A Framework of Controlled Robot Language for Reliable Human-Robot Collaboration

Dang Tran[1], Fujian Yan[1], Yimesker Yihun[2], Jindong Tan[3], and Hongsheng He[1(✉)]

[1] School of Computing, Wichita State University, Wichita, KS 67260, USA
hongsheng.he@wichita.edu

[2] Department of Mechanical Engineering, Wichita State University, Wichita, KS 67260, USA

[3] Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, Knoxville, TN, USA

Abstract. Effective and efficient communication is critical for human-robot collaboration and human-agent teaming. This paper presents the design of a Controlled Robot Language (CRL) and its formal grammar for instruction interpretation and automated robot planning. The CRL framework defines a formal language domain that deterministically maps linguistic commands to logical semantic expressions. As compared to Controlled Natural Language, which aims for general knowledge representation, CRL expressions are particularly designed to parse human instructions in automated robot planning. The grammar of CRL is developed in accordance with the IEEE CORA ontology, which defines the majority of formal English domain, accepting large range of intuitive instructions. For sentences outside the grammar coverage, CRL checker is used to detect linguistic patterns, which can be further processed by CRL translator to recover back an equivalent expression in CRL grammar. The final output is formal semantic representation using first-order logic in large discourse. The CRL framework was evaluated on various corpora and it outperformed CRL in balancing coverage and specificity.

1 Introduction

Reliable communication between humans and intelligent robot is a critical need especially for human-robot collaboration, human-agent teaming, and multiple agent coordination. For most robotic applications, reliable human-robot communication will significantly reduce the chances of unpredictable catastrophes and fatal damages. In addition to physical interaction, natural language has the potential to become the main communication channel for instructing robots, representing contextual knowledge, and providing feedback. From a psychological perspective, *trust* is the grand obstacle preventing human and robot from

This work was partially supported by NSF CMMI 2129113 and WSU URCAF 2019.

H. Li et al. (Eds.): ICSR 2021, LNAI 13086, pp. 339–349, 2021.
https://doi.org/10.1007/978-3-030-90525-5_29

communicate effectively [1]. A robot with dynamic consciousness and optimized precision does not necessarily gain trust from its users. We believe that the lack of reliable natural language communication is one of the main factors that hinder the advancement of human-agent teaming.

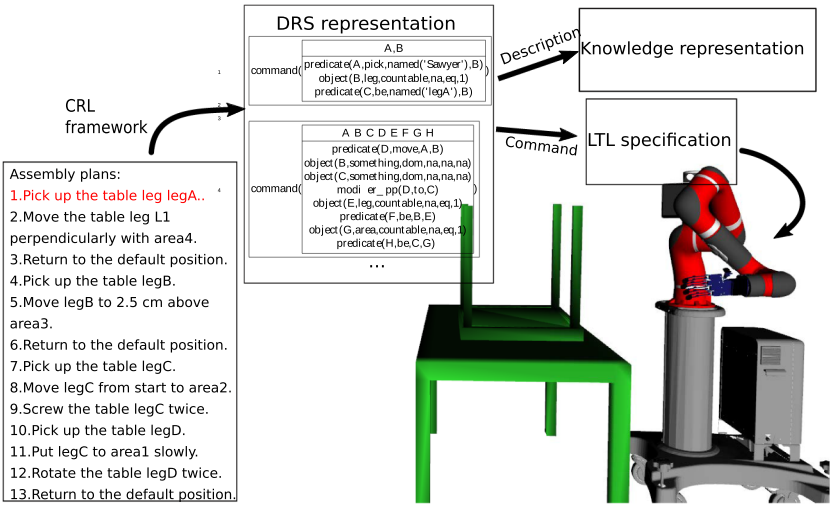

Fig. 1. Parsing of natural language instructions using CRL. The linguistic instructions from user are converted into appropriate semantic representations using Discourse Representation Structure with variables and explicit logical statements. Command-type DRS expressions can be used to control robot through planning.

