HAL
open science

Using a SMT solver for risk analysis: detecting logical mistakes in texts

Florence Dupin de Saint-Cyr, Marie-Christine Lagasquie-Schiex, William Raynaut, Patrick Saint Dizier

To cite this version:

Florence Dupin de Saint-Cyr, Marie-Christine Lagasquie-Schiex, William Raynaut, Patrick Saint Dizier. Using a SMT solver for risk analysis: detecting logical mistakes in texts. International Conference on Tools with Artificial Intelligence - ICTAI 2014, Nov 2014, Limassol, Cyprus. pp. 867-874, 10.1109/ICTAI.2014.133 . hal-01147277

HAL Id: hal-01147277

https://hal.science/hal-01147277v1

Submitted on 30 Apr 2015

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.

Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 13266

To link to this article : DOI :10.1109/ICTAI.2014.133
URL : http://dx.doi.org/10.1109/ICTAI.2014.133

To cite this version : Dupin De Saint Cyr - Bannay, Florence and Lagasquie-Schiex, Marie-Christine and Raynaut, William and Saint-Dizier, Patrick *Using a SMT solver for risk analysis: detecting logical mistakes in texts*. (2014) In: International Conference on Tools with Artificial Intelligence - ICTAI 2014, 10 November 2014 - 12 November 2014 (Limassol, Cyprus).

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Using a SMT solver for risk analysis: detecting logical mistakes in texts

F. BANNAY, MC. LAGASQUIE-SCHIEX, W. RAYNAUT, P. SAINT-DIZIER
IRIT-UPS
Toulouse, France
Email: {bannay, lagasq, dizier}@irit.fr

***Abstract*—The purpose of this paper is to describe some results of the LELIE project, that are a contribution of Artificial Intelligence to a special domain: the analysis of the risks due to poorly written technical documents. This is a multidisciplinary contribution since it combines natural language processing with logical satisfiability checking.**

This paper explains how satisfiability checking can be used for detecting inconsistencies, redundancy and incompleteness in procedural texts and describes the part of the implemented tool that produces the logical translation of technical texts and realizes the checkings.

***Index Terms*—Knowledge Representation, Reasoning, AI in Natural Language Processing and Understanding**

I. INTRODUCTION

Companies maintain a large number of procedural texts in various sectors which may lead to risky situations. Poor requirement compliance in procedures often leads to accidents, with major health and ecological consequences. Social and psycho-social problems (*e.g.* due to poor management requirements) are also often encountered and, obviously, negative financial situations may be critical. The negative consequences of bad or poor requirements and procedures have been investigated in depth. Industry data show that approximately 50% of product defects originate from incorrect or unreadable procedures. Perhaps 80% of the rework effort on a development project can be traced to requirements defects. Because these defects are the cause of over 40% of accidents in safety-critical systems (see [1]), poor requirements and procedures have even been the ultimate cause of both death and destruction.

The LELIE[1] project was realized from 2011 to 2013[2]. It was funded by a special track of the ANR[3], the ANR Emergence, combining ergonomy, language processing and artificial intelligence with an applicative orientation. The main goal of this project is to detect potential risks in industrial processes based on language processing and logic-based artificial intelligence techniques. This has been realized via the analysis of risks indicators of different kinds (health, ecology, economy, etc.) on the basis of written technical documents.

In this paper, we present an original approach proposed in LELIE where natural language processing combined with AI is used for an analysis of inconsistencies, redundancies and incompleteness typical of technical documents. We concentrate on *procedural documents and requirements* (*e.g.* for installation, production, maintenance) which are, by large, the main types of technical documents.

Given a set of procedures over a certain domain produced by a company, and possibly given some domain knowledge (ontology or terminology and lexical data), the goal is to detect and model these errors and then to annotate them wherever potential risks are identified. Procedure authors could then be invited to revise these documents. Risk analysis is based on several types of considerations:

1) Inappropriate ways of writing that may lead to potential risks: texts including a large variety of complex expressions, fuzzy terms, implicit elements, scoping difficulties (connectors, conditionals), lack of cohesion, inappropriate granularity level, etc. These inappropriate ways were established by cognitive ergonomic simulations and analysis (see [2]).
2) Incoherence among procedures: detection of unusual ways of realizing action (*e.g.* unusual instrument, temperature, length of treatment, etc.) with regard to similar actions in other procedures. This was based on a repository of actions from previously processed procedures (Arias software) on a given domain.
3) Lack of compliance of procedures with regard to domain security requirements and regulations, therefore leading to risks. Inconsistencies or incompleteness situations have often been observed between requirements and procedures.

Only Point 3 is developed in this paper (the study of Points 1 and 2 is out of the scope of this paper since it is mainly based on the way documents are written[4]). This point mainly deals with content aspects and requires several types of inferences and reasoning in relation with risk analysis.

