"Are you telling me to put glasses on the dog?" Content-Grounded Annotation of Instruction Clarification Requests in the CoDraw Dataset

Brielen Madureira[1] **David Schlangen**[1,2]
[1]Computational Linguistics, Department of Linguistics
University of Potsdam, Germany
[2]German Research Center for Artificial Intelligence (DFKI), Berlin, Germany
{madureiralasota,david.schlangen}@uni-potsdam.de

Abstract

Instruction Clarification Requests are a mechanism to solve communication problems, which is very functional in instruction-following interactions. Recent work has argued that the CoDraw dataset is a valuable source of naturally occurring iCRs. Beyond identifying when iCRs should be made, dialogue models should also be able to generate them with suitable form and content. In this work, we introduce CoDraw-iCR (v2), extending the existing iCR identifiers with fine-grained information grounded in the underlying dialogue game items and possible actions. Our annotation can serve to model and evaluate repair capabilities of dialogue agents.

Introduction If someone requests you to put glasses on a dog, you may doubt yourself: *Is that really what I am supposed to do?* Before attempting that, you'd likely seek confirmation, for instance, by posing a clarification request. In real life, dogs do fine without glasses, but, as we see in Figure 1, that is indeed a correct action in the context of a scene construction dialogue game.

In instruction following settings, ambiguous or underspecified instructions may elicit clarification requests when the instruction follower realises they cannot act properly without further information. These are Instruction Clarification Requests (iCRs), namely CRs that occur in Clark's 4th level of communication (Clark, 1996), when an utterance (here, an instruction) is understood generally, but not at the level of uptake (Schlöder and Fernández, 2014).

We have recently argued that the CoDraw dataset (Kim et al., 2019) is a rich and large source of spontaneous iCRs (Madureira and Schlangen, 2023). We identified iCRs among all instruction follower utterances and proposed using the annotation to model the tasks of knowing *when* to ask and to reply to an iCR. However, knowing *what* and *how* to ask are also topical devices for a competent instruction follower dialogue model. To account for that, we continue this initiative by adding information

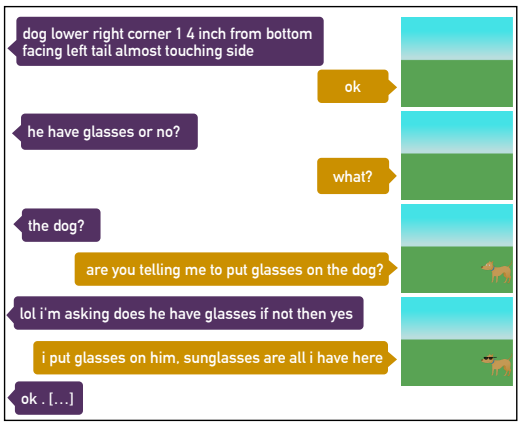


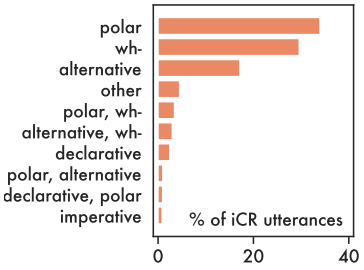
Figure 1: A communication problem occurring and being resolved with the aid of clarification requests in an instruction following interaction (CoDraw, ID 9429, CC BY-NC 4.0, scene from Zitnick and Parikh (2013)). When an instruction is not clear enough, the instruction follower asks for clarification, in order to act accordingly (here, placing cliparts in the scene).

about the *content* and *form* of iCRs, in order to allow modelling and evaluating the subsequent task of *generating* iCRs, not yet explored in this corpus.

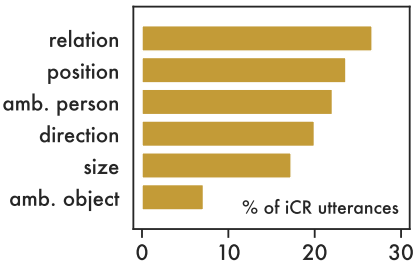
Our annotation complements CoDraw-iCR (v1) by adding mood categories and by mapping each utterance to its corresponding objects and action-related attributes. We show that this sample is an appealing ensemble of mostly unique surface forms through which interesting relations in co-occurring objects and attributes emerge, making it a handy resource for further CR research. The data and documentation is available for the community at https://osf.io/gcjhz/, which also contains a link to an extended version of this summary.

Background Clarification Requests are a multi-faceted phenomenon in dialogue, with vast literature on categorising, documenting and modelling their various realisations as well as their relations to other utterances and to the context (Purver et al., 2003; Gabsdil, 2003; Rodríguez and Schlangen,

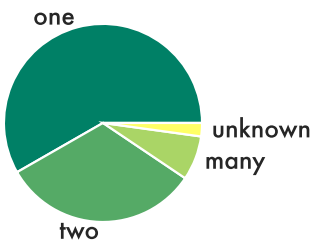
Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue, August, 16–17, 2023, Maribor.



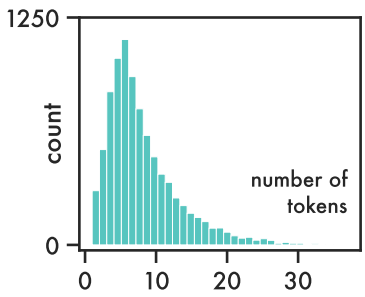
(a) Ten most common moods



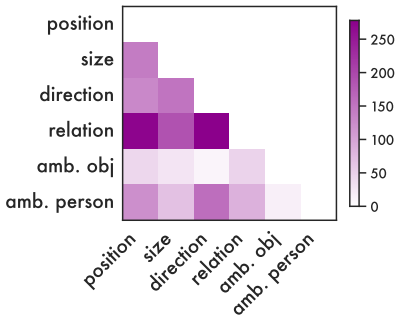
(b) Relative frequency of attributes



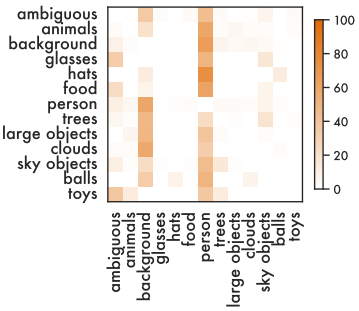
(c) Quantity of mentioned cliparts



(d) Distribution of length



(e) Frequency of co-occurring attributes.



