientific papers written in English by French native speakers.

3.2 Defining and Classifying Errors

Before presenting how errors are analyzed in our research, we need to clarify our definition of the concept, from which the criteria used for the detection of errors will stem.

At its most basic, an error is defined as "an unsuccessful bit of language" (James, 1998). The most common criterion for declaring a segment of language "unsuccessful" is grammaticality, that is to say whether or not the segment follows the rules of grammar. Acceptability is another criterion, which focuses on whether or not the segment might be produced by a native speaker in an appropriate context (Lyons, 1968).

The concepts of *competence* and *performance* have also been explored in the definition of errors. Competence errors are attributable to a lack of knowledge in the language, while performance errors are due to external factors, such as lack of attention, stress or fatigue (Corder, 1967). However, researchers have highlighted the fact that even though this distinction is theoretically relevant, it is practically impossible to distinguish competence errors from performance errors (Thouësny, 2011).

We adapt these criteria to the objectives of our project as well as to the nature of the documents in our corpus to define the types of segment identified as errors:

- Grammaticality: segments that don't follow morpho-syntactic rules (e.g. plural agreement), lexico-syntactic rules (e.g. choice of preposition after a verb), or other basic grammar rules are identified as errors.
- Acceptability: requirements need to be written as clearly and as intelligibly as possible so as to eliminate ambiguities and confusions, which leads us to stretch the notion of acceptability to include clarity and intelligibility. Segments that introduce ambiguities, lack clarity, or require an effort from the user to understand the intended meaning (e.g. stacking of modifiers in composite nominals) are identified as errors. In this case there is an overlap between errors as we define them and the structures that are prohibited in some controlled languages.
- Performance errors/competence errors: spelling errors, since they are most likely due to a lack of attention, (especially since in other instances the correct spelling is used) and can easily be eliminated with the use of a spellchecker, are identified as performance errors. However, their presence in our corpus warrants their inclusion in our categories.

Error categories usually rely on a set of criteria, used in combination or alone:

- the linguistic domain of the error (e.g. morphology, syntax, spelling, etc.);
- the part of speech bearing the error or that needs to be modified to correct the error;
- the linguistic system linked with the error (e.g. agreement, verb complementation, etc.);
- the description of surface phenomena (e.g. word omission, extra word, word order, etc.).

One of the objectives of classifying errors in our research is to obtain precise and comparable data about errors found in requirements. The type of classification that we estimate as the most adequate in order to yield comparable data should include the rank of the linguistic unit that needs to be taken into account for the error to become apparent (i.e. the *error domain*, Lennon, 1991), usually *Noun Phrase*, *Verb Phrase*, *Clause*, etc., as well as a number of sublevel categories giving more detailed information, such as *Missing article*, *Preposition selection*, etc.

As a supplement to these domains and categories, we also include categories such as *Lexicon*, *Continuity and coherence of discourse*, and *Punctuation*, since errors in these categories are present in the corpus and have an effect on the overall quality and intelligibility of the requirements. Additionally, we have chosen to flag all errors, regardless of whether they could be successfully detected and corrected by a standard grammar checker, in order to give a neutral snapshot of the distribution of errors in requirement. Moreover, as stated below, the requirements used in our corpus have already gone through the initial stages of editing, and the presence of these errors seems to indicate that the use of grammar checkers is not a systematic practice in all industries.

4. Corpus Compilation and Validation

Ideally, a representative corpus of requirements should include the following specificities:

- requirements in the corpus must represent various professional activities where requirements are relevant: product definition, functions, properties and design, maintenance, production launch, safety;
- requirements must come from various industrial areas;
- requirements must follow various kinds of authoring recommendations, business style and format imposed by companies;
- requirements must vary in complexity, and must have been produced by different types of staff (technicians, stakeholders, engineers), representing steps in the proofreading and certification process.

We have strived to compile a corresponding corpus, but the sensitive nature of the documents concerned means that some aspects are difficult to control for. For example, for some portions of the corpus, it is difficult to know how many engineers have made corrections to them.

