Facilitating the application of Controlled Natural Language (CNL) to standardize communication in logistics and supply chain management

Joel Cedric Lengeling
Chair of Logistics
Technical University of Berlin
lengeling@logistik.tu-berlin.de
alternatively joel.cedric@lengeling.eu

Abstract

Communication protocols allow to standardize communication. They are typically implemented to standardize the exchange of messages in the area of information systems. Nevertheless, by applying Controlled Natural Language (CNL), it is possible to implement communication protocols that allow both; participation by humans, while still enabling information systems to accurately and efficiently process the very same messages. Thus allowing computers and humans to communicate in unison. Here an artifact that allows applying formal CNLs for communication in the domain of logistics is presented.

1 Introduction

It is a truism: communication plays an important role in logistics. Delays, production schedules, missing spare parts ... and the information thereof has to be communicated. Although communication plays such an important role, it is surprisingly neither always standardized, nor is it - partly result of that missing standardization - always automated and misunderstandings are commonplace. This contribution aims to mitigate that situation by promoting the application of Controlled Natural Languages (CNLs) in logistics communication introducing an artifact.

To assess this artifact some observations on the logistics industry and some observation on, more or less, typical communication situations in that industry are presented. Part of those observations are based on interviews conducted with some arbitrarily chosen experts in the field. In addition some casual conversations with industry veterans do influence that picture: logistics is an highly competitive market, profit margins are low, and the fear for competition is commonplace. Logistics business arrangements are often only short term and the small organizations forming a large part of that market are rarely big enough to allow for an "automation" department. The average employee is typically non-academic and the environment is multi-lingual and multi-cultural. In addition the following observations on communication in the domain of logistics have been made: communication may either be scheduled or non-scheduled, it may often be granular, it might often reoccur regularly, and the variance between occurrences may often be rather low.

Some reasons identified by the experts why communication is typically not yet fully automated have been: 1) the incompatibility of information systems, 2) the sizes of participating organizations, 3) the mentality towards information transparency, 4) the mentality towards digitalization, and 5) the cost of implementation. As a result opportunities for digitalization are missed.

Standardizing communication in the domain of logistics applying CNLs may potentially improve that situation. Kuhn mentions that there is no generally agreed-upon definition of CNL and describes the insight that "CNLs can be conceptually located somewhere in the gray area between natural languages on the one end and formal languages on the other"(cf. Kuhn, 2014). Here we work with CNLs that are on the formal language end. ISO/IEC/IEEE 24765:2017(E) defines a formal language rather conversational as "language whose rules are explicitly established prior to its use"(24765, 2017 p. 188). Mateescu defined formal languages less conversational applying Σ and Σ^* (cf. Mateescu and Salomaa, 1997 pp. 10-11). The alphabet Σ is a finite nonempty set of which the elements are called letters or symbols and Σ^* is a set of all words or strings consisting of zero or more letters of Σ. Subsets, finite or infinite, of Σ^* are referred to as formal languages over Σ. We work with CNLs that are formal languages that apply words or strings from one specific natural language - the base language - in such a way, that the essence of texts written in that CNL may be understood by the average employee that understands the base language from which the words or strings originate. In formal language theory a grammar of a language is a mechanism that allows the production of sets of strings in that language (cf. Harrison, 1978 p. 13). Essentially, formal languages are described by their syntax. The semantics of a formal language is, at least in computer science textbooks, rarely discussed. The semantics of a formal CNL is here taken from the base language that provided the words or the strings to that CNL. Extended Backus–Naur form (EBNF), an extension of Backus–Naur form (BNF), allows to express the grammar of the formal languages in mind. *A Restricted English for Constructing Ontologies (RECON)* is actually an example of a CNL that has been expressed by BNF (Barkmeyer and Mattas, 2012). Standardizing communication applying a CNL of which the grammar is e.g. expressed applying EBNF would allow the automation of that communication, while at the same time allowing "participation" of a non-academic workforce speaking the base language of the CNL. Nevertheless, as of today, standardizing communication applying CNLs is rarely discussed in both the logistics scientific community as well as in the logistics industry. Here an attempt resulting in an artifact that may lead to a more often application of CNLs in logistics is presented. This attempt is conducted from the viewpoint of a software engineer. The goal is to present an artifact that allows simple and flexible application while being rather maintainable and independent of other systems.