Significant research progress has been made to address the importance and challenge of reliable natural language communication in robotic domain [2]. Robots are deemed to understand natural language if the robot can either (i) extract correct information or (ii) have a logical semantic representation for the context. From the former perspective, natural language understanding is considered as shallow parsing low-level knowledge for action control. The most common approaches in this branch include probabilistic models [3] and neural-network methods [4], which have demonstrated robust performance in detecting linguistic patterns and triggering low-level actions. These approaches focus on partial information only, which is not representative enough for describing complete planning scenarios in realistic cases. The latter perspective parses the natural language in a richer way, where the semantics of natural language expressions are representable in logic form [5].

Despite the fruitful research progress of human-robot communication, natural language based interfaces in robotics are still limited due to the trade-off between *expressiveness* and *reliability*. Linguistic models that can handle a large range of natural language expressions are less deterministic or reliable; on the other hand, systems that can understand natural language inputs in-depth are limited

in expressiveness. A common language domain could bridge the gap between expressiveness and reliability in human-robot interaction.

In this paper, we propose a *grammar model* named Controlled Robot Language (CRL) that interprets general human instructions into Discourse Representation Structures (DRS) [6], which is a semantic format that can capture various partial information into the same data structure. More importantly, the CRL is designed for general-purpose that represents a *deterministic* and *expressive* linguistic domain rather than an ad-hoc development, inspired by Controlled Natural Language (CNL). The CRL framework defines CRL grammar that syntactically captures a majority of English expressions. For dialect expressions that do not follow the grammar, we developed a CRL checker, which contains flexible set of common linguistic patterns, to detect correctable grammar errors. In the CRL domain, expressions following the grammar are parsed into corresponding formal representations, which contain all essential linguistic information for robotic planning with reference to IEEE CORA standards [7]. As shown in Fig. 1, given a sequence of natural language instructions, the CRL parser generates the corresponding syntactic structure for expressions without errors in grammar, or corrects fixable patterns for expressions with errors in grammar. The CRL framework will translate the instructions into knowledge representations or robot action planning.

We plan to address two fundamental challenges. The first challenge is to design a linguistic model that is comprehensive and unambiguous. To the best of our knowledge, there is no work so far addressing the importance of these two properties equivalently and simultaneously. The second challenge is to automatically transform or represent formal representation into robot knowledge and actions. A primary objective of human-robot communication is to enable high-level mutual understanding and automated planning. There is limited research addressing on expanding the natural language domain in human-robot collaboration. The main contributions of this paper are:

1. We designed and implemented a linguistic grammar tailed for robotic applications, which achieves both reliability and expressiveness; and
2. We developed a complete framework and proposed a methodology in translating natural language instructions into corresponding robot actions.

2 The Framework of Controlled Robot Language (CRL)

In view of the lacking of linguistic interfaces that satisfy reliability and expressiveness, we aim to implement a framework of Controlled Robot Language (CRL) for robot understanding and planning. Motivated by Controlled Natural Language, the proposed model maintains essential properties: certainty and generality. The proposed framework contains three fundamental components: CRL grammar, a CRL parser, and a CRL translator. The CRL grammar defines a formal language domain – a set of common English instructions. CRL parser is used to assign equivalent syntactic structures to input commands. From syntactic structures, partial linguistic information, such as *subject, object, predicate, noun*

modifier, and predicate modifier, can be extracted and constructed back into formal representations. The CRL grammar is designed toward a deterministic and reliable interpretation with no ambiguity. We limited the set of CRL grammar for a compact and efficient grammar core. The sentences outside the domain are further processed by the CRL parser in accordance to predefined dialect patterns – set of sentence patterns that exist in daily conversations but are not generalized enough to be represented as grammar rules. These dialect patterns can be recognized and transformed by the CRL translator, which returns a valid expressions but in CRL grammar domain.

2.1 CRL Grammar and Dialect Patterns

Grammar is the core component of the proposed framework. We designed the CRL grammar as a general-purpose and deterministic grammar in English. The CRL grammar defines an unambiguous formal language domain, which can be computational efficiently processed by a computer, but still expressive enough to allow natural usage. We defined and developed the CRL grammar in terms of Context Free Grammar (CFG) with selective rules to avoid unnecessary ambiguity.

