

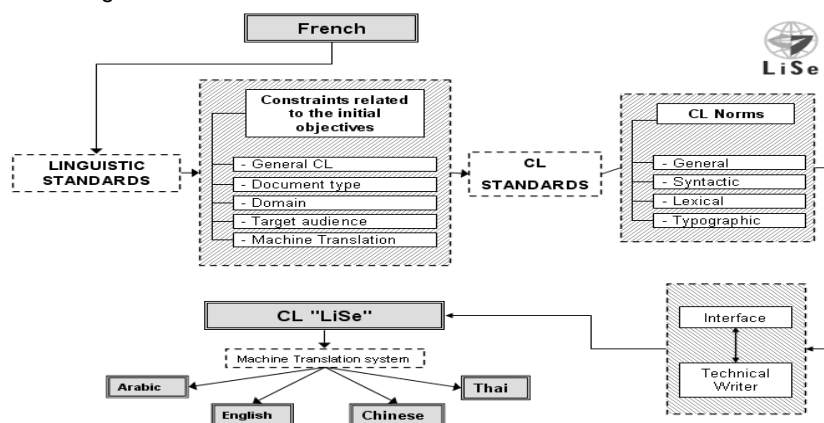
**JULIE RENAHY**  
**DILBER DEVITRE**  
**IZABELLA THOMAS**  
**ALEKSANDRA DZIADKIEWICZ**  
 Centre Lucien Tesnière  
 Research center in Linguistics and in Natural Language Processing  
 Université de Franche-Comté  
 Besançon, France  
 julie.renahy@gmail.com, dilber\_devitre@yahoo.com  
 izabella.thomas@univ-fcomte.fr, aleksandra.d@hotmail.fr

***Controlled language norms for the redaction of security protocols: finding the median between system needs and user acceptability***

**1. Introduction**

Research done on CLs is dominated by the English language; rules simplifying its vocabulary and grammar are developed in order to facilitate and improve comprehension for native and non native English speakers and allow quality translations into other languages. Analogous studies applicable to the French language are rare. This can be explained by the fact that most CLs are created for intercultural communication, that is to say, simplified English is used as the international lingua franca, especially in technical domains. The CL that we have developed, "LiSe", is different in so far as it is addressed to a monolingual public, be it the general public or specialists who communicate in French or in any of the other target languages of the project. Project LiSe<sup>1</sup> fills the void in the field of CLs in French by adding an original contribution, both from a theoretical point of view and its application. The challenge of this project, which began in January 2007, is to create a French controlled language that would facilitate speedy and reliable writing of protocols and alerts in the domain of security and translate these protocols automatically into 4 target languages: Arabic, Chinese, English and Thai. Topics covered range from general security and aeronautics to medicine. Another challenge of this project is to create an easy-to-manipulate interface that allows technical writers to use "LiSe" to its maximum potential without having to learn all the CL rules by heart. This article retraces the problems and the constraints that were taken into account while creating this CL, and the solutions that we propose in order to attain our objectives.

Research has shown that there is no universal CL. Even the numerous CLs that exist in English vary, because their design depends on several criteria. Apart from typical CL constraints such as clarity, readability and non-ambiguity, the design of a CL depends on several other factors such as – 1. the type of document to be written, 2. the domain, 3. the target audience, 4. machine translation (O'Brian 2003). Each of these factors imposes certain constraints. Given below is the diagram that synthesises the CL procedure and the factors taken into consideration while creating "LiSe".



*CL procedure taking into account the initial objectives*

Every linguistic norm is examined closely with respect to each constraint (document type, domain, target audience, machine translation). This gives us a set of controlled language norms, which are subdivided into 4 categories – general, typographical, syntactic and lexical. An interactive user interface is then developed keeping in mind 2 main objectives:

- a) to ensure the quality of the results;
- b) to facilitate the acquisition of "LiSe" and its application (the actual redaction of protocols), especially by a non-linguist.

<sup>1</sup> "Linguistique, normes, traitement automatique des langues et Sécurité : du « data et sense-mining » aux langues contrôlées" (2006-2009), project financed by The French Research Agency (ANR) and currently undertaken at Centre Tesnière, Université de Franche-Comté (France).

The last step (facultative) involves the implementation of a machine translation system that translates protocols written in “LiSe” into the 4 target languages.

## 2. From a normalised French to a controlled French and a machine translation system:

In order to create “LiSe”, we start with a normalised, standard version of French as per the existing linguistic norms (for example: “Write in correct French”, “Do not make any spelling mistakes” etc.). This standard version is then analysed keeping in mind the aforementioned constraints. These constraints often overlap, but we have structured the CL rules into a typology.