2 A limited literature review

A limited search was conducted in three research databases: A) *Business Source Complete (via EBSCO Host)*, B) *Web of Science Core Collection*, and C) *IEEE Xplore Digital Library* during the summer of 2021. The following search string has been applied: *("controlled natural language" OR "cnl" OR "domain specific language" OR "dsl" OR "formalized language" OR "processable language") AND ("logistics" OR "scm" OR "supply chain" OR "operations management")*. By taking that approach 58 results in A), 5 results in B), and 32 results in C) have been identified. After filtering for duplicates, non-scientific publications, and publications that have been considered off-topic, 19 publications remained. Of those 19 publications, 17 mentioned DSL, 1 mentioned CNL, and 1 mentioned both, DSL and CNL. Following Deursen a DSL is "a programming language or executable specification language that offers, through appropriate notations and abstractions, expressive power focused on, and usually restricted to, a particular problem domain"(Van Deursen et al., 2000). From the viewpoint taken here 1) a CNL may be also classified as a DSL, if the CNL is a formal language, and 2) a DSL may be also classified as CNL, if the DSL incorporates natural language and is expressive enough to allow usage by anybody fluent in that natural language. Thus some languages may be classified as both, CNL and DSL. Of the 17 contributions mentioning DSLs, 1 contribution introduces a DSL which at least from our perspective may also be classified as CNL. That language is called *Logistics Task Language* and it allows "to describe intra-logistics material flow processes" (Detzner et al., 2019). The identified contribution that mentions CNL, specifically does mention CNL as part of a system that supports manual assembly planning (Manns et al., 2018). Nevertheless, from the perspective taken here, both do not introduce a generalized system for logistics communication. The contribution which mentions both DSL and CNL does this, by discussing a use case related to logistics in the context of ProjectIT, an approach and tool for requirements engineering (Da Silva et al., 2007). Nevertheless, from the perspective taken here, this approach does not seem suitable to generalize communication in logistics as described before. The earlier mentioned language RECON was not found by applying the aforementioned search strings showing the limit reach of the literature review. Nevertheless RECON may still be applicable in logistics. At the end of this short overview, it might be important to point out, that there may be more CNLs applied in the actual field that are not subject to scientific publications. One example coming to mind are Pick-By-Voice systems that are most likely implemented applying CNLs, but nevertheless no publication discussing how to implement such a system was found within the limited search here.

3 Applying CNL to standardize communication in logistics

To simplify the task at hand this work builds on a flexible and comprehensible model of a *communication protocol* out of the domain of computer science presented by Holzmann. The advantage of that model is, that it allows a rather flexible approach to implement standardization without going into as much detail as many other technical standards in the field do; e.g. for application in logistics it suffices to specify that messages are transmitted by (e-)mail including an (e-)mail address instead of specifying the communication down to the bit level over the wire. Another benefit of that approach is it's ability to just describe the exchange of one message as part of one protocol. A complete message exchange may be described applying multiple instances of one protocol. Following Holzmann a protocol is the sum of all rules, all formats, and all procedures that have been agreed upon, between at least two computers in order to communicate (cf. Holzmann, 1991 pp. 19-21). Here, that definition is slightly extended, to also cover communication between either two humans or an human and a computer. A message that contains the content of the communication may be created by applying the rules, formats, and procedures specified as part of that protocol. Following Holzmann a protocol specification consists out of five distinct parts: 1) the *service* to be provided by the protocol, 2) the *assumptions* about the environment in which the protocol is executed, 3) the *vocabulary* of messages used to implement the protocol, 4) the *encoding* (format) of each message in the vocabulary, and 5) the *procedure rules* guarding the consistency of message exchanges.

The *Design Science Research Process Model* (DSRPM) as introduced by Vaishnavi and Kuechler (cf. Vaishnavi and Kuechler, 2015 pp. 14-18) has been applied to develop the artifact. This model comprises five process steps which are frequently iteratively performed: *awareness of problem*, *suggestion*, *development*, *evaluation*, and *conclusion*. The awareness of the problem described before originated from casual conversations with industry professionals. As an initial solution, a system that allows end user development of simple problem specific programming languages for communication that are designed applying everyday language and that thus may be understood by the average employee, has been first suggested in a casual setting to an industry professional during the summer of 2020. Since then, multiple DSRPM iterations have been conducted. The artifact version discussed here was subject of more extensive and structured interviews with industry experts during the spring of 2021. For those interviews 9 experts have been reached out to, 7 responded, and 5 confirmed, that they were willing to allocate the necessary time. The conducted interviews took up to two hours and had 3 stages: 1) explanation of the theory, 2) presentation of tangible examples from logistics, and 3) a detailed discussion. In order to receive comparable results such a discussion was guided by a pre-prepared presentation containing questions and discussion points that dealt with the applicability of the artifact in logistics, communication in logistics, and the technical implementation of the artifact. A more extensive discussion of the conducted interviews will be published in a more extensive contribution. 1 expert is the CEO of an automotive supplier, 2 experts have a leadership role in departments that are responsible for business innovations; 1 at a global logistics service provider and 1 at a global manufacturing cooperation. 1 expert is a consultant currently working on a communication automation project in logistics, and the final expert is a purchasing agent at a pharmaceutical trading company. This diverse field was approached to gain rather diverse feedbacks.

