



Resources for Controlled Languages for Alert Messages and Protocols in the European Perspective

Sylviane Cardey¹, Krzysztof Bogacki², Xavier Blanco³, Ruslan Mitkov⁴

¹Centre Tesnière, Université de Franche-Comté (FR)

²Uniwersytet Warszawski (PL)

³Universitat Autònoma de Barcelona (ES)

⁴RILP, University of Wolverhampton (GB)



JLS/2007/CIPS/022

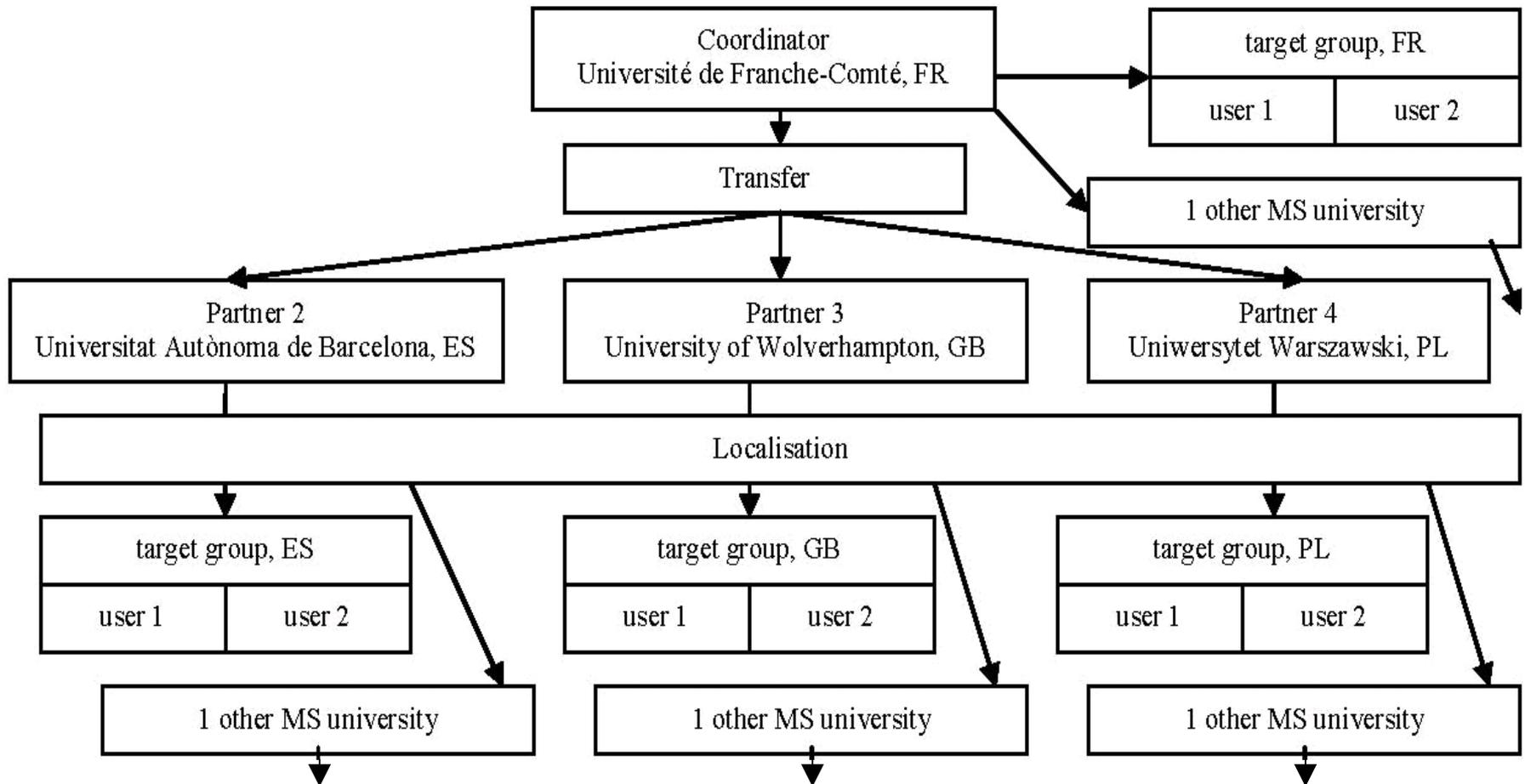
With the support of the Prevention, Preparedness and Consequence Management of Terrorism and other Security-related Risks Programme European Commission - Directorate-General Justice, Freedom and Security