The goal of this paper is threefold. First it demonstrates the feasability of a tool that can automatically detect potential risks in natural languages technical documents. It also shows the benefits of logic-based tools like SMT-solver and theoretical concepts such as ATMS for industrial-oriented applications. Third it shows that it is possible to combine a natural language analysis with a logical handling of inconsistency. Moreover, we show that those tools allow us not only to detect the

[1]LELIE: An intelligent assistant for the analysis and the prevention of risks in industrial processes.

[2]by IRIT in collaboration with CRTD-CNAM, Paris.

[3]ANR: Agence Nationale pour la Recherche; a French government agency.

[4]That also corresponds to a major problematics in the industry, see [3].

existence of a problem but also to point out the parts of the text that are responsible for it. The identification of major errors in procedures with regard to related requirements is a very important result achieved by the LELIE tool.

Note that there are two main components in the LELIE tool: a "linguistic component" and a "logical" component. This paper is concerned with only the last one. Nevertheless, for a best understanding of the tool, Section II briefly explains the first step of the process: the automatic analysis of the structure of requirements and of procedures done with the system TEXTCOOP in order to produce a translation into a logical form. TEXTCOOP and LELIE are developed in [3]; processing technical documents results in the annotation of those structures which are typical of technical documents, *e.g.*: titles, instructions, prerequisites, warnings, explanations. Then the core of this paper, the logical handling part, is presented in Section III based on the outputs of the language processing step carried out with TEXTCOOP. The corresponding implemented tool is described in Section IV. Section V gives some related works, suggests directions for further research and concludes the paper.

II. NATURAL LANGUAGE ANALYSIS

Procedural texts and requirements are written in specific forms and are often very well structured, hence they are less complex, in terms of structure and ambiguity, to analyze and translate into a logical form. Indeed, procedures are often presented under the form of a list of instructions, each instruction being expressed in a simplified and standard way following guidelines.

A. TEXTCOOP *engine*

LELIE is based on the TEXTCOOP system (see [4]), a system dedicated to language analysis, in particular discourse (including the taking into account of long-distance dependencies). The kernel of the system is written in Prolog SWI, with interfaces in Java. LELIE realizes an annotation of the different discourse structures useful for our purpose. To avoid the variability of document formats, the system input is an abstract document with a minimal number of XML tags as required by the error detection rules. Managing and transforming the original text formats into this abstract format is not dealth with here.

Briefly, TEXTCOOP identifies the following structures which are of interest for the investigations presented here.

- titles, instructions, requirement statements, prerequisites, definitions, warnings, advice and some form of explanation which are proper to technical texts,
- thematic structures: theme, topic, strength (for requirements),
- within instructions and requirements: the main verb and its complements, in particular instruments of means (with equipment or product names) or adjuncts such as amounts which are numerical values (Ph, Volts, weights, etc.) and temporal complements.

Each of these types of values has a specific annotation in XML possibly with attributes (see an analysis and a description of the performances of TEXTCOOP in [3]).

B. *Procedural texts tagged by* TEXTCOOP

Going into more details about the content analysis, there are three kinds of input data, given in text files:

- requirements: information describing the context and the precautions with which a certain action (included into a procedure) must be carried out,
- a procedure: which is an ordered sequence of instructions,
- a list of synonyms in order to restrict the vocabulary to manage the term matching aspects between requirements (prescriptive style) and the related procedures (injunctive style).

Examples are provided below. Note that real examples correspond to confidential data given by industrial partners of the project. These data are technical texts containing hundreds of lines and thousands of symbols. So it is not possible to present them in this paper. Thus, for the sake of readability, the examples presented here are kept minimal and very simple but are sufficient for illustrating the tagged results given by TEXTCOOP and for introducing the formal version of these texts.

Example 1: Here is a short extract of a tagged instruction:

```
< procedure>
   < predicate> use </predicate>
        < object> a rope </object>
   to tie the harness.
</procedure>
```

If we consider the instruction of this procedure, the verb "use" is a predicate that takes as arguments a subject (here the person who executes the procedure, called the operator, `op` for short) and a complement/adjunct (here "the rope"). So, this sentence can be formally written as: `use(op,rope)`. The remainder of the instruction is ignored.

Another point must be taken into account: the theme of the procedure to properly relate requirements and procedures. For instance:

```
< procedure> in order to
   < theme> sweep a chimney</theme>
   < predicate> climb </predicate>
       < location> on the roof </location>
</procedure>
```

There are two parts in this instruction; the first one can be translated as previously whereas the second one needs a special translation since it gives the theme of the procedure, *i.e.* the execution context of the procedure; in this case, it can be formally expressed by the formula `is(theme,sweep_a_chimney)`.

Note that this theme can be extracted from the texts by TEXTCOOP, using for instance the title of the documents.