### 2.1 General CL constraints (comprehension, readability, non-ambiguity)

The first classification includes general constraints related to the use of any CL, i.e. as a general rule, CLs must be easy to read and understand, and completely non-ambiguous. To meet these objectives, the following restrictions have been imposed:

- a) **Information structuring:** Rules reminding the technical writer to respect the chronological and logical order in structuring the information have been included even though these may seem obvious. Moreover, every sentence can express only one idea or action i.e. it can contain only one piece of information.
- b) **Syntactic control:** First of all, only one verb is allowed per sentence. This rule may seem very restrictive but it is highly appropriate in the context of procedural texts<sup>2</sup>. Not only does it ensure that the CL rule “one sentence = one piece of information” is respected, but it also simplifies syntactic structures to a great extent. Consequently, it is no longer possible to have sentences with coordinate or subordinate clauses. Secondly, authorised verbal structures are limited and predetermined. The number and nature of arguments is predefined for every verbal head word. Therefore, each verbal structure thus defined can have one and only one meaning, and reciprocally, each meaning can be expressed by one and only one verbal structure<sup>3</sup>.

Exemple (1):

*Qqch passer sous quelque chose* (« si la fumée passe sous la porte »)

v/s

*Qqn passer à quelque chose* (« passer à l'étape suivante »)<sup>4</sup>

- c) **Lexical control:** the most important rule is the attribution of one and only one meaning to each lexical headword. This eliminates the possibility of any lexical ambiguity. For example, the word “prise” is polysemous in French but can only be used in its “prise électrique” meaning and not by any other meaning such as “prise de judo” or “prise de médicaments”.<sup>5</sup>

These general constraints are completed with more specific ones that shall be discussed in the following paragraphs.

### 2.2 Constraints related to the type of document

“LiSe” has been created to facilitate the redaction of a specific type of documents i.e. security protocols. We define this as follows:

*A protocol is a text that aims to communicate to an end-user, be it a specialist or a non-specialist, actions that must be executed under certain conditions. The protocol can be written for immediate execution, with different levels of urgency, or in order to instruct the end-user on what he should do in the case of an emergency.*<sup>6</sup>

One must remember that the person to whom the protocol is destined is not simply a ‘reader’, but a ‘doer’. The aim of the protocol is to make the end-user execute precise instructions, and knowing that there are always different ways of expressing the same information, it is primordial to establish structures that are in keeping with the semantic contents that one wishes to express. Moreover, even though most protocols are written to be read before hand, very few end-users bother reading them unless they really have reason to do so.

Protocols, therefore, have an internal structure that has been identified and formalised based on a representative corpus, which has been constituted with texts on civil security found on various official websites, and medical protocols provided by The University Hospital (CHU) in Besançon. There are 5 types of functional structures in protocols and each of these activates a specific, finite subset of syntactic, text formatting and page layout rules. The subset of lexical rules is identical across all functional structures and only differs according to the domain selected:

- a) **Titles:** short sentences or clauses that state the title of the protocol; syntactic structures for titles are restricted to five.
- b) **Sub titles:** short sentences or phrases that introduce specific sections of the protocol; syntactic structures for sub titles are the same as those for titles and differ only in the formatting.
- c) **Instructions:** injunctive sentences that describe an action that must be performed by the end-user. We decided to authorise only verbs in the infinitive form to express an injunction. Therefore every syntactic structure used to write an instruction has an infinitive verb.

<sup>2</sup> See paragraph 2.2.

<sup>3</sup> This constraint is also due to automatic translation, see paragraph 2.4.

<sup>4</sup> “Something passes through under the door” v/s “to pass,(i.e to go) to the next step”

<sup>5</sup> The French word “prise” can mean “a plug”, “a judo hold” or “the intake (of medication etc.)”

<sup>6</sup> This definition is based on Heurley, 2001/2 and Bouffier, 2006.

- d) **Conditions:** this functional structure follows the traditional 'if X then Y' pattern, with its variants. Special rules for co-ordinated conditions (if X and Y, then Z) and disjoint conditions (if X or Y, then Z) have also been included. Conditional syntactic structures differ from injunctive syntactic structures because they allow the use of conjugated verbs.
- e) **Explanatory notes:** this functional structure allows the technical writer to explain the reasons behind certain instructions if they are not self-explanatory, in order to ensure that the end-user does execute the action. Explanatory notes are limited in number and are always preceded by an introductory word such as 'explication', 'aim' ('explication', 'but' in French) etc. and a colon. However, technical writers are reminded to use these explanatory notes sparingly in order to avoid overcharging the text with superfluous information and distracting the attention of the end-user.