This project has been funded with support from the European Commission. This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein

The MESSAGE Project

The MESSAGE project (full name Alert Messages and Protocols) <http://message-project.univ-fcomte.fr> involved the development and transfer of a methodology created by the coordinator Centre Tesnière for writing safe and safely translatable alert messages and protocols by means of a consortium of four partners to their four European member states in their languages (ES, FR (Coordinator), GB, PL), in such a manner so that this transfer could be extended to other member states.

The MESSAGE Project's Transfer of Controlled Language (CL) Technology



Target Groups

The target groups concerned were drawn from the project consortium's member states' law enforcement and emergency community involved in writing, controlling and evaluating alert messages and protocols, these latter being aimed at variously professionals and the general public (who could be citizens of other member states).

The specific target groups concerned were:

- law enforcement
- local government
- civil protection
- fire fighting
- transport
- aeronautics
- meteorology
- chemistry
- emergency medical personnel

[LREC 2010](#)

Dissemination

- The MESSAGE project provided training courses in controlled languages delivered by the Coordinator and targeted at linguists in the 3 other partners who in turn localised and trained target group personnel in four member states (ES Spain, FR France, GB Great Britain, PL Poland).
- The project contributed in part to ISMTCL - International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages <http://www.ismtcl.org/> where there were target group delegates not only from the consortium's member states but also from:
 - AT Austria
 - BG Bulgaria
 - CZ Czech Republic
 - GR Greece
 - IT Italy
 - PT Portugal
 - RO Romania.
- The project prompted research concerning the transfer to other EU languages (Bulgarian, Greek).

[LREC 2010](#)

Resources for Controlled Languages for Alert Messages and Protocols in the European Perspective

Sylviane Cardey, Krzysztof Bogacki, Xavier Blanco, Ruslan Mitkov

Underlying Methodology

The underlying methodology was developed by Centre Tesnière in collaboration with the aircraft industry, the health profession, and emergency services within the French National Research Agency (ANR) funded LiSe project (ANR-06-SECU-007) for establishing standards concerning the writing of alert messages and protocols for safety-critical applications and which can be safely translated.

<http://projet-lise.univ-fcomte.fr>

Alert Messages and Protocols

Alert Message

An alert message is a message transmitted by local or national authorities to the general population (civil or professional) to warn them of an impending emergency:

- meteorological emergencies
- industrial disaster
- terrorist threats
- abduction etc.

Alert messages can be distributed by means of:

- radio and television stations
- traffic variable-message signs
- digital signage
- GPS traffic systems
- webpages
- e-mail etc.

Protocol

A protocol is a text that aims to communicate to a specialist or a non-specialist, actions that must be executed under certain conditions. Protocols can be written for immediate execution, with different levels of urgency, or to instruct on what to do in the case of an emergency.

Controlled Languages for Safety and Security (i)

As the human is an integral part of numerous systems, the weakest link in many of these is human communication, e.g.:

- the Tenerife air crash of 1977:

The Dutch pilot of the Boeing 747, in saying "we are at take-off", because he used a similar structure in Dutch, thought that he had indicated that he was ready for take-off whilst in fact he had indicated only that he was in position for take-off. He thus interpreted, incorrectly, the approval by the control tower as an authorisation for take-off. 583 people lost their lives.

- Because of the lack of rapid translation of instructions on medicines sent to Asia following the 2004 Indian Ocean tsunami, these same medicines could not be used.