Example 2: This example corresponds to a set of requirements:

```
< requirement> in case of
   < theme> work at a height </theme>
   < predicate> do not use </predicate>
        < object> ropes </object>
</requirement>
```

Here, there are also two aspects: the verb (with its subject and complements/adjunct) and the theme. Each of these elements can be easily formalized by predicates:

```
is(theme,work_at_a_height),
¬use(op,rope))
```

but these elements are linked:

```
if is(theme,work_at_a_height)
then ¬use(op,rope);
```

this link can be expressed by a logical implication:

```
is(theme,work_at_a_height) → ¬use(op,rope);
```

Example 3: This example gives illustrates a synonym file:

```
SYN = work at a height
sweep a chimney
work on roof
```

These data are only used for simplifying the texts (requirements or procedures). They must be elaborated from business data. For instance, using this synonym file, the themes of Example 1 and of Example 2 become identical. This will make it easier for us to detect logical incorrectness in them.

Nevertheless, the term "synonym" is too narrow and this method of simplification is a rough one. In a future work, it will be necessary to use domain ontologies. Indeed, linguistically speaking, the expressions "work at a height", "sweep a chimney" and "work on roof" are not synonyms; they are rather in an entailment relationship with each other and this relationship could be extracted from an ontology.

III. Logical mistakes detection in LELIE

The solution proposed for answering to Point 3 given in Section I is a tool using a logical representation of the texts and applying basic AI reasoning principles in order to validate these texts. In our proposal, this validation is done on the three following points: inconsistency, incompleteness and redundancy detection. This is a two-step process. First, we have chosen to translate the written technical documents into a formal language. Then, using an open-source solver embedded into a Java code, we are able to reason on this translated documents.

A. The Logical language

We choose a representation language $\mathcal{L}$ which is a variant of a first-order logic language (see [5]) classically defined with 7 vocabularies:

- V_c (resp. V_v, V_f, V_P) is the set of constants (resp. variables, functions, predicates),
- $\{\neg, \wedge, \vee, \rightarrow, \leftrightarrow\}$ is the set of classical connectors representing respectively *negation*, *and*, *or*, *implication* and *logical equivalence*,
- $\{(,)\}$ is the set of delimitors and $\{\forall, \exists\}$ is the set of quantifiers.

In our case, $\mathcal{L}$ is defined without symbols of function (so $V_f = \varnothing$). The *terms* are classically defined using constants and variables; *ground terms* are special terms using only constants. For any predicate symbol $P \in V_P$ of arity n, $P(t_1, \ldots, t_n)$ is an *atomic formula* (or *atom*) whenever $t_1, \ldots, t_n$ are terms. Moreover, if $t_1, \ldots, t_n$ are all ground terms then $P(t_1, \ldots, t_n)$ is a *ground atom*. $\perp$ is the atomic formula representing the *contradiction*. A *literal* is an atom or its negation. Other non atomic formulas are built by using connectors, and quantifiers applied to variables and the delimitors.

The choice of this language is justified by the following facts:

- Procedural texts are composed by simple sentences with a limited set of terms; it is easy to translate them into a formal language (see Examples 1 and 2).
- First-order logic is a well known language with an interesting basic expressivity.
- There are many possible extensions if we want to increase this expressivity (for instance, the reintroduction of functions).
- Several open-source solvers exist whose efficiency has been proved by decades of research and competitions; in this project, we choose to use the "Z3" solver (see [6]) that respects the formalism that is issued from the Satisfiability Modulo Theory (SMT) area, see [7].

The SMT library allows us to perform automated deduction and provides methods for checking the satisfiability of first-order formulas with regard to some set of logical formulas T (called a theory). By being theory-specific and by restricting to certain classes of first-order formulas, the SMT solvers can be more efficient in practice than general-purpose theorem provers. So, in order to be efficient, the formulas of our language $\mathcal{L}$ will be expressed in the SMT formalism. It is an important point for our tool since the size (in number of formulas and in number of symbols) of the knowledge base corresponding to a real example can very quickly become huge.

B. Translation of the input data

Requirements and procedures are translated into first-order logic. For each sentence, this translation is a three-step process:

1) "cleaning" the text of the sentence by using the lists of synonyms and by removing the articles, and identifying the theme(s),
2) finding the mask corresponding to the sentence and formatting it with regard to this mask,
3) translation of the clean and formatted sentence into first-order logic, using the theme(s).

Several types of masks (reformulations) can be considered according to the form of the sentence:

- masks that are reduced to only one simple component (a simple sentence: a verb, its subject, its complements and some adjuncts),
- masks corresponding to a complex component (a conjunction or disjunction of simple components),
- masks corresponding to a sentence with conditions and exceptions (thus giving a structure with 3 complex components); for instance: *if condition1 then action except if condition2*.

Moreover, each simple component can be instantiated by several variants depending on the semantics of the sentence. For instance:

- the subject of a verb can be a constant or can be quantified universally or existentially;
- the sentence is in direct or indirect form;
- the verb is or is not an action verb ...