We compiled a corpus of requirements written in English by French speakers from documents coming from four areas of industry (space exploration, aircraft maintenance, bio-pharmacy, and collaborative systems for computer networks). The corpus comes from four companies kept anonymous at their request.

At the time of writing, our corpus is composed of the following resources, for a total of 1140 requirements:

- space exploration: 470 requirements; some of these requirements are just 2 lines long whereas others can be half a page long, with equations, charts, figures, etc.,
- aircraft maintenance: 213 requirements; requirements are between 2 to 6 lines long in general,
- bio-pharmacy: 234 requirements; requirements generally include one or two sentences, which are relatively complex in terms of syntax and style,
- collaborative systems: 223 requirements; requirements are quite short, stored in a chart with cross references, level of validation, etc.

These corpora have very different characteristics that offer an adequate representation of how French native speakers write in technical English:

- all texts were written by several authors, from three to six; these texts have therefore been produced by at least 15 different authors,
- all authors are native speakers of French, with intermediate to advanced use of English, with large variations,
- some of these texts have been proofread and validated whereas others are drafts, possibly hastily written,
- authors follow different types of authoring guidelines; the layout is quite diverse and includes in particular charts and diagrams with text, besides simple textual forms;
- these texts represent situations ranging from average to complex, in which requirements are not straightforward to write.

5. Results

The following results were obtained from the analysis of a subset of our corpus, including 550 requirements taken from the "space exploration" section of the corpus (Corpus A) and from the "aircraft maintenance" section (Corpus B).

Table 2 presents a synthetic list of categories of errors found in the corpus, along with authentic examples. Errors are distributed in 5 error domains and 1 *Other* category, with 21 sublevel categories in total. In the case of sublevel categories that have an extra level of detail (ex. *NP Modification*), these are indicated in parentheses next to the sublevel category, but only the total number is shown with the example representing the most numerous type. Errors segments are presented in context and their specific location is underlined for the sake of clarity. The source of the requirement is indicated by a letter preceding the example (A or B for the corresponding corpus). The percentages in the middle column show the proportion of errors from each category in the corpus of errors, with the 6 sublevel categories that each account for at least than 5 % of errors shown in bold.

We found a total of 190 errors in 131 requirements, meaning that roughly 1 in 4 requirements contained at least one error, and 1 in 10 contained more than one error, usually in different categories.

Errors are most often found in the domain of *Noun Phrase*, which totals 58 % of errors, and the category *Other*, with 25.7 % of errors. In each of them, one sublevel category holds the majority of errors: most errors in the *Noun Phrase* category are linked to modification (and mostly to modifier stacking, as shown in the example), with 32.8 % of errors in total, and most errors in the *Other* category are linked to punctuation (mostly to missing commas, as shown in the example), with 10.5 % of errors in total.

In the case of errors linked to the use of prepositions, we have chosen to include them in the error domain of the word that is responsible for selecting the preposition, instead of creating an umbrella "preposition" category, as is often done. The error domain for prepositions only includes errors where the selection of the preposition is linked to semantics rather than to the lexical selection constraints of another word. If all errors linked to prepositions are gathered, they account for 8.5 % of all errors.

