Towards the Use of Controlled Natural Languages in Hazard Analysis and Risk Assessment

Paul Chomicz[1] Armin Müller-Lerwe[2] Götz-Philipp Wegner[2] Rainer Busch[2]
Stefan Kowalewski[1]

Abstract: New safety-critical and software-controlled systems of automobiles have to be developed according to the functional safety standard ISO 26262. A hazard analysis and risk assessment has to be performed for such systems. The sub-activities of this analysis technique are defined by the standard, but informative definitions leave room for subjective variation, and documentation details are left to the car manufacturer. Usually, natural languages are used for the documentation, which are powerful and expressive but also complex and ambiguous. We propose the usage of controlled natural languages for the documentation of the results of the hazard analysis and risk assessment. In a first step, we developed a controlled natural language for the description of the hazardous events. The language reduces ambiguity and improves the consistency across hazard analyses and risk assessments.

Keywords: Controlled Natural Language, Hazard Analysis and Risk Assessment, Functional Safety, ISO 26262, Hazardous Event

1 Introduction

In the automotive industry, new safety-critical functions that are realized by software-controlled electric or electronic systems have to be developed in accordance with the functional safety standard ISO 26262 [IS11]. Its safety lifecycle encompasses principal safety activities during the concept phase, product development, and after start of production.

The hazard analysis and risk assessment (HARA) is a safety activity which is performed during the concept phase. The objective is to identify and categorize the potential hazards of functions and to derive safety goals for the prevention or mitigation of these hazards. The hazard analysis and risk assessment comprises three steps. The first step is the situation analysis and hazard identification. Potential unintended behaviors have to be identified that could lead to a hazard within a specific situation (hazardous event). Afterwards, the risk classification of the hazardous events takes place. The risk of each hazardous event is classified by determining the severity (S), the probability of exposure (E), and the controllability (C). According to these three parameters, the required automotive safety integrity level (ASIL) is assigned in the last step. The ASIL specifies the level of risk reduction for achieving an acceptable residual risk with ASIL D representing the highest and ASIL A the lowest level [IS11].

[1] RWTH Aachen University, Lehrstuhl Informatik 11 – Embedded Software, Ahornstraße 55, 52074 Aachen, {chomicz, kowalewski}@embedded.rwth-aachen.de
[2] Ford Research & Innovation Center Aachen, Süsterfeldstraße 200, 52072 Aachen, {amuell12, gwegner2, rbusch1}@ford.com

The ISO 26262 standard describes which sub-activities are part of the hazard analysis and risk assessment, but it does not describe, for example, how to record the unintended behaviors or how to determine one of the risk parameters. The standard defines severity, exposure, and controllability in a qualitative way that leaves room for subjective interpretation. Due to the fact that usually mutiple new functions use the same actuators, malfunctions could often cause similar hazards. Since new functions and systems are developed by different teams, it is a challenge to assure consistency of the risk classifications between the hazard analyses and risk assessments developed for different vehicle functions. Inconsistency might lead to different levels of safety measures for similar hazardous events.

In analysis techniques like hazard analysis and risk assessment, a natural language is usually used for the documentation. On the one hand, natural languages are powerful and expressive, but on the other hand, they are complex and ambiguous. Same or similar hazardous events and rationales for the classification are often described using different wordings and phrases. This makes it difficult to check for consistency especially across HARAs that are developed by different teams. In order to approach these problems, controlled natural languages are a promising way.

A controlled natural language (CNL) is a subset of a natural language [Ku14]. CNLs are obtained by restricting the vocabulary or the grammar. These restrictions aim to increase terminological consistency and to reduce ambiguity and complexity. Numerous controlled natural languages have been developed for various domains, e.g. Airbus Warning Language [SBC03], Attempto Controlled English (ACE) [FKK08], or Bio-Query-CNL [EY09].

The controlled natural languages for the hazard analysis and risk assessment should allow the description of hazardous events and the rationales of the risk parameters in such a way that it supports the engineers in the development of HARAs, e.g. by a more efficient search for existing ratings of similar hazardous events. It should also reduce the possibility of formulating similar hazardous events or rationales with different wordings and phrases, and additionally, the language should enable or simplify an automatic consistency check between different HARAs.