Controlled Languages for Safety and Security (ii)

- The audiences, the languages and thus the way of writing messages concerning threats and crises vary between the law enforcement and emergency communities of the different member states of the EU.
- At the moment of the threat, which cannot be predicted, alert messages and protocols have to be composed rapidly and these must be clear and unambiguous to the appropriate recipient. Certain of these need to be translated rapidly to one or more other languages of the EU.

Controlled Languages for Safety and Security (iii)

- Without control, inherent ambiguity and variability (e.g. the 'general' language of the civil population opposed to the 'technical' languages of the law enforcement and emergency community) renders interoperable, compatible and portable solutions difficult and quality machine translation impossible.
- Controlled languages (CL) are the means to overcome these problems; they respect standards of writing, lexis and syntax, and can simplify translation and also aid in quality machine translation because, due to the lack of time in a crisis, one can neither 'train' trainable machine translation systems, nor do manual 'pre-edition'.

Controlled Language Evaluation (i)

The CL development methodology follows Centre Tesnière's systemic linguistic analysis. Not only everything must be explicit but also justified, which in this case is by corpus examples which must be representative of both general and particular cases, with samples furnished by the alert messages and protocols writers.

- **internal evaluation** was carried out by means of university students who served as 'guinea pigs' whilst
- **external evaluation** was undertaken by target group personnel who had attended the project's training courses on writing, controlling and evaluating alert messages and protocols. They put this into practice on real alert messages and protocols within their communities.
- uncontrolled and controlled alert messages and protocols were submitted to existing machine translation (MT) systems.

Controlled Language Evaluation (ii)

The criteria and methods were:

- Criteria used for evaluation:
 - quality of the controlled texts
 - quality of the translation
 - the facility with which the CL can be used
- CL evaluation methods:
 - feedback questionnaires filled in by target groups
 - comparison of translations obtained before and after language control
 - time taken to learn the CL
 - time taken by the Target Groups to learn the CL rules and apply them to new texts
- Criteria for assessing that the CL improves human translation and Machine Translation (MT):
 - time measured for human translation
 - time measured for post-editing MT
 - feedback of linguists and target groups about the quality of the MT results
 - amount of changes to the MT documents done by human post-editors using edit distance measures

[LREC 2010](#)

Resources Produced in the MESSAGE Project & contributed to the [LREC2010 Map](#)

“Freely available” Resources

In English, French, Polish and Spanish) at <http://message-project.univ-fcomte.fr>:

- Two are subject to a [Creative Commons Attribution 2.0 France License](#):
 - ‘*Standards for Controlled Languages*’: European member state and specific language independent – meta standards
 - ‘*Add MS Kit*’: transfer with localisation to other European member states,
- Project leaflets
- Extracts from alert message and protocols writing manuals (“From Owner” resources)
- LiSe and MESSAGE projects poster (in English).

“From Owner” Resources

Owner: MESSAGE project consortium - contact message-project@univ-fcomte.fr:

- Course material aimed at linguists covering controlled languages and their evaluation and localisation (written in English and French)
- Course material, writing manual and evaluation techniques aimed at Target Group writers of controlled alert messages and protocols (these written in and localised for the four languages of the partners’ member states).