Domains	Sublevel categories	Rate	Examples
Noun Phrase	Missing article	**12.1**	A. *All components shall meet the environmental requirements of [ ] table presented below.*
	Modification (mod. stacking, mod. morphology, postposed mod.)	**32.1**	A. *The maximum engine casing temperature shall be at 670°C and the inner fixed structure thermal blanket temperature shall be 350°C.*
	Preposition (missing prep., prep. selection)	2.7	B. *For information, in this event, if associated switch is not received after 45-seconds caution message is [ ] be posted.*
	Missing plural	**6.3**	B. *The system shall ramp up/down between two flow schedule*
	Ellipsis of head noun in coordinated structures	3.7	B. *Dirt and rock ingestion shall not damage air System*
	Agreement (mod/head, det/head)	1.1	B. *Each IASC channels shall feature 2 input high speed buses*
Adj. Phrase	Preposition selection after adjective	1.6	A. *The PRV shall be capable to close during the first 5 minutes of the fire test*
	Construction of comparative forms	0.5	A. *Components parts shall be [ ] foolproof as possible*
Prep. Phr.	Choice of preposition when semantically selected only	3.7	A. *Equipment in nacelle and engine area shall comply with its applicable performance standards during and after testing to the conditions defined in X.*
Verb Phrase	Verb group agreement	1.6	A. *Tools and their interfacing aircraft parts or equipment shall be design to permit foolproof installation/removal*
	Verb group construction	0.5	B. *The left full open/full closed micro switches shall and powered from X*
	Negation	1.1	B. *The solenoid shall be not activated as long as the input voltage is lower than 1V.*
	Preposition selection after verb	0.5	B. *The VENTS measurement shall range of at least -40°C to 70°C*
Sentence and clause	Information packaging	2.5	A. *The Supplier shall be responsible to design, integrate, interface, coordinate, manage, develop, test, […] to meet the requirements of this general specification and the individual equipment specification, within the program schedule.*
	Subject/verb agreement	0.5	A. *Demonstrations of compliance based on UL94 standard is no more considered as an acceptable mean of compliance.*
	Missing linker	1.1	A. *The use of GSE and tools shall not impair accessibility to any part, zone or equipment of the aircraft likely to be maintained or serviced [ ] the GSE or tools are in place*
	Ambiguous syntax	2.5	B. *When energized, the pressure required to open the TAPRV shall be less than 1.03 bar rel (15 psig)*
Other	Punctuation (missing commas, extra punct., use of non-standard symbols)	**10.5**	A. *However [ ] justification shall be provided, if, due to program constraints some tests have to be performed on development standard equipment.*
	Coherence and continuity of discourse	2.6	B. *In the event the element is failed stuck in position, and an advisory and associated info messages shall be posted.*
	Lexicon	**5.8**	B. *For information, in this event, if associated switch is not received after 45-seconds caution message is be posted.*
	Spelling	**6.8**	B. *In case of confict the valve closure command shall have priority over the opening command.*
	TOTAL	**100**	

Table 2. Synthesis of error categories with proportions and examples

6. Discussion of Results

6.1 Discussion of Categories

As is the case in most classification systems for language errors, there is an amount of overlap and ambiguity between some categories. For example, in the case of errors attributed to a missing plural on the head noun, it is sometimes difficult to judge whether a given error results from a missing plural morpheme or is actually linked to other factors. This ambiguity arises particularly when we attempt to offer a correction. For example, in *The Supplier shall demonstrate compliance with X for* *all system* *with electrical components or wiring*, the proposed corrections could either be:

- to add a plural ending to the head noun: *for all systems*,
- to add a definite determiner: *for all the system*,
- to reformulate the noun phrase: *for the whole system*,
- to replace the determiner, as the use of *all* could stem from a *calque*-type transfer of the French *pour tout système*, which actually translates to *for any system*.

As is visible from the high number of categories with very low rates, there is a lot of variety in the errors found in requirements. We find low error rates in the domain of the *Verb Phrase*, which accounts for a large portion of errors in other studies (Garnier, 2014.), since verbs are an integral part of sentences. In our corpus, the lack of errors in this domain can be attributed first to the high level of proficiency of the writers and the fact that the requirements are proofread, which may eliminate most agreement errors, and secondly to the use of a controlled natural language prohibiting the use of most complex verb groups. This could also explain the low rate of errors in the domain *Sentence and Clause*, as the constrained use of complex syntax helps limit errors. The sublevel categories with the highest proportion of errors concern errors which are difficult to detect and correct automatically, such as stacked modifiers (24.1 %, not shown in Table 2), missing articles (12.1 %) and missing commas (8.4 %, not shown in table 2). These categories also correspond to "simplification" strategies, with the omission of function words or punctuation that may be perceived by the author as superfluous or expendable.