The remainder of this paper is structured as follows. The next section describes a particular controlled natural language in detail, which was developed and put into operation at Ford Motor Company. Then, the formalization of the hazardous events is explained including the process to accomplish it. The fourth section gives details on the evaluation of the newly created language and first experiences about a productive usage. The last section presents an outlook on future work.

2 Related Work

There are numerous controlled natural languages in various domains [Ku14]. In general, CNLs can be divided into general-purpose languages and domain-purpose languages [Sc10]. A general-purpose language has not been designed for a specific scenario or application domain. An example of such a language is Attempto Controlled English (ACE)

[FKK08]. On the contrary, domain-purpose languages have been developed for a particular application area. The Airbus Warning Language [SBC03] and the Bio-Query-CNL [EY09] are such languages.

To the best of our knowledge, no controlled natural language has been developed specifically for hazard analysis and risk assessment. Some work was done in the area of applying CNLs to (safety) requirements engineering in the automotive domain [PMP11, HMM11, FH14]. In the following, a controlled natural language will be introduced that was developed at the Ford Motor Company. The structure of the language is similar to ours, and especially, the experiences that were made during the development and the productive usage at Ford are of great value.

The Standard Language (SLANG) is a controlled natural language that was developed to write process build instructions for a vehicle [Ry02]. Before using this controlled language, build instructions were written in a natural language causing problems such as ambiguity and inconsistency. Furthermore, Ford's vehicle assembly plants are spread all over the world, and threrefore, several different natural languages were used for the process sheets.

In addition to the language, the Direct Labor Management System (DLMS) was developed to address these problems [O'89]. The tool assists the vehicle assembly process planning. As input, process sheets written in the Standard Language are taken to produce detailed work tasks for each step of the assembly process. Before releasing these tasks to the assembly plants, they are translated into the corresponding language automatically by the system. The usage of the controlled natural language enables or simplifies the automatic machine translation and the precise determination of all work tasks, since the language is constructed in such a way that it can be processed by the system.

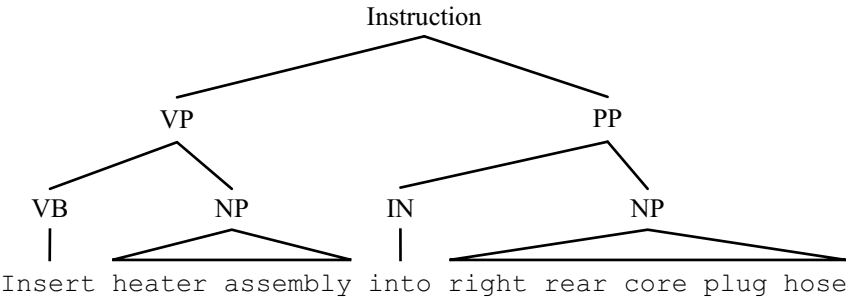

Fig. 1: Instruction written in SLANG [Ry05]

Instructions written in the Standard Language are in an imperative form including a verb phrase (VP), which contains a noun phrase (NP) that serves as the object of the verb. The level of detail can be increased by adding adverbs, adjuncts, or prepositional phrases (PP). The verb is the key word of every instruction, and every verb describes a particular, precisely defined action. The number of verbs in the vocabulary is limited, and certain prepositions have also a specific meaning that will be handled in a pre-determined manner by the system. Every part of the text that is delimited by brackets is ignored by the system

and does not have to conform to the Standard Language. Thus, it is possible to add comments or remarks to the build instructions. Figure 1 shows an example instruction written in SLANG.

The vocabulary of the language is limited to a certain set of words. Since, the vehicle assembly process is very dynamic, it is necessary to extend the vocabulary along with new vehicles and assembly plants that are added to the system. Therefore, process engineers have to request changes which need to be approved by an internal Ford systems organization before they will be added into the language and the system.

As the system simply flags any errors, the process engineers need an enhanced knowledge about the language to be able to fix the errors or to write correct build instructions in the first place. Along with the introduction of the controlled natural language for the productive usage, the process engineers were trained to write the process sheets in SLANG. During the initial implementation, the users resisted to use the language until they were trained and learned how to use it effectively. Another problem that was encountered was the misuse of the commentary function. This feature was used by users to bypass the process of adding new terms into the language [Ry06].