Example 1 (cont): After cleaning, the text becomes (all the non-tagged parts of the text are removed and the synonyms given in Example 3 are used):

```
< procedure>
   < theme>  work at a height </theme>
   < predicate> climb </predicate>
      < location> roof </location>
   < predicate> use </predicate>
      < object> rope </object>
</procedure>
```

Then, for each sentence, a generic mask is identified and the sentence is formatted with regard to this mask. In this example, two masks are used (when an element is missing, it is replaced by NULL, except for the time-step that is encoded by an integer incremented at each instruction):

```
theme is work_at_a_height 0
  (mask: subject state-verb attribute time)

NULL climb NULL NULL roof 1
  (mask: subject action-verb direct-obj
         method place time)

NULL use rope NULL NULL 2
  (mask: subject action-verb direct-obj
         method place time)
```

Each sentence/instruction is considered as a first-order formula that must be true at the moment corresponding to the execution of the instruction. So, this procedure corresponds to the three following first-order formulas:

```
is(theme,work_at_a_height,0)
climb(op,NULL,NULL,roof,1)
use(op,rope,NULL,NULL,2)
```

Since, the chosen solver is the "Z3" solver (see [6]) that uses the SMT formalism (see [7]), this procedure is encoded in the SMT formalism and the resulting program code consists of:

- first, the definition of the different elements used in the language (here `Agent`, `Item`, `Place`, `Attribute` and `Method`); note that temporal elements are encoded as integers (`Int` is a predefined element in the SMT formalism);
- then, for each sentence, there are the definition of the predicate (a function in the SMT formalism), the definition of the constants, the formula.

So, the final translation of this example is:

```
(declare-sort Agent)
(declare-sort Item)
(declare-sort Place)
(declare-sort Attribute)
(declare-sort Method)

(echo "< theme> sweep a chimney</theme>")
(declare-fun
     is (Item Attribute Place Int)
        Bool)
(declare-const
     it_theme
     Item)
(declare-const
     att_theme_work_at_a_height
     Attribute)
(declare-const
     pl_NULL
     Place)
(declare-const
     ag_NULL
     Agent)
(declare-const
     me_NULL
     Method)
(assert
     (is it_theme att_theme_work_at_a_height
         pl_NULL 0))

(echo "< predicate> climb </predicate>
   < object> onto the roof </object>")
(declare-fun
     climb (Agent Item Method Place Int)
           Bool)
(declare-const
     it_roof
     Item)
(assert
     (climb ag_NULL it_roof me_NULL
            pl_NULL 1))

(echo "< predicate> use </predicate>
   < object> a rope </object> ")
(declare-const
     it_rope Item)
(assert
     (use ag_NULL it_rope me_NULL
          pl_NULL 2))
```

Note that the initial sentence (before cleaning) is kept as a comment for an easier reading of the result ("echo" line in the translation).

C. Checking correctness with a SMT solver

We propose 3 kinds of validation for procedures with regard to requirements:

- consistency checking,
- incompleteness detection,
- non-redundancy checking.

All these validations are realized by using the notion of satisfiability[5] of a formula (a set of formulas is handled as the logical conjunction of the formulas of the set). $\phi \models \bot$ denotes the fact that ϕ is unsatisfiable. Checking satisfiability will be done with a solver (here the Z3 solver). In this document, we will use the following notations:

Notation 1: F_i denotes the formula corresponding to the ith instruction of the procedure. R denotes the formula corresponding to the set of requirements. $\text{Lit}(F) = \{l_1, ..., l_n\}$ denotes the set of (positive or negative) literals used in the formula F.

[5]Let ϕ be a first-order formula. ϕ is satisfiable iff there exists a model of ϕ (*i.e.* some assignment of appropriate values to its symbols under which ϕ is evaluated to true)

1) Inconsistency Detection: The detection of inconsistency can be done either on a set of requirements, or on a procedure (a set of instructions), or between a set of requirements and an instruction (or a set of instructions). This detection can be formally defined as follows:

Definition 1 (Inconsistency Detection): Let R be a set of requirements. Let $\{F_1, \ldots, F_n\}$ be a set of instructions.

- There exists an inconsistency in the set of requirements iff $R \models \bot$.
- There exists an inconsistency in the set of instructions iff $F_1 \wedge \ldots \wedge F_n \models \bot$.
- There exists an inconsistency between the set of requirements and the set of instructions iff $R \wedge F_1 \wedge \ldots \wedge F_n \models \bot$.

