

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)

E-mail: sylviane.cardey@univ-fcomte.fr, kbogacki@gmail.com, Xavier.Blanco@uab.es, r.mitkov@wlv.ac.uk

Abstract

This paper is concerned with resources for controlled languages for alert messages and protocols in the European perspective. These resources have been produced as the outcome of a project (Alert Messages and Protocols: MESSAGE) which has been funded with the support of the European Commission - Directorate-General Justice, Freedom and Security, and with the specific objective of 'promoting and supporting the development of security standards, and an exchange of know-how and experience on protection of people'. The MESSAGE project involved the development and transfer of a methodology for writing safe and safely translatable alert messages and protocols created by Centre Tesnière in collaboration with the aircraft industry, the health profession, and emergency services by means of a consortium of four partners to their four European member states in their languages (ES, FR (Coordinator), GB, PL). The paper describes alert messages and protocols, controlled languages for safety and security, the target groups involved, controlled language evaluation, dissemination, the resources that are available, both "Freely available" and "From Owner", together with illustrations of the resources, and the potential transferability to other sectors and users.

1. Introduction

This paper is concerned with resources which have resulted from work undertaken during the MESSAGE project¹ which involved the development and transfer of a methodology for writing safe and safely translatable alert messages and protocols created by Centre Tesnière (Cardey, 2009a) by means of a consortium of four partners here represented by the authors 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 (see Figure 1).

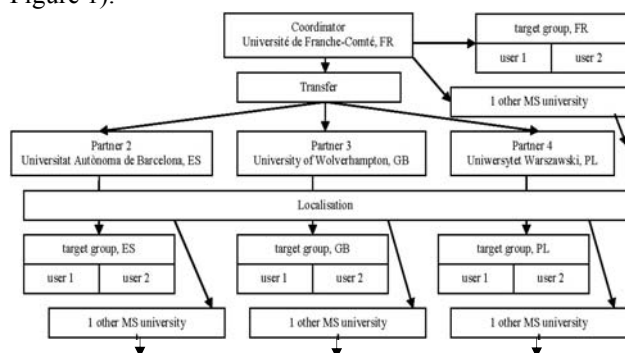


Figure 1: The MESSAGE project's transfer of Controlled Language technology

¹ MESSAGE: Full name Alert Messages and Protocols, 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. See <http://message-project.univ-fcomte.fr>

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 for example). The underlying methodology was itself initially developed by Centre Tesnière in collaboration with the aircraft industry, the health profession, and emergency services within the LiSe project² for establishing standards concerning the writing of alert messages and protocols for safety-critical applications and which can be safely translated (Spaggiari et al., 2005; Renahy, 2009).

2. Alert Messages and Protocols

2.1 Alert Message

An alert message is a message transmitted by local or national authorities to the general population (civilian or professional) to warn them of an impending emergency. Alert messages may be used in the following situations (non exhaustive list):

- meteorological emergencies
- industrial disaster
- terrorist threats
- abduction

² LiSe: Full name « Linguistique, normes, traitement automatique des langues et Sécurité: du « data et sense-mining » aux langues contrôlées » ANR-06-SECU-007 Project financed by the French Research Agency (ANR). See <http://projet-lise.univ-fcomte.fr>

Alert messages can be distributed through different means (non exhaustive list):

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

2.2 Protocol

A protocol is a text that aims to communicate to an end user who can be 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 in order to instruct end-users on what they should do in the case of an emergency.

3. Controlled Languages for Safety and Security

The human being an integral part of numerous systems, the weakest link in many of these is human communication. The Tenerife air crash of 1977 is a convincing example. 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.

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. 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 (Bogacki 2009; Rudas 2009; Temnikova & Orasan, 2009) because, in the context of a crisis, due to the lack of time, one can neither 'train' trainable machine translation systems, nor do manual 'pre-edition'. Furthermore, CL texts can themselves be useful resources in their own right (Catena et al., 2009; Gwiazdecka, 2009).

4. Target Groups

The target groups concerned within the project 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 being aimed at variously professionals and the general public. The specific target groups concerned were as follows:

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

5. Controlled Language Evaluation

The evaluation of CLs developed using the CL development methodology follows directly that of systemic linguistic analysis (Cardey et al., 2006). 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. For CLs, the justifications are made with samples furnished by the alert messages and protocols writers.

In the MESSAGE project, internal evaluation was carried out by means of university students who served as 'guinea pigs' whilst external evaluation has been undertaken by target group personnel who have attended the project's training courses on writing, controlling and evaluating alert messages and protocols. They have put this into practice on real alert messages and protocols within their communities. Furthermore, uncontrolled and controlled alert messages and protocols have been submitted to existing machine translation (MT) systems. The CL evaluation criteria and methods were as follows:

- 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 translation 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 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

6. 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 (see <http://www.ismtcl.org>) (Cardey, 2009b) 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 (Temnikova & Margova, 2009; Papadopoulou et al., 2009; Puig et al., 2009).