The Standard Language is a controlled natural language that was never designed to produce correct grammatical sentences with respect to the English language [Ry02]. The goals of the language were to develop consistent and precisely defined means of communicating and to enable or to simplify machine translation. Therefore, the process sheets get more precise and simpler in terms of automatic processing, but the language is less expressive and loses a part of its naturalness compared to the English language. As a consequence, instructions become less readable and less understandable for humans, especially for untrained users. However, the restrictions on the vocabulary and the grammar improve the quality of the translations. To counteract the negative effects of the newly created language, effort needs to be made to train the process engineers.

SLANG is a domain-purpose language. It is too domain-specific so that it could be reused for our purpose. Whereas, the usage of a general-purpose language would be in general possible, but it would have also some drawbacks. Such a language is not optimized to our domain-specific application. Certainly, the vocabulary would have to be adapted to the automotive domain. Furthermore, the structure of the language might not be suitable with respect to how the descriptions were written before by the safety engineers. Therefore, we decided to develope an own language that is close to existing hazardous event descriptions.

3 Controlled Natural Language for Hazardous Event Descriptions

In this section, the process of the formalization of the hazardous event descriptions and the formalization itself are described. For this purpose, we have analyzed existing hazard analysis and risk assessment documents provided by the Ford Motor Company. Based on this analysis, the controlled natural language was created. Furthermore, the translation of hazardous event descriptions that do not conform to the CNL is explained by means of two examples.

3.1 Analysis Process

To achieve a formalization of the hazardous event description, existing hazard analyses and risk assessments provided by the Ford Motor Company were analyzed. Our approach is bottom-up and iterative. In the first iteration, nine HARA documents have been analyzed. The documents describe the hazard analysis and risk assessment of, among others, an emergency braking system and an electronic controlled differential. The HARAs were performed by different teams and in different countries. However, the English language was used in all cases for the documentation.

From the provided documents, 208 different hazardous event descriptions were extracted, and the structure of the descriptions and the used wording were analyzed to create the controlled language. Table 1 contains an exemplary set of such descriptions that represents how hazardous events are currently described.[3] A hazardous event description consists of at least one hazardous event, which might be caused by another event. Thus, it is possible to construct causal chains.

The structure of the hazardous event descriptions can be divided into two categories. The first category contains descriptions that were written in a bullet-point manner. The hazardous events in the second category were formulated using full sentences. The descriptions 1 and 4 of Table 1 are written in a bullet-point manner, and the other two are part of the second category. Overall, 141 descriptions belong to the bullet-point manner category (67.8 %) and 45 descriptions are part of the full sentence category (21.6 %). 22 hazardous events were formulated using both full sentences and bullet-point descriptions (10.6 %). The categorization was performed manually.

No.	*Hazardous Event Description*
1	`Fire outside passenger compartment.`
2	`The driver is not alerted to a credible threat.`
3	`The system is active at high speed and may not detect objects in relevant distance (due to sensor performance).`
4	`Unintended and unlimited system activation leading to loss of vehicle steerability due to blocked wheels without ABS.`

Tab. 1: Exemplary set of hazardous event descriptions

The intermediate formalization for the first set of HARA documents was reviewed in a second iteration, where seven different documents have been analyzed. Again, the provided data fulfilled the same properties as the first one, e.g. the HARAs were performed by different teams. The data set contains the documentation of the hazard analysis and risk assessment for an electronic clutch and a park assist.

[3] The examples are slightly modified to avoid the disclosure of proprietary information about the analyzed systems.

93 additional hazardous events that are different compared to the first set were extracted from this set. The structure of the descriptions is the same as in the first set, and the used wording is similar apart from system-related words. 76 hazardous events are written in a bullet-point manner (81.7 %) and 12 descriptions are formulated using full sentences (12.9 %). Again, a small portion of the descriptions is formulated using both full sentences and bullet-point descriptions (5.4 %).