(Meta) Standards for Controlled Languages

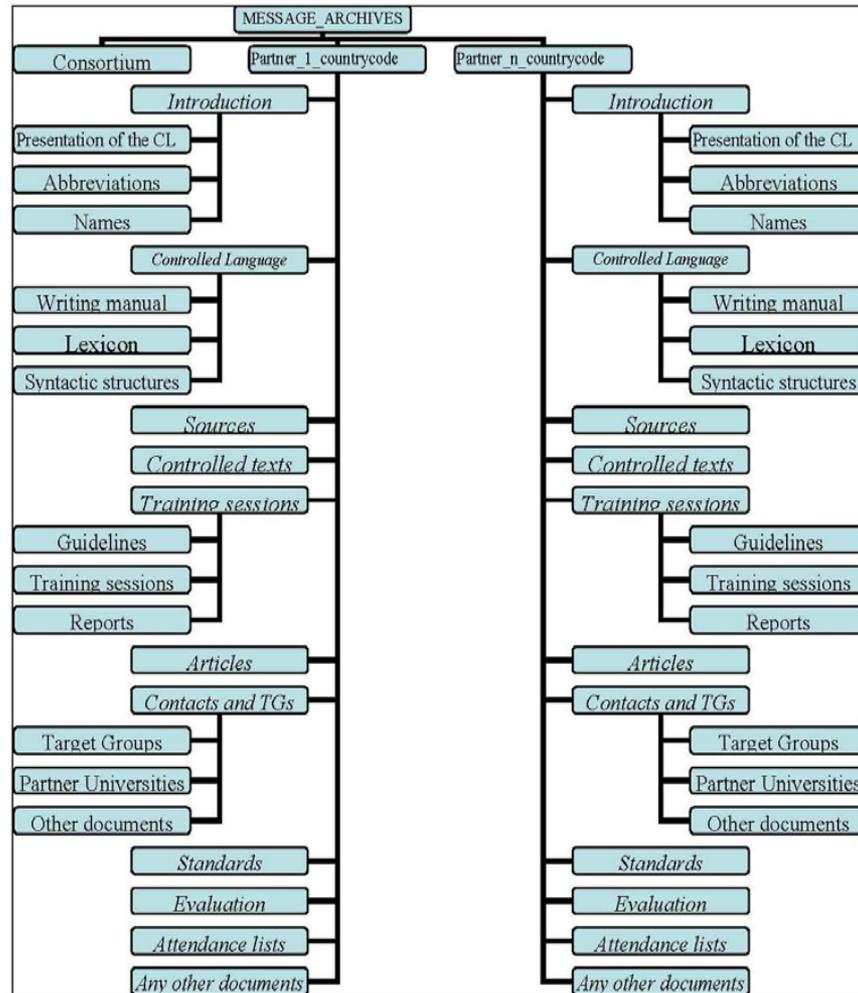
- I. Format standards
 - i. Document standards
 - ii. Standards for archiving the project
- II. Procedural Standards (*in the 'Add MS Kit'*)
- III. CL specific standards
- IV. Linguistic/language standards
- V. Archive organisation standards
- VI. Appendices
- VII. Notations

Language independent CL standards to be respected when creating a CL (English version)

- Identify the domains.
- Identify the language.
- Identify the nature of the text to be controlled.
- Collect a representative corpus of domain related documents in the language.
- Apply general CL rules as and where applicable.
- Create language specific rules.
- Always explain the reasons behind the rules.
- Explanation: This is done for the linguist's benefit in order to trace the CL. A simple guide without rule motivation may be created for non linguists, e.g. technical authors.
- Create a list of permitted lexicon and permitted syntactic structures.
- Assign a unique identifier to every rule.
- If rule identifiers are concatenations of different abbreviations:
 - Please define all abbreviations.
 - Please provide explicit examples alongside every rule.
- If you explicitly forbid a particular word or syntactic structure:
 - Always provide a permitted substitution for forbidden structures/words.
- Check the CL with domain specialists.

Schematic representation of the MESSAGE_ARCHIVES

Serves as a model for the archive organisation of a new project involving the transfer of the Controlled Language technology to another member state (part of 'Add MS Kit')



The member state/language specific standards are organised in a common manner, independent of member state and language

LREC 2010

“Add MS Kit” Content

I. Requirements

i. Human resources: *skills & role; for example:*

*Linguists trained in Controlled Language (CL) techniques,
Linguists fluent in the MS language*