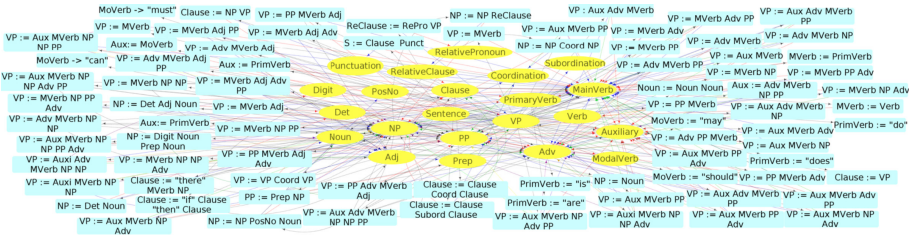

Fig. 2. CFG productions for CRL grammar – implicit descriptions of CRL grammar.

Given a set of terminal nodes associated with a set of terminal symbols $\mathcal{T}$ and nonterminal nodes associated with a set of nonterminal symbols $\mathcal{N}$, we defined grammar using CFG paradigm. CFG grammar is a collection of linguistic productions in the form of

$$X \rightarrow \{Y_i\}_i^n \ \{\alpha_j\}_j^m \tag{1}$$

where $X \in \mathcal{N}$, $Y_i \in \mathcal{N} \cup \mathcal{T}$ and $\alpha_j \in \mathcal{T}$. To support common instruction structures and IEEE CORA ontology, we formulate the CRL grammar by a set of about 330 productions[1], including a set of standard Penn Treebank POS tags and 30 additional new nonterminals. A fragment of grammar is visualized in Fig. 2. The grammar defines constraints on language domain, aiming to optimally avoid unnecessary ambiguity. The grammar captures a majority of common linguistic

[1] The complete grammar and parsers are available at: https://github.com/hhelium.

expressions in real-world scenarios, e.g., *"A robot picks a red apple on the table."* and *"Which apple is red?"*.

CRL grammar was initially developed to express the core grammar production in form of S → NP VP and its variants. During the grammar developing process, attachment ambiguity and coordination ambiguity appeared. To avoid solving ambiguities by designing a complex disambiguation model, dialect pattern recognition is used to detect patterns that can potentially lead to ambiguities, and transform them back to equivalent CRL linguistic expressions. These dialect patterns are manually designed based on WikiHow instructions [8]. This dynamic set can be flexibly expanded and modified in accordance to the chosen language domain. As shown in Table 1, to make CRL applicable in WikiHow corpus domain, input sentences must be normalized and transformed into equivalent representation following each dialect pattern's correcting rule.

Table 1. Grammar dialect patterns for CRL checker.

Dialect Pattern	**Examples**	**Linguistic Correction**
Your/our/my pattern	*your* hand	change to a possessive pronoun
You/I/we pattern	*you* can move	replace with a subject pronoun
Plural nouns	*cubes*	singularize
Metrics	3 kilograms, 1 ton	map to standard metric units
Literal quantity	one half; quarter; dozen	map to standard metric units
Imperative	*rotate* the leg	*"robot"* as the default subject
Compound noun	*table leg*	last noun as the main noun
Consecutive adjectives	a *small red* apple	add conjunction "and"
Consecutive adverbs	*gradually slowly* move	add conjunction "and"
Verb+obj +to+verb	click the screen to start	restructure
Verb + to + verb	have to wait	add modal
Verb+gerund	consider stopping	gerund as the main verb
From-to pattern	*from 0.1 to 0.2 cm*	standardize the structure
Passive voice	is picked by	convert to active voice
Progress description	*by grasping its hand*	convert the gerund to a verb

2.2 CRL Parser

With the defined CRL grammar, we constructed a syntactic parser to analyze syntactic structures of natural language descriptions. We utilized a common Bottom-Up dynamic programming parsing algorithm Cocke-Kasami-Younger (CKY) to find all syntactic structures for a given sentence with efficient run time complexity – $\mathcal{O}(n^3 \cdot |G|)$ where n is the length of input, and $|G|$ is the size of CFG grammar. Although CRL grammar was carefully selected to avoid unnecessary syntactic ambiguities, multiple parsing structures for NL input are

inevitable. To maintain the determinism property of the robot system without building a disambiguation model, appeared ambiguities are manually resolved by user's instructions. These instructions can be collected, trained by a probabilistic model and automatically used in the future cases.