Example 4: Input data are the followings:

- requirements:

```
< requirement> in case of
< theme> work at a height </theme>
< predicate> be protected </predicate>
< predicate> do not use </predicate>
    < object> ropes </object>
</requirement>
```

- instructions:

```
< procedure> in order to
< theme> sweep a chimney </theme>
< predicate> climb </predicate>
   < location> onto the roof </location>
< predicate> use </predicate>
   < object> a rope </object>
</procedure>
```

The logical translation of these data corresponds to the following formulas:

R: (is(theme, work_at_a_height) $\rightarrow$ is(op, protected)) $\wedge$(is(theme, work_at_a_height) $\rightarrow \neg$use(op, rope))
F_1: is(theme, work_at_a_height)
F_2: climb(op, roof)
F_3: use(op, rope)

Here, requirements are consistent ($R \not\models \bot$), instructions are consistent ($(F_1 \wedge F_2 \wedge F_3) \not\models \bot$) and there is an inconsistency between requirements and instructions ($(R \wedge F_1 \wedge F_2 \wedge F_3) \models \bot$).

Moreover, using an ATMS (see [8]), it is possible to identify the origin of the inconsistency. This can be done very simply by first translating every formula in one or several clauses[6] then introducing a new predicate of arity 0 (called assumption predicate) *for* each clause and *in* each clause. The ATMS is able to detect the *nogoods* of the knowledge base, *i.e.* subsets N of formulas such that:

1) the formulas of N are only assumption predicates,
2) the set N is inconsistent with the knowledge base,
3) N is minimal with regard to set-inclusion among the sets respecting 1) and 2).

Example 4 (cont): In this example, the knowledge base contains 5 clauses (R produces two clauses) completed with the assumption predicates $\mathtt{R}_1$, $\mathtt{R}_2$, $\mathtt{F}_1$, $\mathtt{F}_2$, $\mathtt{F}_3$:

[6]A clause is a disjunction of atomic formulas.

($\mathtt{R}_1 \wedge$ is(theme,work_at_a_height)) $\rightarrow$ is(op,protected)
($\mathtt{R}_2 \wedge$ is(theme,work_at_a_height)) $\rightarrow \neg$use(op,rope)
$\mathtt{F}_1 \rightarrow$ is(theme,work_at_a_height)
$\mathtt{F}_2 \rightarrow$ climb(op,roof)
$\mathtt{F}_3 \rightarrow$ use(op,rope)

Then, using an ATMS, the set $\{\mathtt{R}_2, \mathtt{F}_1, \mathtt{F}_3\}$ is a nogood that gives the origin of the inconsistency between requirements and instructions: if someone works at a height then he cannot use ropes ($\mathtt{R}_2$); someone works at a height ($\mathtt{F}_1$); and he uses ropes ($\mathtt{F}_3$).

Note that the computation of the nogoods is already partially implemented in the SMT-solver Z3, since it is possible to assign a name for each assertion and to extract unsatisfiable cores (*i.e.*, a subset of assertions that are mutually unsatisfiable). However this set is not guaranteed to be minimal and only one set is returned even when there are several possible causes for inconsistency[7]; so using the Z3 function, we can directly have a set containing one of the nogoods.

2) Checking non-redundancy: The check of non-redundancy consists in verifying that the addition of a new instruction to a set of instructions allows the inference of new formulas (otherwise it is the symptom that this new instruction is useless). This check can be formally defined as follows:

Definition 2 (Non-redundancy check): Let $\{F_1, \ldots, F_j\}$ be a set of instructions. Let F_k be a new instruction. If $F_1 \wedge \ldots \wedge F_j \wedge F_k \not\models \bot$ then F_k is not redundant iff $F_1 \wedge \ldots \wedge F_j \wedge \neg F_k \not\models \bot$[8].

Example 4 (cont): In this example, let us consider that the third instruction F_3 (which produces an inconsistency) has been replaced by the following new instruction F_3':

```
< predicate> climb </predicate>
   < location> onto the roof </location>
```

This instruction is exactly the instruction F_2. So there is a redundancy that is detected as follows: $F_1 \wedge F_2 \wedge F_3' \not\models \bot$ (no inconsistency in the procedure) and $F_1 \wedge F_2 \wedge \neg F_3' \models \bot$ (so $F_1 \wedge F_2 \models F_3'$). This means that F_3' is inferred by $F_1 \wedge F_2$ and so F_3' is useless.

Note that it is possible to explain the source of redundancy (as done for inconsistency in Section III-C1) by extracting unsatisfiable cores containing the negation of the new instruction. In the previous example we would obtain that F_3' is redundant with $\{F_2\}$ (since $\{F_2, \neg F_3\}$ is an unsatisfiable core).

3) Incompleteness detection: Searching for incompleteness corresponds to two distinct options that can be formally define as follows:

Definition 3 (Incompleteness detection): Let R be a set of requirements. Let $\{F_1, \ldots, F_j\}$ be a set of instructions and F_k be a new instruction.

- There exists an incompleteness in the set of requirements iff there is at least a ground literal $l \in \mathrm{Lit}(R)$ such that $R \models l$.

[7]A solution has been proposed by Liffiton and Malik [9] but it is not yet available in the standard solver.

[8]That means that F_k is not inferred by $\{F_1, \ldots, F_j\}$.