7. Resources

We describe here firstly the resources' availability and we then provide more details and illustrations concerning certain of the resources.

7.1 Resources Availability

The resources which have been produced in the MESSAGE project are available in two forms: "Freely available" and "From Owner".

7.1.1. "Freely available" Resources

These resources (in English, French, Polish and Spanish) are available on the project's web-site (<http://message-project.univ-fcomte.fr>):

- Two resources which are subject to a [Creative Commons Attribution 2.0 France License](#):
 - o 'Standards for Controlled Languages' for writing safe and safely translatable alert messages and protocols in the context of terrorism and other security related risks. These standards are European member state and specific language independent; the member state/language specific standards developed are "From Owner" Resources.
 - o 'Add MS Kit' explaining how the transfer with localisation can be achieved to other European member states, this together with an 'Add MS Kit Calendar'.
- Project leaflets.
- Extracts from alert message and protocols writing manuals (these latter being "From Owner" resources)
- LiSe and MESSAGE projects poster (in English).

7.1.2. "From Owner" Resources

The Owner of these resources is the MESSAGE project consortium. Interested parties are requested to contact message-project@univ-fcomte.fr. These resources are:

- 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).

7.2 The Resources in Further Detail

One of the aims of the MESSAGE Project was to establish standards concerning the writing of alert messages and protocols for safety-critical applications. This is achieved through the creation of controlled languages (CLs) for particular domains. An established CL is by itself a standard on how to write a text as per a given set of rules. The "Freely available" 'Standards for Controlled Languages' for writing safe and safely translatable alert messages and protocols are European member state and specific language independent; and the member state/language specific standards are "From Owner" Resources. The 'Standards for Controlled Languages' contains meta standards that must be respected in order to create a CL, and these are listed in Figure 2 (English version).

- | |
|---|
| <ul style="list-style-type: none">I. Format standards<ul style="list-style-type: none">i. Document standardsii. Standards for archiving the projectII. Procedural Standards (<i>in the 'Add MS Kit'</i>)III. CL specific standardsIV. Linguistic/language standardsV. Archive organisation standardsVI. AppendicesVII. Notations |
|---|

Figure 2: Standards for Controlled Languages

The CL specific standards are language independent and must be respected when creating a CL; these are reproduced in Figure 3 (English version).

- | |
|--|
| <ul style="list-style-type: none">• 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.<ul style="list-style-type: none">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:<ul style="list-style-type: none">Please define all abbreviations.• Please provide explicit examples alongside every rule.• If you explicitly forbid a particular word or syntactic structure:<ul style="list-style-type: none">Always provide a permitted substitution for forbidden structures/words.• Check the CL with domain specialists. |
|--|

Figure 3: Language independent CL standards to be respected when creating a CL

The ‘Standards for Controlled Languages’ also contains a schematic representation of the MESSAGE_ARCHIVES (see Figure 4). For a new project involving for example the transfer of the Controlled Language technology to another member state (see the “Freely available” ‘Add MS Kit’), this representation can serve as a model for the new project’s archive organisation.

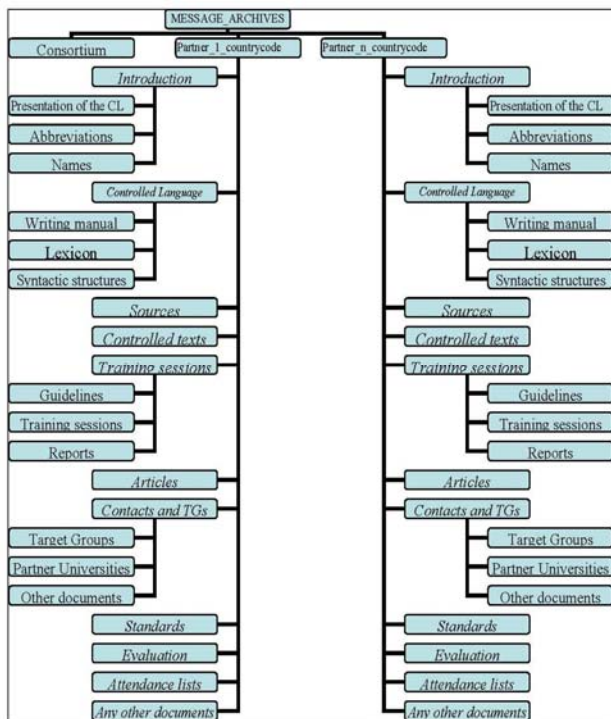


Figure 4: Schematic representation of the MESSAGE_ARCHIVES

As will be seen in the MESSAGE_ARCHIVES (Figure 4), the member state/language specific standards are organised in a common manner, independent of member state and language.