The artifact is supposed to be applicable for the standardization of communication on the operational level, the 'day-to-day' communication, between organizational dyads in logistics. Thus here, we are neither dealing with communication within one organization, nor with communication within a higher level polyad. Following the psychologist Tomasello, the three cooperative social motivations for communication among humans *requesting*, *informing*, and *sharing* do exist (Tomasello, 2008 pp. 82-88). The motivation for communication that may be standardized applying the artifact should be *requesting* or *informing*. Finally to allow for standardization the communication should be either scheduled or non-scheduled and in addition, it should be granular, it should reoccur regularly, and it should have a low variance between occurrences.

The artifact builds on the previously discussed model by Holzmann. The experts were asked if that model seems applicable to logistics and the tendency of the experts has been, that the model is plausible. Protocols that originate applying this artifact may be easily modified due to the limited number of partners. A new version of the protocol is created when it is modified. The information that is needed for each one of the five distinct parts of a protocol for the domain of logistics has been worked out during the iterations of the DSRPM. For each distinct part the experts have been asked to provide feedback. Eventually the following information requirements have been identified for 1) service, 2) assumptions, and 5) procedure rules: **1) service**: a) the version and name of the protocol, b) the motivation for the communication, c) the responsible contact persons at each organization and the two participating organizations, d) a reference to the base language of the applied CNL, and, if applicable, a list of equivalent protocols that apply different CNLs, and e) some keywords allowing to identify and find the protocol in a protocol management system or database; **2) assumptions**: a) how messages applying the protocol are communicated, b) the encoding technologies of the CNL (how the grammar is defined, e.g. (E)BNF, Grammatical Framework, ...) c) how messages applying the protocol are created (by human/ by machine and the solution for the writability problem), and d) how messages applying the protocol are processed; **5) procedure rules**: a) when, which organization initiates the communication, and b) if there is a follow up to the communication. The **3) Vocabulary** is a plain language area for: comments, feedback, and explanations. According to the International Plain Language Federation, communication is in plain language "if its wording, structure, and design are so clear that the intended readers can easily find what they need, understand what they find, and use that information"(Onl). The *vocabulary* is written for a neither technical nor academical audience containing all the information deemed necessary to apply the protocol in the field, helping to prevent errors and misunderstandings. The semantics of a formal CNL is here taken from the base language. Thus, due to e.g. homonyms or jargon misunderstandings are still possible. It is important that as part of the *vocabulary* semantics of such cases may be explained. During the specification of the protocol it is important to determine what needs to be part of the *vocabulary*. Nevertheless, should an error or misunderstanding occur, a new protocol version preventing such an occurrence in the future may be simply created. The *vocabulary* should aim to be as compact as possible and it shouldn't deter by seemingly looking to complex, e.g. by containing anything but plain language, or by looking to extensive, e.g. by being some kind of lexicon. The **4) Encoding** has to describe the syntax of the messages exchanged. Either a new case specific CNL may be expressed or an existing CNL, e.g. RECON, may be applied. If a new CNL is expressed the grammar of the CNL needs to be expressed; e.g. a syntax diagram may be depicted. If an existing CNL is applied an external link to the specification may be provided or the applied aspect of that CNL may be cited. The syntax should ensure that created messages allow identification of the applied protocol. Similarly to the *vocabulary*, the *encoding* should aim to be as compact as possible and it shouldn't deter by seemingly looking to complex or to extensive. More complex communication may be broken down into multiple smaller communication units, for which separate protocols with more case specific CNLs may be created. In addition a process chain diagram from the domain of logistics as depicted in Fig. 1 may be specified as part of the protocol specification if the communication is scheduled communication. Finally a diagram called 'stakeholder communication diagram' depicted in Fig. 2, developed, based on feedback provided during the interviews and inspired by so-called 'dialogue trees' (see Adams, 2010 p. 186), may also be part of the specification. The example diagram is part of the specification of protocol B (indicated by the red arrow). The Initial Stakeholder is able to communicate with Stakeholder 1 applying Protocol A and with Stakeholder 2 applying Protocol B. Stakeholder 2 is optionally - as indicated by the dashed arrow - able to respond back to the Initial Stakeholder applying Protocol C. Stakeholder 2 is also able to communicate to Stakeholder 3 - 'forwarding' the information - applying Protocol D.

4 Conclusion

The tendency of the expert opinion was, that there is a place for the presented artifact in logistics. Nevertheless, the tenor was, that adequate information systems are necessary to allow the application by an average employee. Such a

Figure 1: A process chain diagram depicting a communication process and the processes surrounding this process. Own depiction based on Baumgarten and Inga-Lena, 2000

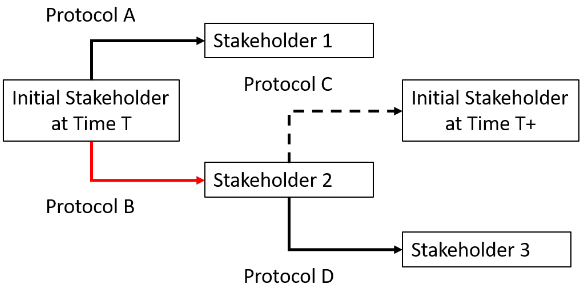

Figure 2: A stakeholder communication diagram.

system requires a protocol creation and versioning component. Protocols implemented applying the artifact should be integratable into systems that allow integration. Due to the limitation - only communication between dyads - and the flexibility of the encoding, modifications are easily implemented. The participating organizations are in charge of which information they share due to the same reason. Communication in a multi-lingual and multi-cultural environment should also be possible by creating equivalent protocols applying CNLs applying different base languages. To summarize, the artifact promotes the application of formal CNL in the domain of logistics.

There are multiple possible future research directions. More research into the application of the presented artifact in the field is required. This type of protocol discussed, may also be applicable in different domains. Additionally, research towards information systems, that would allow the application by the average employee, is needed. During the interviews multiple experts remarked, that it would be helpful, if the protocols are able to apply machine learning to further develop themselves. Thus dynamic protocols, that apply machine learning to further develop themselves, may also be an interesting topic for future research.

References

Iplf, plain language. www.iplfederation.org/plain-language/. Accessed: 2021-07-05.

ISO/IEC/IEEE 24765. 2017. Iso/iec/ieee international standard - systems and software engineering–vocabulary. *ISO/IEC/IEEE 24765:2017(E)*.

Ernest Adams. 2010. *Fundamentals of game design*, 2nd edition. Pearson Education.

Edward Barkmeyer and Andreas Mattas. 2012. *A Restricted English for Constructing Ontologies (RECON)*. NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD.

Helmut Baumgarten and Darkow Inga-Lena. 2000. Prozesskettenmanagement als basis für logistikmanagement. In *Logistik-Management Strategien Konzepte Praxisbeispiele*, 1st edition, chapter 2.3.1. Springer-Verlag.

Alberto Rodrigues Da Silva, João Saraiva, David Ferreira, Rui Silva, and Carlos Videira. 2007. Integration of re and mde paradigms: the projectit approach and tools. *IET software*, 1(6):294–314.

Peter Detzner, Thomas Kirks, and Jana Jost. 2019. A novel task language for natural interaction in human-robot systems for warehouse logistics. In *2019 14th International Conference on Computer Science & Education (ICCSE)*, pages 725–730. IEEE.

Michael A. Harrison. 1978. *Introduction to Formal Language Theory*. Addison-Wesley Series In Computer Science. Addison-Wesley Publishing Company.

Gerard J Holzmann. 1991. *Design and validation of computer protocols*. Prentice hall Englewood Cliffs.

Tobias Kuhn. 2014. A survey and classification of controlled natural languages. *Computational linguistics*, 40(1):121–170.

Martin Manns, Klaus Fischer, Han Du, Philip Slusallek, and Kosmas Alexopoulos. 2018. A new approach to plan manual assembly. *International Journal of Computer Integrated Manufacturing*, 31(9):907–920.

Alexandru Mateescu and Arto Salomaa. 1997. Formal languages: an introduction and a synopsis. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages: Volume 1 Word, Language, Grammar*, pages 1–39. Springer-Verlag New York, Inc., New York, NY.

Michael Tomasello. 2008. *Origins of Human Communication*. MIT press.

Vijay K Vaishnavi and William Kuechler. 2015. *Design science research methods and patterns: innovating information and communication technology*, 2nd edition. CRC Press.

Arie Van Deursen, Paul Klint, and Joost Visser. 2000. Domain-specific languages: An annotated bibliography. *ACM Sigplan Notices*, 35(6):26–36.