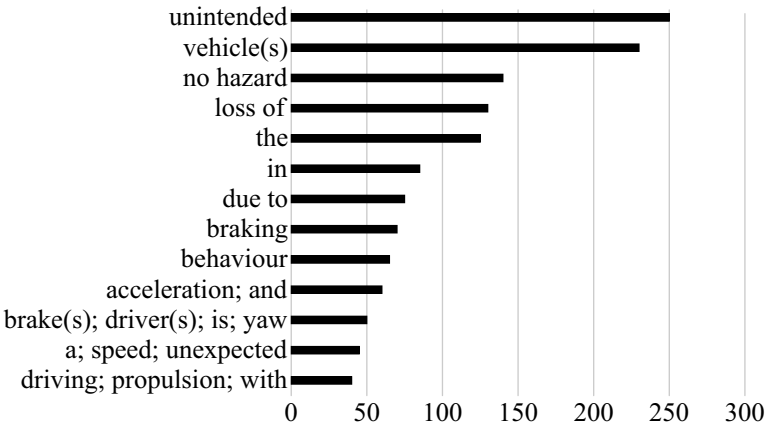

Fig. 2: Most frequently used words and phrases in hazardous event descriptions

Figure 2 depicts a frequency count of the words and phrases that were mostly used for the description of hazardous events. Most of the words are conjunctions, prepositions, adjectives, and nouns. Verbs were rarely used for the descriptions since most of them were written in a bullet-point manner and not as full sentences.

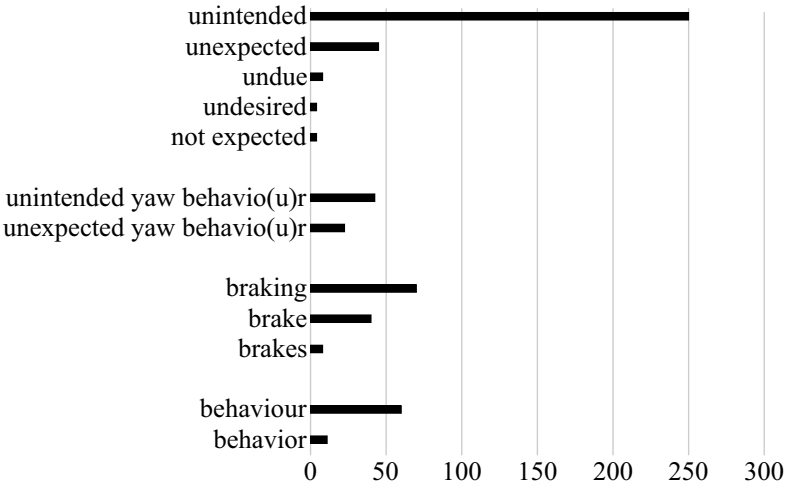

Fig. 3: Synonyms and similar words and phrases in hazardous event descriptions

In addition to this frequency count, an exemplary set of synonyms and words with the same or similar meaning but different spelling, e.g. American and British English, are

depicted in Figure 3. The first group contains words and phrases that are synonyms. The meaning might be slightly different between these words and phrases, but in our context, the differences can be ignored. An example is contained in the second group. This group consists of two different hazardous event descriptions that share nearly the same meaning. Another syntactically different description with the same semantic that is not contained in the analyzed HARAs is “unintended steering input”.

The third group contains words with the same word stem. The hazardous event “unintended braking” and “unintended brake activation” describe the same event using different wordings based on a word with the same word stem. The last group shows an example of the difference between American English and British English. Same words with slightly different spelling are contained in either language. This is an additional fact that has to be considered during the creation of a CNL.

3.2 Formalization

In this subsection, the results of the formalization of the hazardous event descriptions are presented. The restrictions for the developed controlled natural language are made on both the structure of the descriptions, so the grammar of the language, and the vocabulary. The language was developed in a bottom-up approach, and therefore, it is closely related to the provided data.

The grammar only allows to write the hazardous event descriptions in a bullet-point manner. Noun phrases are used to describe an event or a characteristic of a system. The phrase has a noun as its headword and can contain additional adnominals, like adverbs, adjectives, or noun adjuncts. The usage of pronouns and clauses in these noun phrases is prohibited. Certain prepositions and conjunctions are used to connect single hazardous events to build up more complex descriptions and to be able to create causal chains. The prepositions are divided into semantical categories, e.g. the prepositions “between” and “in front of” indicate position information or “by” and “to” indicates the point of view.

The controlled natural language does not provide the possibility to use full sentences for the descriptions. Therefore, verbs are not needed and not part of the language. This restriction further reduces the complexity of the language in comparison to the complete English language. Without verbs, the distinction between active and passive voice does not have to be considered, and further on, grammatical tenses are also omitted.

The first example from Table 1 conforms already to the grammar of the controlled language. The description consists of a noun phrase with a single noun and a prepositional phrase giving additional position information. Figure 4 depicts the example along with the classification of the part of speeches.