2.3 Translating CRL Descriptions to Knowledge Representations

Leveraging the parsed syntactic structures, the CRL framework constructs a contextual semantic representation of natural language expressions. These semantic representations are essential for robot understanding and planning. Each semantic representation can be either classified into one of three categories: contextual descriptions, command instructions, and queries. The instructions correspond to robot planning, and the contextual descriptions specify environment context, temporal and spatial constraints. We developed a program that automatically translated these semantic expressions into corresponding LTL specifications, which can be executed directly by robot planners.

Given the syntactic structure, the framework extracts fundamental linguistic components such as *object*, *predicate*, *property*, *adjunct,* and *phrase*. The extracted information is, however, discrete and exclusive. To unify them into a single semantic representation, we need semantic rules to depict combining procedures. These rules are best described using symbolic language as Prolog [9]. The main challenge of this process is the ability to create general rules that can unify low-level knowledge. Appropriate selection of semantic rules goes beyond a combination task: solve anaphoric problem, identify quantification property, and describe temporal constraints. In this paper, we focus on a set of semantic rules for discourse representations, which also handle anaphora binding and determiners quantifying. The theoretical details of these rules can be found at Kamp's lecture [10]. The developed CRL system expanded the semantic rules as defined in [11] by unifying discrete components and adding a few abstraction layers. These abstraction layers provide convenient ways to identify objects, trigger actions, or query. By using additional abstraction layers, we can easily classify all possible formal representation into three main types: (i) *Description-type*: used to describe an event, robotic environment, and knowledge base; (ii) *Command-type*: used to trigger robot actions; (iii) *Query-type*: extract information back from the knowledge base.

3 Experiment

We evaluated the CRL framework in parsing performance, formal robot planning, and user acceptance. The CRL framework is compared with two linguistic frameworks, Attempto Controlled English [12] and Stanford Constituency parsing [13], on instruction-based corpora WikiHow [8] and Collaborative Manipulation [14]. The effectiveness of the CRL framework in automated planning is also

demonstrated in an example of instruction-based furniture assembly. At last, we quantified the user acceptance of CRL in terms of correctness, complexity, ambiguity, readability, and efficiency.

3.1 Performance of Natural Language Parsing

The CRL framework was tested on instructions in Collaborative Manipulation [8] and WikiHow corpus [14] for planning domain. The Collaborative Manipulation is a standard dataset, which is a corpus of 1670 natural language sentences. The WikiHow dataset is a collection of human describing procedural task using step-by-steps instruction style. The dataset consists of 9520 instruction sentences. The performance of CRL in grammar coverage and semantics representation is compared with ACE framework [12] and Stanford Constituency parsing [13]. As showed in Table 2, the CRL is outperforming ACE in expressiveness in term of grammar coverage, but not as competitive as the induced-grammar-based methods Stanford CoreNLP. Nonetheless, CRL can express all parsed instructions into complete semantic representations, whereas Stanford CoreNLP was not designed with this capability. This is the conundrum in solving the contradiction between coverage and formal representation. To represent a linguistic instruction into formal semantics, the domain grammar needs to be structured; on the contrary, structured grammar cannot cover all free-form natural language. Though defined in general language domain, the ACE framework cannot understand any sentences in the WikiHow corpus, due to the extensively use of invalid expressions in ACE grammar.

Table 2. Performance comparison on standard corpora.

	Collaborative Manipulation (1670)	
	Grammar coverage	Formal semantics
ACE [12]	140 (8.38%)	140 (8.38%)
CRL (this paper)	739 (44.25%)	739 (44.25%)
CoreNLP [13]	1670 (100%)	N/A
	WikiHow (9520)	
	Grammar coverage	Formal semantics
ACE [12]	0 (0%)	0 (0%)
CRL (this paper)	3898 (40.95%)	3898 (40.95%)
CoreNLP [13]	9520 (100%)	N/A

The CRL framework has a flexible set of linguistic patterns, allowing flexible adaption to various scenarios. The experiment showed the same trail in the Collaborative Manipulation corpus. It is important to guarantee deterministics and reliability in human-robot collaboration for predictable system behavior.

3.2 CRL-Based Robot Planning

We implemented an automated planning system based on CRL for the assembly of an IKEA table by following natural language instructions. The robotic system consists of a Sawyer manipulator with an AR10 robotic hand, which support context-aware task-oriented manipulation [15,16]. We employed the RRTConnect [17] planner for low-level control, and developed a Python modules for the CRL interface. The control and communication are implemented as ROS services. We also utilized the Spot [18] to handle LTL specification by building reactive system from DRS instructions, and used RViz for simulation and visualization.