The sublevel category we have identified as "modifier stacking" gathers segments in which a noun phrase is composed of a head noun and a string of modifiers to its left, sometimes with their own embedded modifiers (the term of "stacking" is used loosely here, and not in its specific syntactic meaning). Here are a few examples of this error type:

- *the following probable average operational duty cycle of the X*
- *The system shall include a locking in full closed position device.*
- *the reference computed ventilation flow (flight leg computed minimum reference flow i.e. X).*

Even though the use of several noun modifiers in an NP is becoming increasingly common in English, especially in technical and journalistic English (Pastor-Gomez, 2011), the fact that they rely on implicit information that needs to be reconstructed by the reader results in cognitive overload for the reader and possible loss of meaning (Biber et al., 1999).

Finally, there is a category of segments that was initially included in the error corpus; these segments all contain a form of ellipsis that renders them ungrammatical. A few examples are shown below:

- *In case of PDPS detected failed*
- *On ground, when PACK is selected OFF, corresponding RARV shall stay in position*
- *Both FCVs are selected open*
- *During PACK starting sequence the FCV is not commanded fully open*
- *In the event the RARV is failed stuck in position*

From the point of view of surface syntax, these segments have different forms (e.g. a past participle followed by a participial adjective, a past participle followed by a preposition, a past participle followed by one adjective or by an adjective phrase composed of a head adjective and a modifying adverb or a PP complement), but from a semantic point of view they are built on the same model, which is close that of verbal expressions such as *to turn on*, *to switch off*: the second term or phrase indicates the position or situation of a "mobile" element, such as a switch (e.g. *OFF*, *ON*, *open*, *closed*, *failed*), while the first one either specifies the action leading to that situation, or the observation of that situation (e.g. *selected*, *detected*). This type of segment was found only in Corpus B, with 31 instances (which made it one of the most numerous categories), but coming from only one company. It is imaginable that this type of phrasing is accepted and even expected in the company for which the requirements were written. This type of segment would therefore be an example of an ungrammatical but acceptable phrasing.

These segments were removed from the list of errors, even though we did find an alternative phrasing in one instance (*selected* ***to*** *OFF*), suggesting that the practice is either not accepted or not stable. Furthermore, the ellipsis of prepositions and other function words, which may conversely be seen as increasing concision and simplicity of expression, also creates a gap that must be filled by the reader, and may therefore lead to ambiguities if the use of such structures is not homogeneous and well-mastered.

6.2 Comparison with Other Error Corpora

We compared the proportion of the two most common error types in our requirements corpus with those found in two corpora of English written by native speakers of French with an intermediate to advanced level. One of them is a corpus of argumentative student essays on broad topics while the second one is a corpus of scientific papers.

In the case of missing articles, we notice that results in the two comparison corpora are similar, and are much lower than in the requirement corpus. However, determination errors in general account for 24.1 % (student essays) and 15.8 % (scientific papers) of all errors in the two comparison corpora, indicating that authors produce a more varied range of determination errors in these types of writing than in requirements, where determination errors other than missing articles are marginal to non-existent. Again, this can be interpreted as the result of "simplification" strategies having to do with the genre or requirements.

Error / **Corpus**	Missing article	Mod. stacking
Student essays	7.3 %	2.9 %
Scientific papers	7.7 %	14.3 %
Requirements	12.1 %	24.2 %

Table 3. Comparison of error distribution in three L2 English corpora

In the case of modifier stacking, there is a progression in the number of such errors found in the three corpora, with them being marginal in the corpora of student essays. This is consistent with other studies on such structures (e.g. Pastor-Gomez, 2011), which identify them as a feature of technical, scientific or journalistic English. They are also more varied and complex in the corpus of requirements, with up to five modifiers on the left of a head noun (see examples above).

We can conclude from this comparison that the distribution of errors seems to reflect the specificities of the genre of technical documents, warranting the collection and use of data in this genre, and specific treatment as an L2 production.