The last example from Table 1 is almost a correct description according to the controlled language. The description starts with a noun phrase consisting of additional adjectives followed by two causal phrases (CP). Causal phrases are phrases that start with a causal

linking phrase (CLP) followed by a noun phrase. Such phrases enable the user to formulate causal relationships.

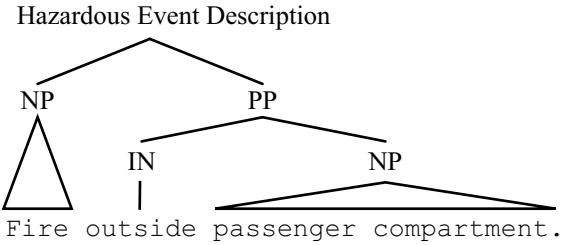

Fig. 4: Correct example with regard to the CNL

In this case, two different causal linking phrases have been used describing slightly different causal relationships. “A leading to B” expresses that an event A is the cause of an event B, whereas “A due to B” expresses that an event B is the cause of an event A. Both can be used interchangeably. However, describing a situation with more than one cause and using both expressions together might lead to a lack of causal relationship information.

In the example depicted in Figure 5, the structure of the description is “A leading to B due to C”. From this formulation, it can be interpreted that the events A and C are the causes of event B but nothing is said about the relationship between A and C. Considering the events in more detail, the event A (“unintended and unlimited system activation”) happens before the event C (“blocked wheels without ABS”). To avoid such a lack of information and to reduce misunderstandings, we restrict the controlled natural language to use only the causal linking phrase “due to”.

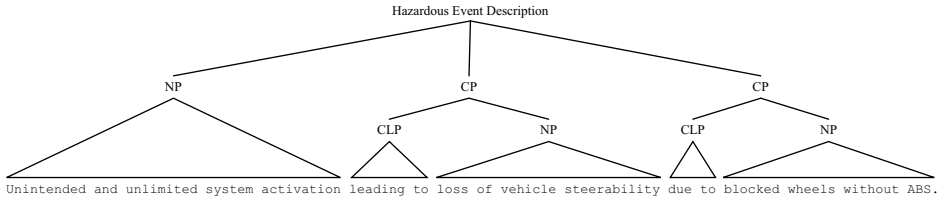

Fig. 5: Almost correct example with regard to the CNL

Since the usage of such a new controlled natural language might be difficult at the beginning, a commentary functionality is introduced to enable the user to write parts of the description using the complete English language. The users should be able to write the first ideas of the description or parts of it in the language they know and afterwards to translate it into the required form. This functionality is rather intended for the productive usage of the language than being a part of the language itself. Still it can be used to provide additional information, which is related to the hazardous event but is not part of its description.

The vocabulary of the CNL is restricted to the words that have been used in the provided documents and which were not removed during the formalization. In this context, words have been identified that share the same semantics. For example, the words “unintended”,

"unexpected", "unwanted", and "undesired" are semantically equivalent in our context as already mentioned above. Only one of the words ("unintended") is contained in our vocabulary, and the other synonyms are prohibited.

3.3 Translation

Hazardous event descriptions that were formulated using full sentences can be translated into bullet-point manner descriptions which are semantically equal and conform to the developed controlled natural language. In this subsection, the hazardous events 2 and 3 of Table 1 will be exemplarily translated into our language.

The general translation procedure is as follows. First, the determination of the part of speech of every word has to be performed [Br00]. Based on the given parts of speech, the translation is performed by removing words with certain parts of speech or transforming words into related words with a different part of speech. Some of the transformations will be explained by the two following examples.

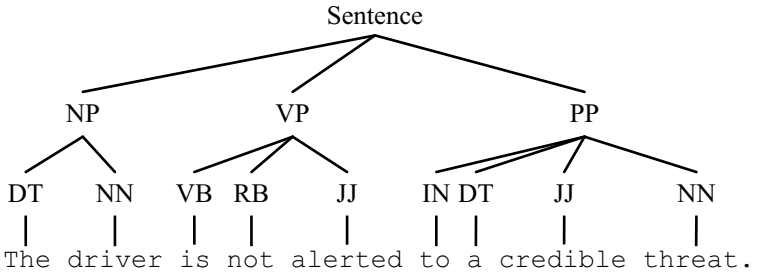

Fig. 6: Incorrect hazardous event description as a full sentence

The first example is shown in Figure 6 along with the classification of the parts of speech. The first part of the description is a simple sentence with a subject, verb, and adjective. In such a case, the "to be" verb can be simply removed and the order of noun and adjective switched resulting in "not alerted driver". If the adjective is modified by an adverb, then this will be moved along with the adjective just like in this example. The prepositional phrase consisting of a preposition and a noun phrase can remain unchanged. As a result, the hazardous event description "not alerted driver to a credible threat." conforms to the controlled natural language.