	Types of functional structures	Examples of the authorised syntactic structures	Example found in a protocol
a	Title	'Que faire en cas de' + Gn_sa + '?'	<i>Que faire en cas d'incendie ?</i>
b	sub-title	Gn_sa	<i>Ouragan.</i>
c	instruction	opt(neg(Neg)) + Vinf + Arg1+ '.'	<i>Nettoyer les conduits.</i>
d	condition	'Si' + Arg0 + Vconj	<i>Si l'ouragan approche :</i>
e	explanatory note	'Explication' + ' : ' + Gn_sa + '.'	<i>Explication : Risque d'étouffement.</i>

*Examples of functional structures*

Division of the protocol into five distinct types of functional structures ensures that it is executed in the correct order and that the information to be communicated is correctly transmitted. Moreover, associating specific page layout and text formatting rules to each functional structure improves readability of protocols and helps to demarcate clearly and rapidly between the different sections.

### 2.3 Domain related constraints

Constraints imposed by the domain, such as general security (natural catastrophes, accidents, acts of terrorism), aeronautics and the medical domain, are divided into three types:

- a) **Security related constraints:** In high security domains, the need of the hour is precise instructions, immediate comprehension and zero ambiguity because the slightest error can prove fatal.<sup>7</sup>
- b) **Constraints imposed by protocol redaction norms:** CL rules must respect, as far as possible, the prevailing practices that are used in each domain. That is why each rule must be ratified by a specialist of the domain.
- c) **Linguistic constraints:** The lexicon differs for each domain. Authorised syntactic structures are by and large identical across domains but may be modified slightly if needed. For example, every noun can have only one complement in the civil security domain, whereas several more are authorised in the medical domain; here since the order of each complement is important, it is predetermined.

Example (2) 1 seringue de 5 mL d'Alteplase 2 mg / 2 mL.  
2 ampoules de 10 mL de NaCl injectable 0,9 %.<sup>8</sup>

As far as the lexicon is concerned, two different meanings may be attributed to the same word in two different domains.

Example (3) In the civil security domain, *raviver*<sub>1</sub> ('tendre un feu plus vif') :  
*Ne pas raviver<sub>1</sub> les braises d'un barbecue avec de l'alcool.*  
In the medical domain, *raviver*<sub>2</sub> ('remettre à nu la chair vive') :  
*Raviver<sub>2</sub> la plaie.*<sup>9</sup>

Consequentially, even if domain related constraints exist, "Lise"LiSe has been conceived in such a way that it can be easily expanded to allow the writing of protocols in other domains.

### 2.4 Translation related constraints

These constraints aim at ensuring correct machine translations. They also ensure that the zero ambiguity rule is respected, not only in the source language but also in the target languages. This is by no means easy because we have decided to create one unique controlled language in French which will be translated simultaneously into all the target languages of the project. This decision is justified by two criteria – a. redaction time, and b. user ergonomics. Having different controlled languages for each target language would mean having as many source

<sup>7</sup> The fatal consequences of an error due to lack of vigilance and especially faulty translations can be seen in the radiation incident that took place at a hospital in Epinal, France, where 5000 people were over exposed to radiation because of a faulty translation of the technical documents from English to French. 5 of the victims died and several others suffered from irreversible complications (Source (amongst others): *Le scandale de l'hôpital d'Epinal s'aggrave*, Plusnews.fr, 28 March 2007, [http://www.plusnews.fr/articles/000305\\_irradiations\\_cancer.php?part=contexte](http://www.plusnews.fr/articles/000305_irradiations_cancer.php?part=contexte) ).

<sup>8</sup> 1 x 5 mL syringe of 2 mg / 2 mL Alteplase, 2 x 10 mL phials of 0,9 % injectable NaCl

<sup>9</sup> *raviver*<sub>1</sub> = rekindle, (Do not rekindle barbeque embers with alcohol), *raviver*<sub>2</sub> = expose raw flesh, (Expose the raw flesh around the wound).

texts to type out and as many language specific rules. This goes against the urgency criterion that must be considered.

This decision however implies that the impact of every lexical and syntactic rule on every target language must be taken into account.

Translation imposed rules are also divided into three subcategories:

- a) **Syntactic rules:** Syntactic structures have been imposed on source and target languages in order to reduce divergences as far as possible. Certain structures are forbidden if they are difficult to translate into one or several of the target languages. For example, the passive voice is forbidden because of the complications that arise during the transfer to Arabic.
- b) **Lexical rules:** The lexicon is subdivided into two lists, authorised and forbidden, in order to avoid any ambiguity in the target language. For example, the verb 'jeter' is authorised only with a solid verbal argument, whereas 'verser' must be used in case the argument is liquid. This is because Chinese and Thai have two different verbs that make this distinction, whereas French does not.
- c) **Morphological rules:** Verb modes and tenses are limited to the present indicative and the infinitive, so as to avoid problems in translating tenses and aspects.