- There exists an incompleteness of the instruction with regard to the set of requirements iff there is at least one ground literal $l \in \text{Lit}(R \wedge F_k)$ such that $R \not\models l$, $F_1 \wedge \ldots \wedge F_j \wedge F_k \not\models l$, $R \wedge F_1 \wedge \ldots \wedge F_j \wedge F_k \models l$.

The first point of Definition 3 is not, strictly speaking, an "incompleteness" (it rather means that the set R is too strong deductively), whereas the second point exactly corresponds to an incompleteness since it means that the union of requirements and instructions allows the inference of new formulas that are not inferred by the instructions alone (which means that these instructions are too weak deductively).

Example 4 (cont): In this example, let us consider only the requirements R and the first instruction F_1. There is an incompleteness of this instruction with regard to requirements. Indeed, considering the ground literals that can be defined from $R \wedge F_1$, we have:

ground atom v	$\models$ by $R \wedge F_1$	$\models$ by F_1	
`is(theme,work_at_a_height)`	Yes	Yes	
`is(op,protected)`	Yes	No	$\star$
`use(op,rope)`	No	No	
`¬is(theme,work_at_a_height)`	No	No	
`¬is(op,protected)`	No	No	
`¬use(op,rope)`	Yes	No	$\star$

Using only the instruction, it is not possible to deduce the ground atoms indicated with the $\star$ symbol. This means that the instruction is incomplete with regard to the requirements. Indeed, the procedure lacks at least an instruction in order to be protected and another one for forbidding the use of ropes.

IV. Functional description of the tool

The tool (described on the French website [10][9]) has been realized in the Java language; it mainly implements four features:

1) Project definition: a project gathers several textual procedures, requirements and synonyms files, the user can create and modify projects.
2) "Cleaning" and cutting sentences: it consists in suppressing useless words and replacing some words by their standard synonyms, translation of tagged sentences issued from TextCoop into formatted sentences according to different masks (the tagged sentences correspond either to requirements or to instructions). At this stage "manual correction" is enabled: the user can propose other synonyms or disagree on the mask chosen.
3) Consistency, completeness and non-redundancy checking: these functionalities are available after a translation of formatted sentences into first-order formulas using the SMT formalism. Then the user can run inconsistency detection in requirements, or in instructions, or between requirements and instructions (using a sliding window on the set of instructions). After having checked consistency, the user can run a completeness check in the requirements and for an instruction with regard to the requirements. The user can also run a non-redundancy check inside the instructions.
4) Miscellaneous: two other functionalities have been proposed, the possibility to automatically load the requirements associated with a procedure by using its theme(s) and the possibility to take into account numeric interval values checking.

[9]Note that this tool is in French language.

A. Project definition

In order to provide a convivial tool, several files can be gathered in one project. Those files and the ones that will be generated will be stored in the same directory. The interface for the project handling divides the screen in three parts, requirement files, procedure files and synonyms files. The tool enables the user to add or remove files.

B. "Cleaning" and cutting

This functionality consists of removing the words that are not tagged by TextCoop. In a second time all articles and prepositions are removed. Moreover the text is updated in order to reduce at most the vocabulary used, this is done by using synonyms files. Before this update, the tool checks for the consistency of synonyms files in order to avoid problems like "word A should be replaced by word B", and "word A should be replaced by word C"; then the tool does a transitive closure of the synonyms files in order to simplify cases where "word A should be replaced by word B" and "word B should be replaced by word C".

The cutting part consists in matching the clean sentence with a predefined mask. In practice this is done by studying the tags that have been given by TextCoop in order to fit the mask.

It is possible for the user to browse the different files, and to open a detailed view of these files in the main part of the screen (see Figure 1). The different stages of the cleaning and cutting processes are shown to the user who can check for the validity of the current translation (and may alert the system if there is a wrong mask selected).

Since the tool is in an experimental stage, it may produce some mistakes. Those mistakes may come from a bad tagging by TextCoop due to the use of too complex sentences, or ambiguous vocabulary, or simply to unpredicted language use. This leads us to allow for manual corrections: the user is enabled to give new synonyms, to alert about some mask mistakes and to propose another tagging of some words and eventually to write an explanation/comment for the TextCoop administrator.