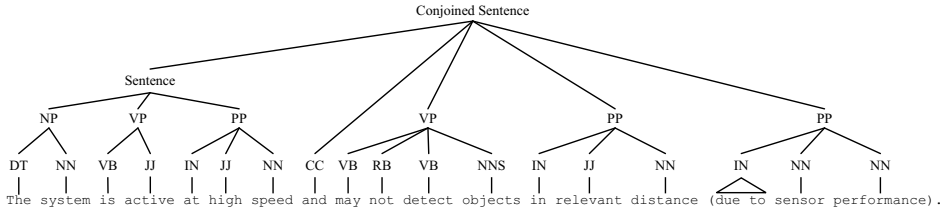

Fig. 7: Incorrect hazardous event description as a conjoined sentence

Figure 7 displays a more complex example. The first part of the conjoined sentence can be translated accordingly to the first example to “active system at high speed”. The second part contains an object after the verb followed by a prepositional phrase. The verb needs to be translated into an adjective that describes the effect on the object. The object is used as the noun phrase resulting in a passive description. This leads to “undetected objects”. The prepositional phrase can be simply taken as it is and added at the end. The last part of the description is written in brackets. The intention behind the usage of the brackets is not defined and gives room for various interpretations. In this case, it might be additional information or an assumption for the root of the hazard. A complete translation of this example might be “active system at high speed and undetected objects in relevant distance by the system due to sensor performance.”.

4 Evaluation

In total, 301 hazardous event descriptions were extracted from 16 different hazard analysis and risk assessment documents. 217 descriptions were already written in a bullet-point manner (72.1 %), and 57 hazardous events were described in full sentences (18.9 %). The remaining 27 descriptions were written in a mixed version (9.0 %).

The newly created controlled natural language for the hazardous event descriptions has been evaluated against the provided data. 156 out of the 217 descriptions that were written in a bullet-point manner are already in line with the CNL (71.9 %). Another 48 hazardous events could be translated into a correct form by replacing a synonym with the correct word that is part of the vocabulary (22.1 %). The other descriptions, including those that were written in full sentences or a mixed version, could be all translated into semantically equivalent hazardous event descriptions that conform to the language, as exemplarily shown in subsection 3.3. These results show that the language is closely related to the provided data but still expressive enough to describe every hazardous event that has arised in the considered HARAs.

Furthermore, the new language was prototypically applied in hazard analyses and risk assessments for new systems to make first experiences in a productive usage. Different engineers use the CNL during the development of three new HARAs within the domains steering, fuel cell, and powertrain. By extending the vocabulary with domain-specific terms, it was feasible to describe all hazardous events according to the CNL. It turned out that using the language leads to more consistent descriptions within the HARAs compared to existing HARAs.

On the other hand, it is difficult to write valid descriptions according to the CNL without good knowledge about the vocabulary and the grammar. Therefore, instantaneous support while entering data seems to be essential to enable engineers to use the controlled language effectively when performing hazard analyses and risk assessments.

As a last point, two examples are presented to show in which way the controlled natural language is able to avoid inconsistency. The vocabulary does not contain synonyms. Therefore, it reduces the possibility to write semantically equal descriptions syntactically

differently. For example, the two descriptions “unintended acceleration.” and “unwanted speed-up.” have the same meaning. The words “unintended” and “acceleration” are part of the vocabulary, but “unwanted” and “speed-up” are not, since these are synonyms of the other two words.

The second example concerns the structure of the descriptions. Using the English language, the two descriptions “vehicle may pull towards the opposite lane or the side of the road due to understeering behavior.” and “lane departure due to understeering.” with the same meaning were permitted, but only the second one conforms to the controlled natural language. The restrictions on the vocabulary and on the structure of the descriptions are intended to unify the descriptions and to reduce ambiguity and complexity.