(f) Co-occurrence of thematic entity categories, distribution by row in %.

Figure 2: Overview of the distributions of annotated categories in CoDraw iCR utterances.

2004, *inter alia*). Still, it remains an open research area; in particular, we cannot delineate yet to what extent CR mechanisms can be learnt via data-driven methods (Benotti and Blackburn, 2021), and dealing with underspecifications is still hard for pre-trained language models (Li et al., 2022).

Benotti and Blackburn (2021) have recently raised awareness to the different world modalities upon which clarifications can be *grounded*, like vision, movement or physical objects. Still, few works exist that systematically map the content of CRs to elements related to the context where they occur (Gervits et al., 2021). Some examples are Benotti and Blackburn (2017), who use a methodology to classify CRs according to why they make implicated premises explicit (*e.g.* wrong plan, not explainable plan or ambiguous plan in instruction giving), in a corpus that is further analysed in Benotti and Blackburn (2021) with a recipe to detecting *grounded clarifications*. Gervits et al. (2021) propose a fine-grained annotation schema for CR types related to the environment (object location, feature, action, description, etc). The small size of these corpora, however, does not meet the needs of current data-driven methods.

Corpus Overview 8,765 utterances (7,710 types) were identified as iCRs in CoDraw. Figure 2 presents an overview of the annotation. The immediately preceding instruction giver utterance is the source utterance (*i.e.* the utterance where the communication problem manifests) for 80.26% of the iCRs and 78.49% of the iCR utterances get a response from the instruction giver in the immediately following turn. For 63.85% of them, both conditions are true. They are realised in many surface forms, ranging from short and generic (*sorry?*), to very specific (*owl is med?*), to long and verbose (*is the girl sitting or standing i need to know as there are multiple options and her expression as well*). Besides, the iCRs cover all available objects and are well distributed among actions.

Outlook Given the need for large scale corpora for data-driven methods, trading some of the ecological validity in the annotation process for machine-learnability was necessary. Still, even in its controlled environment with a limited number of actions and objects, the resulting iCR utterances are very diverse in surface form and very fertile in content. With the release of the annotation, the community gains a larger resource with sequential, spontaneous iCRs in turn-based dialogues. We aim to encourage more research on modelling CRs in instruction following interactions, and also to enable detailed evaluation of iCR generation.

Acknowledgements

We are very thankful to our student assistants, Sebastiano Gigliobianco, for performing the initial identification and implementing the GUI and Sophia Rauh, for doing the fine-grained annotation. We also thank the anonymous reviewers for their feedback.

References

Luciana Benotti and Patrick Blackburn. 2017. Modeling the clarification potential of instructions: Predicting clarification requests and other reactions. *Computer Speech & Language*, 45:536–551.

Luciana Benotti and Patrick Blackburn. 2021. A recipe for annotating grounded clarifications. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4065–4077, Online. Association for Computational Linguistics.

Herbert H Clark. 1996. *Using language*. Cambridge university press.

Malte Gabsdil. 2003. Clarification in spoken dialogue systems. In *Proceedings of the 2003 AAAI Spring Symposium. Workshop on Natural Language Generation in Spoken and Written Dialogue*, pages 28–35.

Felix Gervits, Antonio Roque, Gordon Briggs, Matthias Scheutz, and Matthew Marge. 2021. How should agents ask questions for situated learning? an annotated dialogue corpus. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 353–359, Singapore and Online. Association for Computational Linguistics.

Jin-Hwa Kim, Nikita Kitaev, Xinlei Chen, Marcus Rohrbach, Byoung-Tak Zhang, Yuandong Tian, Dhruv Batra, and Devi Parikh. 2019. CoDraw: Collaborative drawing as a testbed for grounded goal-driven communication. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6495–6513, Florence, Italy. Association for Computational Linguistics.

Haau-Sing Li, Mohsen Mesgar, André FT Martins, and Iryna Gurevych. 2022. Asking clarification questions for code generation in general-purpose programming language. *arXiv preprint arXiv:2212.09885*.

Brielen Madureira and David Schlangen. 2023. Instruction clarification requests in multimodal collaborative dialogue games: Tasks, and an analysis of the CoDraw dataset. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2303–2319, Dubrovnik, Croatia. Association for Computational Linguistics.

Matthew Purver, Jonathan Ginzburg, and Patrick Healey. 2003. On the means for clarification in dialogue. In *Current and new directions in discourse and dialogue*, pages 235–255. Springer.

Kepa Joseba Rodríguez and David Schlangen. 2004. Form, intonation and function of clarification requests in german task-oriented spoken dialogues. In *Proceedings of Catalog (the 8th workshop on the semantics and pragmatics of dialogue; SemDial04)*.

Julian Schlöder and Raquel Fernández. 2014. Clarification requests at the level of uptake. In *Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue - Poster Abstracts*, Edinburgh, United Kingdom. SEMDIAL.

C Lawrence Zitnick and Devi Parikh. 2013. Bringing semantics into focus using visual abstraction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3009–3016.