7. Applied Uses

7.1 Error Detection Using LELIE

One possible application of the results of this study is the inclusion of errors patterns for these error types in LELIE, as a grammar checking module complementing the system. Linguistics-based strategies are often overlooked in favor of statistical methods in standard grammar checking, but recent research has shown that using finely-tuned patterns based on linguistic research yields satisfactory results when confronted with very specific errors (Garnier, 2012; Garnier, 2014). Since they include information about the error pattern, and can incorporate annotations, pattern-matching methods also facilitate the inclusion of remediation strategies (i.e. corrected segments or guidance in the correction) and the creation of feedback messages for corrections or warnings. This aspect is a known weak spot of statistical methods. Early strategies for the detection or errors linked to modifier stacking are presented in (Garnier, 2014).

The implementation of this tool entails reusing or developing the following resources:

- the linguistic resources required to develop error patterns, and the features they must be associated with, especially action verbs, deverbal nouns, modal auxiliaries, adverbs, and determiners as well as their subclasses (these resources have already been developed within the framework of TextCoop and LELIE, and they are freely available for French and English),
- the grammatical or discourse structures that must be processed (conditionals, causes, illustrations, reformulations, purposes, circumstances, etc.). The recognitions of these structures enables the detection of errors within specific constructs. Since technical documents form a constrained linguistic genre, the recognition of these structures is high (Saint-Dizier, 2014),
- the type of error (or alert) message that is produced and the type of correction or suggestion. Explanations may also be included to help authors. Error messages and explanations are tailored to the user and to the domain, similarly to what is done in LELIE. This is implemented in the user profile and may be updated.

Evaluating such a system is a complex process, and includes multiple elements. The first criterion for evaluation is the adequacy of the patterns to detect errors, checking for precision and recall. Research in computer-assisted language learning has shown that checkers targeting non-native speakers should aim at maximizing precision, even at the detriment of recall, in order to avoid false positives that might undermine the user's confidence in their knowledge and use of the language, and risk fossilizing the use of erroneous structures (Tschichold, 1999), due to a belief in the authority of the automatic system. This guideline applies to requirements authoring as well, especially since the checker is not intended as a replacement for the several levels of proofreading done by human editors, but as assistance in the editing and correcting process.

The adequacy of the resources is another crucial criterion in the evaluation of such a system, since gaps in the lexical resources used can lead to low recall rates. Adapting the system to one given domain yields better results in the form of higher recall rates, but also translates into a lack of effectiveness when applied to other technical domains. Additionally, working with controlled English also means adjusting to the guidelines of the chosen CNL in terms of grammar.

The adequacy of the error alerts for requirements authors is also a necessary part of the evaluation, and can be done through the analysis of users' behavior during the use of the system. This is generally achieved using questionnaires, interviews and direct observation of the use of the writing aid, with the help of researchers in ergonomics.

7.2 In-Person Training

As mentioned previously, recent years have seen the creation of organizations focused on improving requirements by providing training and certifications to requirement engineers. The results of this research can therefore serve as the basis of coursework or exercises to be used by training providers or trainees, with the necessary adjustments.

Error categories should be adapted to be more user-oriented than research-oriented:

- the number of categories should be limited, in order to focus on central error types only;
- the phrasing of categories should be changed in order to be usable by trainees with a less extensive knowledge of syntactic concepts and jargon;
- remediation possibilities should be systematically included and diversified in order to provide a clearer illustration of error types.

Our corpus and error categories can also be used to create tailored grammar and reformulation exercises, a type of training activity that has been requested by requirement training professionals.

The evaluation of this applied use of the research would most likely be done as part of the evaluation of the training program performed by the organization offering it. Evaluation could include the results of a post-treatment test, compared with a pre-treatment test, as well as satisfaction questionnaires for trainees and instructors.