5 Conclusion and Outlook

The formalization of the hazardous event descriptions is the first step towards the utilization of controlled natural languages for the hazard analysis and risk assessment according to ISO 26262. The controlled natural language defines a restricted common structure for the descriptions along with a limited vocabulary. Therefore, the complexity and ambiguity are reduced in the documentation of the HARA resulting in less inconsistency. Furthermore, the common structure simplifies the search for existing same or similar hazardous event descriptions.

The evaluation shows that all existing hazardous event descriptions of the provided HARA documents can be translated into the controlled language. Furthermore, a large portion of the descriptions was already compliant with the language. The CNL was prototypically applied to three newly created HARAs at the Ford Research & Innovation Center Aachen. It turned out that the language was applicable after extending the vocabulary with domain-specific terms, and all hazardous events could be described in that language. Currently, the vocabulary is restricted to the used words in the analyzed HARAs, and it needs to be extended beyond the scope of the provided documents.

Besides the description of the hazardous events, the rationales for the ratings of the parameters severity, exposure, and controllability are an essential part of the documentation of the hazard analysis and risk assessment. Therefore, controlled natural languages for the three rationales shall be provided to complete the set of languages. After the completion, all the languages shall be implemented in a prototype tool to further examine the usage of such languages for the HARA. Based on the prototype tool, a case study will be performed to gather more user experiences that shall help to improve the languages and their usage.

References

[Br00] Brill, Eric: Part-of-Speech Tagging. Handbook of Natural Language Processing, pp. 403–414, 2000.

[EY09] Erdem, Esra; Yeniterzi, Reyyan: Transforming Controlled Natural Language Biomedical Queries into Answer Set Programs. In: Proceedings of the Workshop on Current Trends

in Biomedical Natural Language Processing. Association for Computational Linguistics, pp. 117–124, 2009.

[FH14] Fockel, Markus; Holtmann, Jörg: A Requirements Engineering Methodology Combining Models and Controlled Natural Language. In: Model-Driven Requirements Engineering Workshop (MoDRE), 2014 IEEE 4th International. IEEE, pp. 67–76, 2014.

[FKK08] Fuchs, Norbert E.; Kaljurand, Kaarel; Kuhn, Tobias: Attempto Controlled English for Knowledge Representation. In: Reasoning Web, pp. 104–124. Springer, 2008.

[HMM11] Holtmann, Jörg; Meyer, Jan; Meyer, Matthias: A Seamless Model-Based Development Process for Automotive Systems. In: Software Engineering (Workshops). pp. 79–88, 2011.

[IS11] ISO: , ISO 26262-3: Road Vehicles – Functional Safety – Part 3: Concept Phase, 2011.

[Ku14] Kuhn, Tobias: A Survey and Classification of Controlled Natural Languages. Computational Linguistics, 40(1):121–170, 2014.

[O'89] O'Brien, John; Brice, Henry; Hatfield, Scott; Johnson, Wayne P; Woodhead, Richard: The Ford Motor Company Direct Labor Management System. In: Innovative Applications of Artificial Intelligence. volume 1, 1989.

[PMP11] Post, Amalinda; Menzel, Igor; Podelski, Andreas: Applying Restricted English Grammar on Automotive Requirements – Does It Work? A Case Study. In: International Working Conference on Requirements Engineering: Foundation for Software Quality. Springer, pp. 166–180, 2011.

[Ry02] Rychtyckyj, Nestor: An Assessment of Machine Translation for Vehicle Assembly Process Planning at Ford Motor Company. In: Conference of the Association for Machine Translation in the Americas. Springer, pp. 207–215, 2002.

[Ry05] Rychtyckyj, Nestor: Ergonomics Analysis for Vehicle Assembly Using Artificial Intelligence. AI Magazine, 26(3):41, 2005.

[Ry06] Rychtyckyj, Nestor: Standard Language at Ford Motor Company: A Case Study in Controlled Language Development and Deployment. Cambridge, Massachussets, 2006.

[SBC03] Spaggiari, Laurent; Beaujard, Florence; Cannesson, Emmanuelle: A Controlled Language at Airbus. Proceedings of EAMT-CLAW03, pp. 151–159, 2003.

[Sc10] Schwitter, Rolf: Controlled Natural Languages for Knowledge Representation. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, pp. 1113–1121, 2010.