The “Freely available” ‘Add MS Kit’ is a procedure and thus a standard in its own right describing how, within a prospective project the transfer of the Controlled Language technology together with the necessary localisation can be achieved to other European member states (for example see Figure 1). The content of the ‘Add MS Kit’ is shown in Fig 5 with brief explanations in *italics* (English version).

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

Figure 5: Content of the “Add MS Kit”

The “Freely available” ‘Add MS Kit Calendar’ is provided as aid to project planning, and contains suggested values for the durations of the various activities. This is reproduced here in Table 1 (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

Table 1: Add MS Kit Calendar

Concerning the “From Owner” resources ‘*Course material targeted at linguists*’, this includes course material (theoretical and practical) covering variously controlled languages, how to control a text, lexical control, controlled language rules and controlling for translation.

Concerning the “From Owner” resources ‘*Course material, writing manual and evaluation techniques aimed at Target Group writers of controlled alert messages and protocols*’, we reproduce here as an example in Figure 6 the description of the MESSAGE Training Sessions aimed at Target Group writers (extract from the MESSAGE Training Sessions GUIDELINES GB).

<p>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 by:</p> <ul style="list-style-type: none"> - theoretical sessions (TS) <ul style="list-style-type: none"> - 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) <ul style="list-style-type: none"> - 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»

Figure 6: MESSAGE Training Sessions aimed at Target Group writers (GB)

As a further example Table 2 contains extracts from the alert messages and protocols writing manuals resulting from the project and aimed at Target Group writers of controlled alert messages and protocols.

Extracts from the CL Writing Manual (ES)	
Evite los adjetivos posesivos.	<p>Mantenga su cabeza girada. Llame al teléfono de emergencias 112.</p> <p>→</p> <p>Mantenga la cabeza del paciente girada. Llame al teléfono de emergencias 112.</p>
Adecúe el protocolo al vocabulario específico de la disciplina.	<p>Si la víctima está inconsciente: Coloque a la víctima de lado.</p> <p>→</p> <p>Si la víctima está inconsciente : Coloque a la víctima en Posición Lateral de Seguridad.</p>

Extracts from the CL Writing Manual (FR)	
<p>Quand vous rédigez une instruction :</p> <p>Utiliser l'infinitif présent. Exprimer 1 action uniquement.</p>	<p>Se moucher avec des mouchoirs en papier à usage unique jetés dans une poubelle³</p> <p>→</p> <p>Quand vous vous mouchez : Utiliser un mouchoir à usage unique en papier. Jeter le mouchoir à la poubelle.</p>
<p>Pour indiquer la matière d'un objet :</p> <p>Utiliser la préposition « en ».</p>	<p>1 feuille de papier.</p> <p>→</p> <p>1 feuille en papier.</p>
Extracts from the CL Writing Manual (GB)	
<p>If there are sub-sections:</p> <p>Write the first sub-section title. Jump 2 new lines before the first sub-section title. Write the subsection (conditions, instructions, comments). Write the next sub-section titles. Jump 2 new lines after each subtitle. Jump 2 new lines after each subsection.</p>	<p>Chemical attack</p> <p>Attack outdoors Go to the closest building. Take shelter quickly. Close all windows and doors.</p> <p>Attack indoors Follow chemical attack plans: Open windows. Breathe fresh air.</p>
<p>Avoid passive voice.</p>	<p>Make sure 999 is called.</p> <p>→</p> <p>Call 999.</p>
Extracts from the CL Writing Manual (PL)	
<p>Jeśli piszesz instrukcję: Używaj trybu rozkazującego lub bezokolicznika.</p>	<p>Powinieneś przestrzegać zasad pisania protokołów.</p> <p>→</p> <p>Przestrzegaj zasad pisania protokołów. Przestrzegaj zasad pisania protokołów.</p>
<p>Jeśli używasz struktury "Jak + Struktura Nakazująca(Vinf) + ?" : Używaj tylko formy twierdzącej.</p>	<p>Jak nie zarazić pacjentów?</p> <p>→</p> <p>Jak chronić pacjentów przed zarażeniem? Jak uniknąć zarażenia pacjentów?</p>

Table 2: Extracts from the alert messages and protocols writing manuals

8. Potential Transferability to Other Sectors and Users

The main aim of the MESSAGE project was to transfer a ‘general’ Controlled Language (CL) to the Partner countries, i.e., a CL that can be adapted to different domains by making certain changes (adaptation of the constraints to professional practices and terminological work). The project has shown that the methodology developed by the Coordinator (for French) is adaptable to 3 other languages of the EU (English, Polish, Spanish). In addition, experiments conducted by the ES and GB

³ Extract from Plan national « Pandémie grippale », Fiche C.2 – Principales règles d’hygiène face au risque épidémique, 2009/04/29 http://www.pandemie-grippale.gouv.fr/IMG/pdf/Fiche_C2-4.pdf

partners showed that the technology can be easily used for respectively Greek and Bulgarian. Experiments conducted by the Coordinator have shown that the technology can easily be transferred to Portuguese and also to Persian (Farsi). This permits one to be optimistic about the adaption to other languages of the EU, and furthermore, to languages other than EU languages.

9. Conclusion

The transfer of this controlled language methodology towards the 3 project partner countries (ES, GB, PL) through the training sessions dispensed by the MESSAGE project coordinator (FR) to the partners and thereafter by all of the partners to the target groups has permitted sharing and extending the methodology developed by the coordinator at the European level. The harmonisation of formal and linguistic norms at a European level, achieved through the standardisation of the lexicon, syntax and semantics is a reliable way to assure a sure, efficient, quick and reliable way of diffusing protocols and alert messages at the EU level and hence to assure an optimal level of security.

The controlled language resources which are the outcome of the MESSAGE project are available in two forms "Freely available" and "From Owner". Multilingualism and transfer are explicitly treated in these resources which are intended to ensure an interoperable, compatible and portable approach aiding subsequent extension to other member states and their languages. As a result one will have a much clearer view on the possibilities for effective harmonisation at the EU level.

10. Acknowledgements

The authors wish to acknowledge in particular Peter Greenfield, Julie Renahy, Angels Catena, Joanna Cholewa, Dilber DeVitre, Ewa Gwiazdecka, Constantin Orasan, Lena Papadopoulou, Zuzanna Rudas, Gabriel Sekunda and Irina Temnikova, and also the members of the various and varied target groups without whose interest, involvement and constructive observations the project would not have been possible.

The MESSAGE project, full name Alert Messages and Protocols, 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.

11. References

- Blanco, X. (2009). Remarks about Linguistic Analysis, Normalization and Translation of Spanish "What to Do in Case of Fire" Texts. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 43-48.
- Bogacki, K. (2009). Controlled Languages and Machine Translation. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 49-55.
- Cardey, S., Greenfield, P. (2006). Systemic Linguistics with Applications. In E.M.Bermúdez and L.R. Miyares (Eds.), *Linguistics in the Twenty First Century*. Cambridge Scholars Press, United Kingdom. ISBN 1904303862, 2006, pp. 261-271
- Cardey, S. (2009a). Controlled Languages for More Reliable Human Communication in Safety Critical Domains. In *Proceedings of the 11th International Symposium on Social Communication*, Santiago de Cuba, Cuba, 19-23 January 2009, pp. 330-335.
- Cardey, S. (2009b). Preface. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp ix-x.
- Catena, A., Demenkova, M. (2009). Construction and Application of Indexes for Medical Record Classification. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 90-93.
- Cholewa, J. (2009). Remarks Concerning Writing Texts in Controlled Polish. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 62-68
- Gwiazdecka, E. (2009). Annotation of Terminology from Protocols in Polish Controlled Language. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 121-126.
- Papadopoulou, E., Puig, M. (2009). Abduction Alerts in Greek and Spanish. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 185-189.
- Puig, P., Papadopoulou, E. (2009). Treatment of the Imperative Forms in Machine Translation between Catalan, Spanish and Greek. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 198-202.
- Renahy, R. (2009). Controlled Languages: a Scientific Popularization through the Example of the Controlled Language "LiSe". In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 215-222
- Rudas, Z. (2009). Polish Controlled Language and its Machine Translation into French. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 231-235.
- Spaggiari, L., Beaujard, F. Cannesson, E. (2005). A controlled language at Airbus. In *Lingvisticae Investigationes*, Special issue: Machine Translation, Controlled Languages and Specialised Languages; Tome XXVIII :I (2005), John Benjamins, ISSN 0378-4169, E-ISSN 1569-9927, pp 107-122.
- Temnikova, I., Margova, R. (2009). Towards a Controlled Language in Crisis Management: The Case of Bulgarian. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 244-248.
- Temnikova, I., Orasan, C. (2009). Post-editing Experiments with MT for a Controlled Language. In *Proceedings of the ISMTCL*, PUFC, ISBN 978-2-84867-261-8, pp 249- 255.