8. Conclusion

In this paper, we proposed a method based on corpus analysis aimed at identifying and documenting errors produced by native speakers of French when writing requirements in English. Errors are collected and classified using comparable categories based primarily on the domain of the error. The distribution of errors differs from what has been observed in similar corpora from different genres, indicating the need to take into account the specificities of requirements in order to provide adequate remediation. The results of this research can serve as the basis for the creation of a linguistics-based writing aid implemented in the LELIE system, and for in-person training activities. Because of the high proportion of errors linked to the Noun Phrase, and the specificity of the phenomenon of modifier stacking, we intend to focus our research on this error domain in order to obtain more precise data on these errors. In addition, the research will be extended to at least two additional industrial domains in order to yield more extensive comparable data in relation to the grammar of requirements.

9. Acknowledgements

This research was conducted with the support of a grant from the *International Requirements Engineering Board*, as part of the *IREB Academy Program* 2015-2016.

10. Bibliographical References

Barcellini F., Albert, C., Saint-Dizier, P. (2012). Risk Analysis and Prevention: LELIE, a Tool dedicated to Procedure and Requirement Authoring. In *Proceedings of LREC 2012*,p. 698-706.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Fuchs, N.E. (2012). First-Order Reasoning for Attempto Controlled English. In *Proceedings of the Second International Workshop on Controlled Natural Language.* Berlin:Springer-Verlag.

Grady, J. O. (2006). *System Requirements Analysis*. Orlando: Academic Press.

Garnier, M. (2012). Automatic Correction of Adverb Placement Errors for CALL. In L. Bradley, S. Thouësny (Eds.), *CALL: Using, Learning, Knowing, EUROCALL Conference, Gothenburg, Sweden, 22-25 August 2012, Proceedings*. Research-Publishing.net, p. 99-103.

Garnier, M. (2014). Utilisation de méthodes linguistiques pour la détection et la correction automatisées d'erreurs produites par des francophones écrivant en anglais. PhD Thesis in English Linguistics. Université de Toulouse. https://tel.archives-ouvertes.fr/tel-01257640.

Granger S., Dagneaux E., Meunier F. and Paquot M. (2009). *The International Corpus of Learner English. Version 2. Handbook and CD-Rom*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Hull,E., Jackson,K. and Dick, J. (2011). *Requirements Engineering*. 3rd ed. NY: Springer-Verlag.

Kuhn, T. (2013). A Principled Approach to Grammars for Controlled Natural Languages and Predictive Editors. *Journal of Logic, Language and Information*, 22(1). p. 33-70.

Kuhn, T. (2014). A Survey and Classification of Controlled Natural Languages. *Computational Linguistics*, 40(1). p. 121-170.

Lennon, P. (1991). Error: some problems of definition, identification and distinction. *Applied Linguistics,* vol. 12, n°2, p 180-196.

Jarvis, S., Pavlenko, A. (2007). *Crosslinguistic Influence in Language and Cognition*. London, New York: Routledge.

Pastor-Gomez, I. (2011). *The Status and Development of N+N Sequences in Contemporary English Noun Phrases*. Bern: Peter Lang.

Saint-Dizier, P. (2012) Processing natural language arguments with the TextCoop platform. *Journal of Argumentation and Computation*, vol 3(1). p. 49-82.

Saint-Dizier, P. (2014). *Challenges of Discourse Processing: the case of technical documents*. Cambridge: Cambridge Scholars.

Saint-Dizier, P. (2015). Features of an error correction memory to enhance technical texts authoring in LELIE. *International Journal of Knowledge Content Development and Technology,*, vol 5(2). p. 75-101.

Tschichold, C. (1999). Grammar checking for CALL: Strategies for improving foreign language grammar checkers. In K. Cameron (Ed.) *CALL: Media, Design and Applications*. Lisse: Swets et Zeitlinger. p. 203-222.

Wyner, A., Angelov, K., Barzdins, G., Damljanovic, D., Davis, B., Fuchs, N., Hoefler, S., Jones, K., Kaljurand, K., Kuhn, T., Luts, M., Pool, J., Rosner, M., Schwitter, R., Sowa, J. (2010). On Controlled Natural Languages: Properties and Prospects. In *Proceeding of Controlled Natural Language Conference*. Berlin/Heidelberg: Springer. p. 281-289.