Towards a Georgian Controlled Language in Crisis Management

Nino Amiridze¹, Anna Chutkerashvili¹,², Besik Dundua¹,³, and Irina Temnikova⁴

¹ Ivane Javakhishvili Tbilisi State University, Georgia
² Georgian Technical University, Georgia
³ International Black Sea University, Georgia
⁴ Sofia University, Bulgaria

Controlled natural languages (CNLs) are engineered languages that are based on natural language, but have their vocabulary, syntax, and/or semantics restricted [9]. The motivation is to have a language that, on one hand, looks as natural as possible and, on the other hand, is simple and unambiguous.

The application area of CNLs is quite broad. There are general-purpose CNLs (e.g., Attempto Common English [6]), whose vocabulary can be adapted to specific areas, and domain-specific ones, which reflect also the domain syntax. Due to space limitations, we just list some interesting practical applications of the domain-specific CNLs: writing technical documentation [3], mathematical authoring [5], crisis management [12, 13, 2], legal texts [4], interchanging business rules among organizations [11], and editing ontologies in OWL [10].

While English dominates the landscape of CNLs ([9] analyzes 100 English CNLs), controlled (sub)languages have been developed for many other languages (e.g., Bulgarian, French, Spanish, Chinese, Russian, German, Greek, Spanish, and Japanese). In this talk we present our project on Georgian controlled language in crisis management, which aims at establishing clear (and unambiguous) communication between different management specialists and the general population.

Georgian is a language from the Kartvelian family of languages, spoken mainly in Georgia, where it is an official language. Georgian has its own alphabet and long literary tradition, which goes back to 5th century AD. From the computational point of view, despite some recent and ongoing attempts, Georgian is still pretty under-resourced (e.g., [1, 8]).

One of the application areas of controlled natural languages is crisis management. During the last decades several big disasters happened in Georgia which had substantial impact on people’s lives. In 2017, 100 hectares of forests were burned in Borjomi Gorge. In 2015, the Vere river was flooded, which caused severe damage on the city’s infrastructure and resulted in at least 19 fatal casualties. In 1991, an earthquake occurred in the province of Racha, which killed more than 270 people and left more than 100,000 inhabitants homeless. Such big disasters require clear communication with large number of people in a stressful situation. Unfortunately, until now, the country does not have a central emergency alarm system, but since 2018 there is a discussion in the government to implement it.

Our project includes: collection of a small corpus of emergency instructions in Georgian, linguistic analysis of the ambiguities in the texts, development of CNL grammar rules (some
of which map to first-order logic) and define the allowed syntactic structures, word order, and lexicon; and evaluation of the resulting CNL.

In this talk we present some of the work done. We present some ambiguity cases and propose alternative clearer solutions. We analyze which of the existing ambiguity cases for other languages appear also in Georgian, and which are typical only for it. Being a complex language not closely related to any well-resourced language, developing a CNL for Georgian is an additional challenge. In fact, Georgian has one of the most complex morphologies: e.g., the verb form is very complex – it has agglutinative and inflection morphemes, and many categories (person, number, tense, mood, voice, etc.) which complicates parsing.

Some of the most frequent types of complexities and ambiguities in emergency texts are: writing too long sentences, not following lexical consistency, using syntactically or lexically ambiguous modifiers [7], using more archaic or complex forms or highly specialized terms, etc.

The examples below show some of the ambiguity cases in Georgian emergency instructions, that we identified in real texts, along with our proposed solutions.

(1) Syntactic ambiguity, which may prevent from correctly executing the instruction:

\begin{verbatim}
(1) Syntactic ambiguity, which may prevent from correctly executing the instruction:
daketet kar+eb-eb-i, panjr-eb-i da napral-eb-i
you.PL.close.them door+PL-PL-NOM window-PL-NOM and opening-PL-NOM
amokolet svel-i pirsaxoc-it.
you.PL.block.them wet-NOM towel-INST

a. ‘Close the doors and windows and block the openings with wet cloths.’
b. ‘Close the doors and block the windows and openings with wet cloths.’
\end{verbatim}

The form *k.ar+eb-eb-i* in (1) is a non-standard, colloquial form referring to the plural referent ‘doors’. The instruction manual from which we cited the example avoided using the standard Georgian plural form *kar-eb-i* (2a), as the latter is ambiguous – in colloquial Georgian it refers to the singular referent ‘door’ (2b).

(2) Lexical ambiguity:

\begin{verbatim}
(2) Lexical ambiguity:
  a. Standard Georgian
      kar-eb-i
door-PL-NOM
      ‘doors’
  b. Colloquial Georgian
      kar+eb-i
door+PL-NOM
      ‘a door’
\end{verbatim}

To avoid both the lexical ambiguity (which the use of *karebi* might cause (cf. (2a) vs. (2b))) and the use of the non-standard *kar+eb-eb-i* (as in (1)), we would prefer to employ the quantifier *qvela* ‘all’ (3):
Our talk will present more intriguing cases, taken from real documents, along with the CNL pattern rules to avoid such ambiguities. Our work is an useful example for many other under-resourced languages.

Acknowledgments. This research has been partially supported by the Shota Rustaveli National Science Foundation under the grant YS-18-1480.

References