Figure 3 shows the results of successful action execution by following natural language instructions to assemble an IKEA table step by step. Four primitive actions were developed and grounded: *pick, place, release,* and *rotate*. The complete implementation of 13 instruction scripts took about 4 minutes to finish, but there is a lot of room for optimization both in CRL parsing and action planning.

3.3 User Acceptance of CRL

We evaluated the acceptance of CRL in terms of the discrepancy between CRL parsing and human perception. We designed questionnaires containing five binary (yes/no) evaluation criteria: *correctness, complexity, ambiguity, readability,* and *efficiency*. 1) The correctness is defined as whether the parsed results are correct in DRS forms, i.e., the parsed DRS representations are correct and semantically consistent with the input sentences; 2) The complexity evaluates whether the parsing results contain additional redundant words or phrases generated by dialect correction; 3) The ambiguity in semantics is used to examine if the CRL can handle a potentially ambiguous natural language input; 4) The readability is used to evaluate whether the CRL parses a potentially ambiguous input in the same way as a human does; 5) The efficiency is used to evaluate whether the parsing results are sufficient for further robotic applications such as task planning. Ten participants were invited to evaluate 350 CRL parsing results. As shown in Fig. 4, the CRL has high acceptance and consistence with human understanding of natural language instructions.

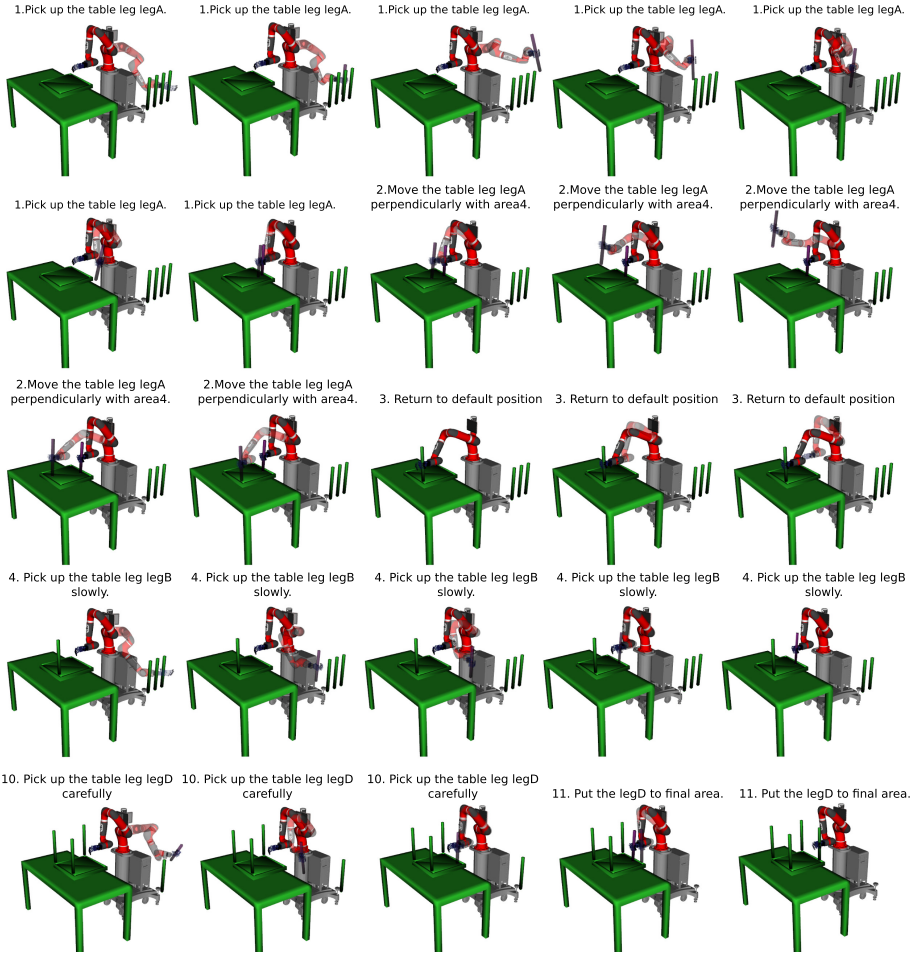

Fig. 3. Assembly of a furniture table by following natural language instructions based on CRL.