Every time a constraint has been imposed, an alternative solution is proposed so that the technical writer is never blocked while writing the protocol.

#### 2.5 The interface-technical writer module

Potential users of "LiSe" are not linguists; they are specialists in their respective domains (doctors, people in charge of security etc.), and even though some of them are appointed as technical writers, it is often as a secondary activity. This makes the redaction of protocols in a CL all the more difficult.

In order to facilitate the acquisition of "LiSe" and to ensure that the actual protocol redaction is in accordance with the rules, two tools have been developed: a redaction guide and an interactive redaction help tool.

**The redaction guide**, written itself in "LiSe", contains all the controlled language rules and the forbidden and authorised lexicon.

- a) **'Compagnon LiSe'** is an interactive, easy-to-use help tool that on one hand, automates a large part of the rules, especially punctuation and page layout rules, and on the other hand proposes an interactive aid for non-automated rules. All the user has to do is either :
  - i. enter his sentences in pre-structured text fields. The syntactic structuring that must be respected is thus taken into account by the interface.
  - ii. choose between available options according to what he wishes to write through click buttons or drop-down menus.

The interactive user interface reduces to a great extent the number of CL rules that the technical writer must keep in mind while writing a protocol. Non-automated rules will be displayed in an adjustable window at the bottom of the screen according to the functional structure that the technical writer is typing.



Screen dump of 'Compagnon LiSe'

Key:

Neg = negation, Qqn/Qqc = someone/something, Comp = circumstantial complement

With 'Compagnon LiSe', we have overcome one of the most pertinent criticisms made against CLs, i.e. the difficulty of their application and usability due to the large number of rules that a technical writer must remember and due to restricted lexicons.

### 3. Conclusion

- "LiSe" allows an efficient redaction of security related protocols.
- CL are often criticised as being arduous and long to learn and requiring lengthy training sessions. The intelligent management of rules by the interface greatly reduces the time needed to learn and write protocols in "LiSe", even for non-linguists.
- We have developed a CL in French. However, subsequent research has shown that several "LiSe" rules are language independent and can be applied to other languages. Based on work done for the French CL, current research is being undertaken to develop CLs for protocol redaction in English, Polish and Spanish.

### Acknowledgments

We would like to thank sincerely all our partners, Laurent Spaggiari, Ciprian Melian and especially Dominique Vuitton for her comments and suggestions. Our very special thanks go to Sylviane Cardey-Greenfield and all the

other members of the LiSe team: Raksi Anantalapochai, Mohand Beddar, Peter Greenfield, Gan Jin, Ziad Mikati, Gabriel Sekunda and also Mathias Lambert and Morgane Tornaboni, for cheerfully going through this project.

### References

Bouffier, A., (2006), *Segmentation et structuration de textes procéduraux pour l'aide à la modélisation de connaissances : le rôle de la structure visuelle*. Schadae, prepublication n°10, fascicule n°1, pp. 79-84, [www.unicaen.fr/services/puc/ecrire/preprints/preprint0102006.pdf](http://www.unicaen.fr/services/puc/ecrire/preprints/preprint0102006.pdf)

Cardey S., Greenfield P., Vienney S., (2005), Machine Translation, Controlled Languages and Specialised Languages *Lingvisticae ligature Investigationes*, Benjamins, année 2005, ISSN 0378-4169/E-ISSN 1569-9927.

Heurley, L., (2001/2002), *Compréhension et utilisation de textes procéduraux : l'effet de l'ordre des informations*, *Revue Française de Linguistique Appliquée*, Volume VI, pp 29-46.

O'Brian, S., (2003), *An Analysis of Several Controlled English Rule Sets*. Presented at EAMT/CLAW2003, Dublin, Ireland, 15-17 May 2003, <http://www.ctts.dcu.ie/presentations.html>

Rationalised French (GIFAS), project presentation: [http://www.imdr.eu/v2/extranet/detail\\_projet.php?id=19](http://www.imdr.eu/v2/extranet/detail_projet.php?id=19)

Reuther, U., (2003), *Two in One – Can it Work? Readability and Translatability by means of Controlled Language*. Presented at EAMT/CLAW2003, Dublin, Ireland, 15-17 May 2003, <http://www.ctts.dcu.ie/presentations.html>

Spaggiari, L., Beaujard, F., Cannesson, E., (2003), "A Controlled Language at Airbus", *Proceedings of the Joint Conference combining the 8th International Workshop of the European Association for Machine Translation and the 4th Controlled Language Application Workshop*, pp. 151-159.