C. Logical correctness

Once the clean and corrected sentences have been associated with a mask, they are translated into logic (as explained in Section III-B), then the logical formula is sent to Z3 solver. The results are parsed in order to give a clear diagnostic and they are presented thanks to a translation into XSL which allows to show the texts in a browser in a more convenient way, using colors and fold/unfold effects. More precisely the

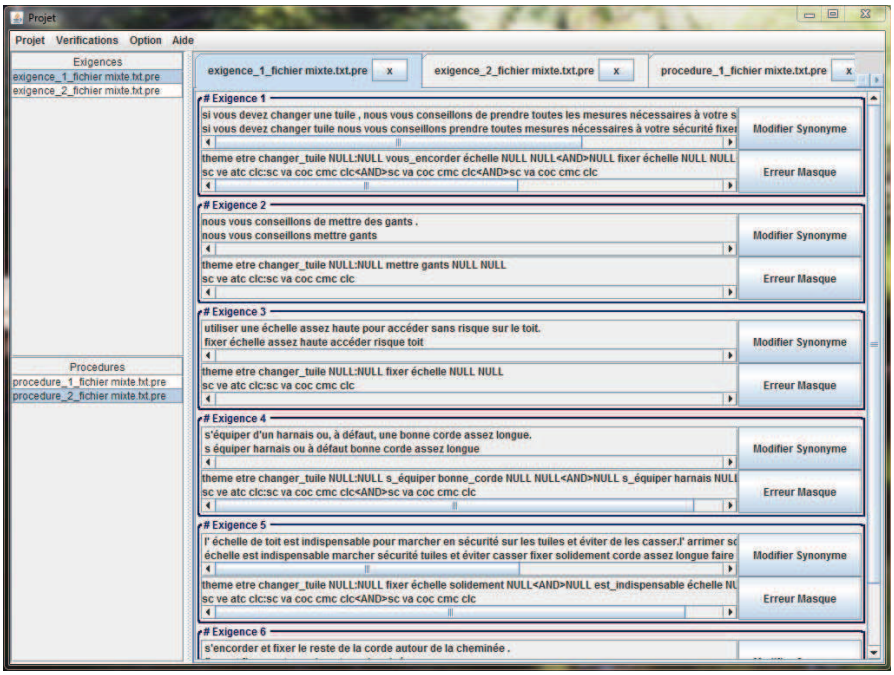

Fig. 1. Results of the different stages of the cleaning and cutting processes in LELIE on a French example

original sentences are shown in the initial order that they had in the procedure, they are coloured by the system (for instance the inconsistencies appear in red while correct instructions are in green, see Figure 2), every item is unfoldable in order to see their different translation stages and it is possible to obtain an explanation of the inconsistency by clicking on the item "inconsistency sources" (see Section III-C1).

A procedure being consistent, its completeness should be checked. An item "Completeness check" is proposed to the user and can be unfolded by clicking on it. The red color is used to signal the requirements that are not fulfilled. If a redundancy is detected then the corresponding text is colored in yellow.

D. Miscellaneous

The automatic selection of the requirements related to a procedure according to its theme is an available feature of the tool.

The procedure consistency can be checked either instruction by instruction or by checking a group of instructions together and by shifting the entire group forward of one instruction.

Another feature is the ability to detect and reason about numerical values and intervals of numeric values. For instance, it is possible to use numbers in instructions or requirements "check that the sensor temperature is equal to 25", "the sensor temperature should be between 5 and 10". This has been done by adding comparison tags to the masks as well as values or interval values. In particular, it is possible to set the time as well as the place of an instruction. The time being represented by an integer, it can be used in comparisons.

Note that due to the many possible file formats for text encoding (specially for French), this encoding should be specified by the user. The tool may run on different operating systems, the operating system is detected automatically (it is necessary for a correct handling of the file storage).

V. Related works and discussion

This paper describes a tool that is based on AI-techniques and automatic natural language processing, more precisely the part of the tool that is able to translate procedural texts into a predicate language in order to detect logical incorrectness. This detection is done thanks to the Z3 SMT-solver. The choice to use a SMT-solver and not a SAT-solver is justified by the fact that it is easier to translate a sentence in natural language by a logical expression using a predicative form than into an expression with propositional variables. Moreover in a SMT-solver it is possible to handle numerical values which are frequent in industrial domains, the availability of quantifiers and function symbols was also one reason for our choice even if we do not use them in the current version of the tool. The use of Prolog could also have been chosen in order to check logical inconsistencies, the benefit of SMT-solver is their efficiency in time (this is due to the SAT research progress that have been stimulated by the international competitions among SAT-solvers [11] and among SMT-solvers [12]).

Our use of the SMT-solver Z3 is a new application for this kind of solvers that were initially designed for software verification and analysis [6], [13]. It also has numerous other applications in automated theorem proving, in hardware verification [14], and in scheduling and planning problems [15] for

- **Exigence 1 : il faut être protégé**
- **Exigence 2 : ne pas utiliser des cordes**
- **Exigence 3 : utiliser des échaffaudages**

- **Theme de la procedure : pour ramoner**
- **Instruction 1 : il faut monter sur le toit**
- **Instruction 2 : utiliser une corde**

- **Sources d'incohérences**
  - **Exigence 2 : ne pas utiliser des cordes**
  - **Theme de la procedure : pour ramoner**
  - **Instruction 2 : utiliser une corde**

Fig. 2. Results of an inconsistency detection with LELIE on a French variant of Example 4-Section III-C1 (in red – or dark grey in the black-and-white document – the unfolded instruction that produces the inconsistency)

instance. But as far as we know this is the first use of Z3 in combination with an automatic handling of natural language in order to detect logical incorrectness in texts.