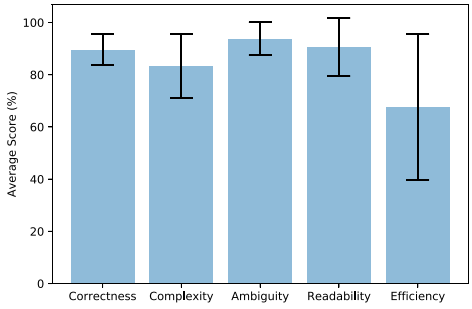

Fig. 4. User acceptance of CRL (higher scores are better).

4 Conclusion

We proposed a CRL framework that ensures reliability and expressiveness for natural language communication. We also demonstrated the procedure to integrate the CRL framework into robotic planning: from building a complete semantic representation to mapping those representations into robotic actions. The experiment showed the performance of the CRL frame in parsing natural language instructions, and demonstrated the effectiveness and flexibility of the CRL framework for automated robot planning.

References

1. Yagoda, R.E., Gillan, D.J.: You want me to trust a robot? The development of a human-robot interaction trust scale. Int. J. Soc. Robot. **4**, 235–248 (2012)
2. Liu, R., Zhang, X.: Methodologies for realizing natural-language-facilitated human-robot cooperation: a review. CoRR, vol. abs/1701.08756 (2017)
3. Salvi, G., Montesano, L., Bernardino, A., Santos-Victor, J.: Language bootstrapping: learning word meanings from perception-action association. CoRR, vol. abs/1711.09714 (2017)
4. Bisk, Y., Yuret, D., Marcu, D.: Natural language communication with robots. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 751–761 (2016)
5. Tenorth, M., Beetz, M.: KnowRob: a knowledge processing infrastructure for cognition-enabled robots. Int. J. Rob. Res. **32**, 566–590 (2013)
6. Kamp, H., Van Genabith, J., Reyle, U.: Discourse representation theory. In: Handbook of Philosophical Logic, pp. 125–394. Springer, Dordrecht (2011). https://doi.org/10.1007/978-94-007-0485-5_3
7. Prestes, E., et al.: Towards a core ontology for robotics and automation. Robot. Auton. Syst. **61**(11), 1193–1204 (2013)
8. Koupaee, M., Wang, W.Y.: WikiHow: a large scale text summarization dataset (2018)
9. Bratko, I.: Prolog Programming for Artificial Intelligence. Pearson Education (2001)
10. Kamp, H., Genabith, J., Reyle, U.: Discourse Representation Theory, pp. 125–394 (2010)
11. Kuhn, T.: Controlled English for knowledge representation. Ph.D. thesis, University of Zurich (2010)
12. Fuchs, N.E., Schwitter, R.: Attempto controlled English (ACE). arXiv preprint cmp-lg/9603003 (1996)
13. Klein, D., Manning, C.D.: Accurate unlexicalized parsing. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1, (USA), pp. 423–430, Association for Computational Linguistics (2003)
14. Scalise, R., Li, S., Admoni, H., Rosenthal, S., Srinivasa, S.S.: Natural language instructions for human-robot collaborative manipulation. Int. J. Rob. Res. **37**(6), 558–565 (2018)
15. Li, H., Tan, J., He, H.: MagicHand: context-aware dexterous grasping using an anthropomorphic robotic hand. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 9895–9901 (2020)

16. Rao, A.B., Krishnan, K., He, H.: Learning robotic grasping strategy based on natural-language object descriptions. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 882–887 (2018)
17. Kuffner, J.J., LaValle, S.M.: RRT-connect: an efficient approach to single-query path planning. In: Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065), vol. 2, pp. 995–1001. IEEE (2000)
18. Duret-Lutz, A., Lewkowicz, A., Fauchille, A., Michaud, T., Renault, É., Xu, L.: Spot 2.0 — a framework for LTL and ω-automata manipulation. In: Artho, C., Legay, A., Peled, D. (eds.) ATVA 2016. LNCS, vol. 9938, pp. 122–129. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46520-3_8