*Technical authors from the emergency services/Professionals involved in writing
security protocols and alert messages*

ii. Material Resources: *corpora & lexical resources*

II. Procedure

i. Preliminary steps: *CL training session by linguists trained in CL to MS linguists (if
needed), Identify the domains in which CL must be developed*

ii. Contact: *Identify potential Target Groups, Initiate contact with Target Groups*

iii. Localisation

iv. Outreach *Training sessions with professionals/technical authors from Target
Groups*

v. Further Dissemination

III. Appendices

i. Abbreviations

ii. Terminology definitions

IV. Notations

Add MS Kit Calendar (English version)

MESSAGE (Alert Messages and Protocols)
Step by step procedure of addition of a MS
Legend: MS = Member State, CL = Controlled Language, TG = Target Group

Project title: MESSAGE (Alert Messages and Protocols)											
Activity	MONTHS										
	1	2	3	4	5	6	7	8	9	10	11
1 Define the consortium	X										
2 Create a set of presentation standards for all official documents, power-point presentations and correspondences	X										
3 Identify the domain in which CL must be developed.		X									
4 CL training session by linguists trained in CL to MS linguists (if needed)											
4.1 Create training material		X									
4.2 Delivery		X									
5 Identify potential TGs.		X	X	X							
6 Initiate contact with TGs			X	X	X						
7 Localisation of general CL rules to MS language by MS linguists			X	X	X						
8 Training session with professionals/technical writers from TGs						X	X				
9 Create network of target group personnel							X				
10 Target group tests/evaluation								X	X	X	
11 Further Dissemination											X

LREC 2010

MESSAGE Training Sessions aimed at Target Group writers

Extract from the MESSAGE Training Sessions GUIDELINES GB “From Owner” resource

MESSAGE Training Sessions have the aim to teach to emergency services writers how to adapt the Controlled Language rules for their purposes.

The Training Sessions are composed of:

- theoretical sessions (TS)
 - TS1 «Introduction to Human Comprehension in Emergency Domains and Controlled Languages»
 - TS2 «The CLCM framework and the controlled language rules for general population»
- practical sessions (PS)
 - PS1 «Re-writing a given text by the facilitator text following the existing controlled language rules»
 - PS2 «First steps into developing controlled language rules for your needs»

Extracts from the CL Writing Manuals (“From Owner” resources)

Extract from the CL Writing Manual (FR)	
Pour indiquer la matière d'un objet : Utiliser la préposition « en ».	<p>1 feuille de papier.</p> <p>→</p> <p>1 feuille en papier.</p>
Extract from the CL Writing Manual (PL)	
Jeśli piszesz instrukcję: Używaj trybu rozkazującego lub bezokolicznika.	<p>Powinieneś przestrzegać zasad pisania protokołów.</p> <p>→</p> <p>Przestrzegaj zasad pisania protokołów.</p> <p>Przestrzegać zasad pisania protokołów.</p>

Conclusion

- The transfer of this controlled language methodology has permitted sharing and extending the methodology developed by the coordinator at the European level.
- The harmonisation of formal and linguistic norms at the European level by means of controlled languages is a reliable way to assure a sure, efficient, quick and reliable way of diffusing protocols and alert messages at the EU level.
- The controlled language resources which are the outcome of the MESSAGE project are available in two forms “Freely available” and “From Owner”, and have been contributed to the [LREC2010 Map](#).
- As a result one has a much clearer view on the possibilities for effective harmonisation at the EU level.



Resources for Controlled Languages for Alert Messages and Protocols in the European Perspective

Sylviane Cardey¹, Krzysztof Bogacki², Xavier Blanco³, Ruslan Mitkov⁴

¹Centre Tesnière, Université de Franche-Comté (FR)

²Uniwersytet Warszawski (PL)

³Universitat Autònoma de Barcelona (ES)

⁴RILP, University of Wolverhampton (GB)



JLS/2007/CIPS/022

With the support of the Prevention, Preparedness and Consequence Management of Terrorism and other Security-related Risks Programme European Commission - Directorate-General Justice, Freedom and Security

This project has been funded with support from the European Commission. This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein