Towards a Controlled Language in Crisis Management: The Case of Bulgarian

Irina Temnikova* and Ruslana Margova

Research Institute in Information and Language Processing
University of Wolverhampton, UK*, 'Geomedia' Journal, Bulgaria

E-mail: I.Temnikova2@wlv.ac.uk*, naruslana@yahoo.com

Abstract

This paper presents a first attempt to develop controlled language rules for Bulgarian for writing messages and instructions for the general public in emergency-related situations. The research is based on a collected corpus of emergency instructions and on parallels with two controlled languages for Greek and Polish. The paper presents the linguistic investigations and the rules derived and also provides an example of a re-written text.

1. Introduction

In the past three years in Bulgaria several big natural and man-made disasters happened which made an impact on the lives of the population. In August 2007 the dam burst of an artificial lake provoked by rains near the city of Tzar Kalojan killed 8 people. In February 2008 eight people were burned alive in the Sofia-Kardam train by an unidentified blaze. In July 2009 in the neighborhood of Sofia – Chelopechene, a military storage facility for ammunition equivalent to 20 tons of trotyl exploded. Fortunately there were no victims. Every summer in the southern part of the country forest fires burn thousands of hectares. Every spring floods inundate many villages and cities all over the country.

For all these calamities there are no early alerts and adequate signals. Information was always presented after the event and only by the media without any participation of the government. Only in March 2009 did the Bulgarian government accept a decree for developing a national early alerting system for emergencies and disasters. The system has to be developed by the Ministry of Emergency Situations by August 2009.

In emergency situations, the clarity of messages transmitted to the population is crucial. This paper aims to assist the development of a European-level early alert system in Bulgaria by providing a set of guidelines for writing clear, concise and unambiguous messages for the general public in emergency-related situations. It also aims to explore whether the controlled language technology developed in the context of Message Project is easily transferable to languages other than French [1].

The paper is structured as follows: Section 1 introduces the general situation with emergency alerts in Bulgaria, Section 2 provides the general context of our controlled language, Section 3 discusses the relevant corpus findings and the controlled language rules, Section 4 presents some results of applying the controlled language rules, Section 5 gives the conclusions and features some future work.

2. Controlled languages

Controlled languages (CLs) are specially designed natural language subsets that obey syntactic and lexical restrictions. The main purposes of controlled languages are to eliminate or reduce natural language complexity.

*Full title: Alert Messages and Protocols, project financed by the European Union (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 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.
and ambiguity, standardize language irregularity and improve document translatability and the retrievability of information.

CLs can be human-oriented and machine-oriented [2]. Machine-oriented CLs aim to improve the performance of NLP applications. Controlled languages for humans can be very general, such as the Plain English Campaign2, or more detailed, such as the French controlled language developed in the context of Message Project. Controlled language restrictions depend on the particular application of the concrete controlled language. The main application field of controlled languages is usually technical documentation. Controlled languages are developed mainly for English, but examples exist also for other languages, such as Esperanto, French, German, Swedish, Spanish, Japanese, Mandarin and Modern Greek [3]. A controlled language for Polish [4] is currently being developed in the context of the Message Project, a European Union funded project which involves developing a controlled language technology for the emergency-related domain with particular focus on easy transferability to new languages.

The current research is related to situations in which human understanding is crucial, thus keeps into account psycholinguistic findings such as the working memory overload [5]. The controlled language we are currently developing has both prescriptive and proscriptive rules and imposes restrictions at all levels: starting from the general formatting and structure of documents and ending with the morphological, lexical and syntactic levels.

No previous attempt for a controlled language for Bulgarian has been ever published. The current research is based on a corpus analysis and on parallels with the existing controlled languages for Modern Greek [6], which is a Balkan language, and Polish [4] – a typical Slavonic language, due to their partial similarity with Bulgarian.

3. Particularities of Bulgarian and the controlled language rules

Bulgarian is an Indo-European language, a member of the Southern branch of the Slavonic languages having characteristics of both Slavonic and Balkan languages.

The current controlled language is based on a study of a corpus containing existing alerts, messages and instructions in Bulgarian for emergency-related situations. The topics cover a variety of typical situations, varying from fires to earthquakes and radioactive fallouts and are written to address both crisis management specialists and the general population. The corpus has been divided in two parts - one for conducting the analysis and one for testing the prepared re-writing rules.

Due to the fact that it shares characteristics common to both the Slavonic and the Balkan languages, the Bulgarian language exhibits rich complexity at all levels. The current study takes into account only the linguistic phenomena of key importance for the particular controlled language under development. The key linguistic phenomena are of a lexical, morphological and syntactical nature.

3.1. Lexical complexity phenomena and lexical rules

The lexical complexity phenomena, present in the examined instructions are the following:

- Bulgarian is rich in neologisms, archaism, clichés, foreignisms, and phraseological combinations which could cause problems of understanding to non-native readers or people with a low reading age. Part of this phenomenon is due to the establishment of a particular style, created in the past 50 years. This formal bureaucratic style is rich in clichés, archaism and meaningless words.

- Bulgarian possesses a very strong system of prefixes and suffixes with many words having more than one root.

The lexical controlled language rules consist of general suggestions and lists of forbidden words and their recommended alternatives. Some lexical suggestions are given below:

1. We suggest that clichés and archaism be avoided. In some cases the problematic terms can be simply omitted, while in others alternative terms are suggested. Some examples are provided in Table 1.

<table>
<thead>
<tr>
<th>Table 1. List of lexical items and their replacements.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Formal word</strong></td>
</tr>
<tr>
<td>отчетливо (distinctnessly)</td>
</tr>
<tr>
<td>За евакуация използвайте стълбища или аварийните такива (For evacuation use the)</td>
</tr>
</tbody>
</table>

---

2Information about Plain English Campaign can be found at http://www.plainenglish.co.uk.
2. We suggest the use of words with only one root and no more than one prefix in order to keep the access to the meaning simple. If words with more than one root have to be used, they can only be specialized terms for which no alternative terms exist. Examples of such allowed multi-root terms are „пожарогасител” (fire extinguisher), „противогаз” (gas mask) and „земетресение” (earthquake).

3. We prohibit the use of metaphors and figures – to avoid ambiguity and because this is exactly the opposite of the idea of an emergency message, which should be straightforward to understand.

3.2. Morphological complexity phenomena and morphological rules

The morphological issues concern mainly the nominal and the verbal systems of Bulgarian [7]. An overview of both of them, together with controlled language suggestions are provided below.

3.2.1. Nominal system. The nominal system is quite complex. Nouns, adjectives and pronouns are inflected for grammatical gender, number, case (to a very limited extent) and definiteness. Adjectives and adverbial pronouns agree with nouns in number and gender. The analysis put into light a variety of issues, some of which are common for English, but others – like the high ambiguity of short pronouns – which are typical for Bulgarian (Ex. „ви” can be a possessive pronoun, but also a personal pronoun in dative or accusative case). Restrictions of the variety and ambiguity of morphological endings are addressed also in the controlled languages for Greek and for Polish. In Greek the suggestion is to restrict the use of pronouns in general. In Polish the homonymy of case forms in controlled language is avoided by using the passive voice or a composed verb.

For Bulgarian we suggest using only the full forms of the possessive pronouns and their derivations („ваш”, „ваша”, „ваше”, „ваш”). and not the short ones (such as „ви”) in order to avoid ambiguity („ви” is also personal pronoun in accusative or dative).

3.2.2. Verbal system. Verbs in Bulgarian vary in aspect, mood, tense, person, number, gender and voice. The verbal system exhibits a set of particularities with the most important ones being the aspect (a category which exists also in Greek and in Polish) and the lack of a real Infinitive form (the same as in Modern Greek).

In ancient Bulgarian and ancient Greek there is a real Infinitive which has disappeared and today the verb form in Present tense, first person, singular is conceived as an Infinitive and cannot be used in instructions (differently from Polish – where it can be said that there exists a real Infinitive form which permits the formation of constructions like “do not open”). The lack of a real Infinitive form imposes the use of the Imperative form of the verbs. The same case is in Modern Greek.

In Bulgarian parallel sets of perfective and imperfective pairs of verbs exist (e.g. “ночна - почвам”). The different semantic and syntactic behaviour of perfective and imperfective verbs, which is typical for languages such as Bulgarian, Greek, Serbian and Polish allows them to be used for different purposes.

Both perfective and imperfective verbs in Bulgarian, Greek and Polish can form Imperative in positive constructions. But in written positive instructions in Bulgarian it is better to use perfective verbs only to define the finite action. The imperfective verbs are the only ones which are grammatically allowed to be used for forming the negative Imperative form (the same case as in Modern Greek, Serbian and Polish). The opposition perfective verb/positive Imperative - imperfective verb/negative Imperative permits definitely to distinguish the two actions – “to do”/“not to do”. Allowing the use of negative Imperative has two advantages - on one hand the rich variety of equivalent negation forms in Bulgarian will be restricted (more than 4 forms exist), on the other, it will avoid the generation of long positive paraphrases (compare “не затваряйте вратата” (”do not close the door”) with “избягайте да затваряте вратата” (“avoid closing the door”).

Another crucial topic is the choice between the informal and the polite form of the Imperative. The problem is that the polite form is longer than the informal one, but tradition and the formality impose the use of the longer form.
The morphological guidelines concern the tense, voice, aspect, mood and the person of the verb, and the use of the non-personal verb forms.

1. The use of imperative in instructions is obligatory because of the lack of Infinitive.
2. The verb should be in second person, plural, active voice.
3. We suggest the use of perfective verb form for the positive imperative.
4. The use of negative is possible – because of the shortness and the definiteness.

3.2.3. Non-personal verb forms. In the current section we examine participles, adverbs, prepositions, particles and interjections. In Bulgarian there are different types of participles but we focused on the present participles - adverbial participle and adjectival participle (депричастия и страдатели причастия). These two kinds of participles could make sentences heavy and unclear (ex. “минавам” – “като минавате”, “слязах” – “който слиза”).

Our suggestions related to participles are the following:
1. We do not recommend the use of present participles.
2. We suggest using the conjugated verb instead of adverbial participle and adjectival participle, if it is necessary.

The adverbial system in Bulgarian is very rich because of the different types of adverb formation – they can be formed from nouns, adjectives, pronouns and verbs. The use of adverbs in instructions is sometimes absolutely necessary for clarity and good understanding. The system of the prepositions is also very rich because of the possibility to combine different prepositions in only one word (e.g. иззада, между, вътре, навсякъде, подпредел, подпредделством). We consider that particles and interjections are absolutely incompatible with instructions.

The rules derived from our observations are the following:
1. We allow the use of simple adverbs: a) formed from adjectives (“просто”,”бързо”); b) formed from pronouns (“стам”,”гова”,”когато”)
2. We allow only composed adverbs for location – отгоре”, ”отдолу”, ”отпред”, ”надясно”, relative indications such as ”отляво”, ”отясно” should be avoided;
3. We allow time adverbs, like ”сега”,”вечерта”.
4. We allow only simple prepositions.
5. Particles and interjections are not allowed (“наля”, ”ли”, ”надаля”, ”а”, ”ами”, ”дя”, ”я”, ”нама”, ”о”; ”ах”, ”мале”, ”олеле”, ”ех”, ”ох”)

6. The only particles allowed are these which are part of verbs – „не”, „да”.

3.3. Syntactic complexity phenomena and syntactic rules

The syntactic complexity issues in Bulgarian concern the rich variety of equivalent syntactic constructions and the ambiguity complications that arise from any attempt to solve this richness. More concrete details are listed below:

The conditions in our corpus are frequently in the form of long syntactically full sentences starting with „Ако” („Ако откриете пожар”). This could be replaced by a nominal phrase with preposition “при” for simplicity and clarity (ex. „При пожар” instead „Ако откриете пожар”).

We also recommend the use of „при” instead of the nominal construction „в случай на”. In some cases the replacement of a syntactically full sentence starting by „ако” is impossible (ex. „Ако сте навън”, „Ако сте вътре”). In these cases it is recommended to use the full construction.

Another problem is the position of the conditional construction: where to put the condition phrase with “If” – before or after the main clause (ex. “Като се по стълбите, ако има дим” or “Ако има дим, се качете по стълбите”. In Bulgarian the more logical position is the condition to be positioned after the main clause but it enters into conflict with the requirement of the Message controlled languages guidelines to put the condition in first. To solve this contradiction, there are two solutions:

Solution 1. To separate the two parts of the sentence and to use a full conditional construction:
„Ако има дим:
- качете се по стълбите”

Solution 2. To separate the two parts of the sentence and to replace the verb phrase with a noun one phrase:
„При дим:
- качете се по стълбите”

The problem with the inversion in the reflexive verb “качете се” – ”се качете” also exists. There is a different linguistic situation - it is possible to say “качете се по стълбите” and „тогава се качете”, but it’s impossible to say „се качете по стълбите” or „тогава качете се”. The proposed separation of the sentence in two parts could solve this problem.

There is no need to provide restrictions for the word order for Imperative, because in general the syntactic order of Imperative in Bulgarian is rigid. We do not go deeper in this paper.
The syntactic rules developed on the basis of the above observations define the allowed sentence types. 
1. Only simple phrases with one action are allowed. 
2. The instructions should follow the logical order of the actions to be performed. 
3. The conditions should be placed before the instructions. 
3.1. The condition should be introduced by a nominal construction with the preposition „при” + “noun”. 
3.2. In some cases the condition could be introduced by the full phrase „ако” only for clarity. 
3.3. The condition should be separated from the other part of the sentence.

3.3. Formatting rules

The formatting rules follow the guidelines provided in the Message Project, in order to test their transferability to other languages. The formatting rules define the structure of the text (titles, subtitles, subsections), the formatting (fonts, shrift, size), the compulsory and optional elements (titles, conditions, instructions, explanatory notes, lists) and their reciprocal order. They also provide guidelines about the punctuation signs to be used in specific situations.

4. Results

The set of lexical, morphological, syntactic and formatting rules has been applied in order to re-write 5 texts constituting the testing part of the corpus. An example of an original text containing instructions for actions to be undertaken in an emergency (left column) and its re-written version (right column) is given in Table 2.

Table 2. Example of a re-written instruction.

<table>
<thead>
<tr>
<th>Original</th>
<th>Re-written</th>
</tr>
</thead>
<tbody>
<tr>
<td>В ситуация на аварии и на извънредни ситуации, които могат да възникнат на място или по време на превоз, членовете на екипажа на превозното средство, са длъжни да предприемат следните мерки, от гледна точка на тяхната безопасност, и практическа възможност:</td>
<td>При аварии и извънредни ситуации: държете ръчната спираща, изключете двигателя, откачете акумулаторните батерии.</td>
</tr>
<tr>
<td>- задействайте ръчната спираща, изключете двигателя и откачете акумулаторните батерии,</td>
<td></td>
</tr>
</tbody>
</table>

The results indicate that the texts have been shortened by an average of 45%. Additionally, 3 human judges asked to select the simpler versions, have agreed that the re-written versions are the clearer ones.

5. Conclusions and future work

The paper aims to develop a human-oriented controlled language in order to assist the development of an early alert system in Bulgaria. Several complications related to the inherent complexity and high ambiguity of the Bulgarian language and their solutions have been discussed.

The paper also showed that negation has its own valid reasons to be kept in a controlled language for Bulgarian and that the formatting rules of Message project are transferable to new languages.

Future work will feature a close collaboration with the Bulgarian Ministry of Emergency Situations which is the main player in the currently developed early alert system in Bulgaria, as well as experiments to determine the allowed length of conditions, instructions and noun phrases in lists.

6. References