The LELIE project is a new approach for analyzing technical texts. In this framework, the existing systems were either only able to correct grammatical mistakes or only dedicated to manage the requirements files and handle requirements traceability (see [16] for a review of the existing softwares). The idea to help people to correct higher-level mistakes like logical ones is completely new in the domain of automatic and interactive correction of written technical texts. The correction checks that are carried out by our tool are crucial to reduce complexity and mistakes in industrial texts hence to prevent industrial risks.

If the studied framework is enlarged to other kinds of text, there exists at least one other tool (see [17]), that also tries to combine natural language processing and logic; this work concerns regulation texts; nevertheless, these texts are more complex that the technical ones and so the formal language used in [17] is a very complex logic, a default temporal logic, without efficient solvers.

Concerning future work, we would like to consider at least the following directions:

- the use of ontologies in order to exploit the hierarchical links between the manipulated objects (for instance to exploit more intelligently the synonym files).
- a user-validation of the set of existing masks on real-cases (for instance, do we need masks containing universal/existential quantifiers?); and, more generally, the use of real-case is needed for validating the principles and the tool, since the scalability of the tool is very important, it is an ongoing task,
- the automatic correction of inconsistencies (possibly by giving priorities to some requirements),
- an applet version of this tool.

References

[1] Health and Safety Executive, "Fatal injury statistics," http://www.hse.gov.uk/statistics/fatals.htm.

[2] F. Barcellini, C. Albert, and P. Saint-Dizier, "Risk analysis and prevention: LELIE, a tool dedicated to procedure and requirement authoring," in *Proc. of LREC*. ACL, 2012.

[3] P. Saint-Dizier, *Challenges of Discourse Processing: the case of technical texts*. Cambridge University Press, 2014.

[4] ——, "Processing natural language arguments with the textcoop platform," *Argumentation and Computation*, vol. 3, no. 1, pp. 49–82, 2012.

[5] M. Fitting, *First-Order Logic and Automated Theorem Proving*, ser. Graduate Texts in Computer Science. Springer, 1996.

[6] L. Moura and N. Bjørner, "Z3: An efficient SMT solver," in *Tools and Algorithms for the Construction and Analysis of Systems*, ser. LNCS, C. Ramakrishnan and J. Rehof, Eds. Springer, 2008, vol. 4963, pp. 337–340.

[7] C. Barrett, A. Stump, and C. Tinelli, "The SMT-LIB Standard: Version 2.0," in *Proc. of the 8th Intl. WS on Satisfiability Modulo Theories*, A. Gupta and D. Kroening, Eds., 2010.

[8] J. D. Kleer, "An assumption-based TMS," *Artificial Intelligence*, vol. 28, pp. 127–162, 1986.

[9] M. Liffiton and A. Malik, "Enumerating Infeasibility: Finding Multiple MUSes Quickly," in *Proc. of CPAIOR*, 2013, pp. 160–175.

[10] W. Raynaut, *Module IA de l'Outil LELIE. Un logiciel intelligent d'aide au diagnostic de risques dans les procédures industrielles*, IRIT, http://www.irit.fr/~Marie-Christine.Lagasquie-Schiex/Lelie, 2013.

[11] M. Järvisalo, D. L. Berre, O. Roussel, and L. Simon, "The international SAT solver competitions," *AI Magazine*, vol. 33, no. 1, pp. 89–92, 2012.

[12] C. Barrett, L. de Moura, and A. Stump, "SMT-COMP: Satisfiability modulo theories competition," in *Computer Aided Verification*. Springer, 2005, pp. 20–23.

[13] S. Lahiri and S. Qadeer, "Back to the future: revisiting precise program verification using SMT solvers," *ACM SIGPLAN Notices*, vol. 43, no. 1, pp. 171–182, 2008.

[14] R. Bruttomesso, A. Cimatti, A. Franzen, A. Griggio, Z. Hanna, A. Nadel, A. Palti, and R. Sebastiani, "A lazy and layered SMT (BV) solver for hard industrial verification problems," in *Computer Aided Verification*. Springer, 2007, pp. 547–560.

[15] P. Gregory, D. Long, M. Fox, and J. Beck, "Planning modulo theories: Extending the planning paradigm." in *ICAPS*, 2012.

[16] C. Ebert and R. Wieringa, "Requirements engineering: Solutions and trends," in *Engineering and Managing Software Requirements*, A. Aurum and C. Wohlin, Eds. http://dx.doi.org/10.1007/3-540-28244-0_20: Springer Berlin Heidelberg, 2005, pp. 453–476.

[17] U. o. Pennsylvania Engineering, "Extracting traceable formal representations from natural language regulatory documents," http://rtg.cis.upenn.edu/extract-fm.