

# Chat-o-matic: an online chat tool for collecting conversations of situated dialogue

Simon Dobnik and Aram Karimi

Department of Philosophy, Linguistics and Theory of Science (FLoV)

Centre for Linguistic Theory and Studies in Probability (CLASP)

University of Gothenburg, Sweden

name.surname@gu.se

## 1 Introduction

One of the major challenges for modelling situated dialogue is a collection of quality natural language data in situational contexts. Such data involves long and sometimes persistent dialogue conversations over situated scenes between several conversational partners. Scenes may also change over time either because new events occur or conversational participants change locations so dialogue history has to be matched with different states in the perceptual environment. Collection of such data requires significant time as conversational participants must interact with each and with the surrounding environment. The amount of data that can be practically collected from a single participant is thus relatively limited which together with a high variability of both linguistic and perceptual information makes situated dialogue an *under-resourced task* in natural language processing.

One possibility taken in constructing corpora is to automatically generate the linguistic dataset from a set of rules (Das et al., 2018) but this has distinct limitations as the scope of language and vision that these rules cover is highly limited and the referring accuracy of these descriptions is unreliable (Aruqi, 2021). Several task-tailored solutions have been developed to collect multi-modal data with targeted participants either in real, e.g. (Dobnik, 2009), or virtual environments, e.g. (Stoia et al., 2008). There are also general solutions, most notably Dialogue Experimental Toolkit or DiET (Healey et al., 2003). This can be run on two or more participant computers where the clients are connected to a server where dialogue tasks are configured and recorded. The user interface resembles online messaging applications, with a chat window in which participants view the unfolding dialogue and a typing window in which participants can type and correct their contributions. In order to extend the pool of participants, over the last 10 years

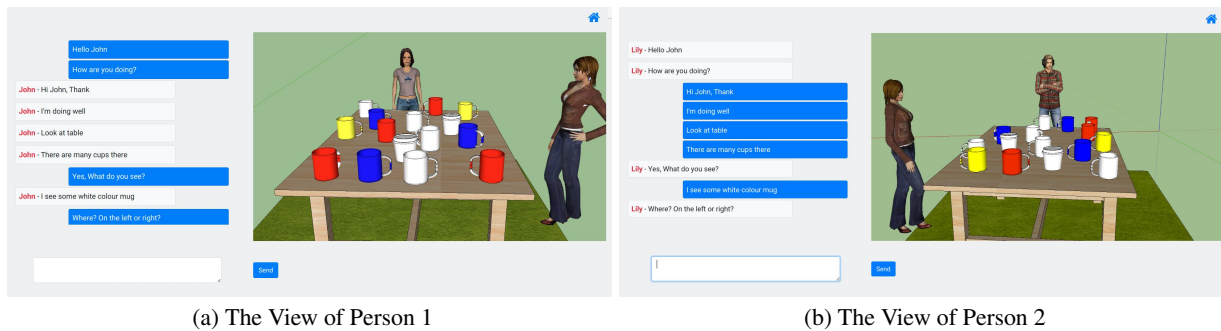
researchers have turned to online crowd-sourcing such as Amazon Mechanical Turk which has shown to produce good quality data based on completion rate and passing manipulation checks (Ipeirotis, 2011).

The crowd-sourcing platforms have been developed with the aim of collecting information from a single person at a time which is presented as a list of choices and input text. Since crowd-workers are normally paid by each input, time is an important issue for them, and they will try to accomplish the task as quickly as possible. Crowd-sourcing has been successfully used in research in computational linguistics, robotics and computer vision for tasks such as describing, annotation and providing speaker judgements. In order to set-up a task and the resulting data it is important to understand other tasks on the platform as well as the demographics and experiences of workers (Hitlin, 2016). Situated dialogue represents several additional changes to the standard crowd-sourcing setup. The first one is matching participants in real time and engage them in longer persistent conversations. The second is connecting visual environments with the textual chat interface. Several tools have been developed already that offer this kind of functionality, for example (Chernova et al., 2011; Manuvinakurike and DeVault, 2015; Das et al., 2017; Schlangen et al., 2018).

## 2 Semant-o-matic chat

We present a demo of an online chat tool that we have developed to collect conversational data from situated dialogue called Semant-o-matic chat or *Chat-o-matic*<sup>1</sup>. This allows us to setup data collection experiments both with targeted participants and with online crowd-sourcing platforms. The main motivation for developing our own tool rather than using existing solutions was to extend our

<sup>1</sup><https://www.dobnik.net/experiments/chat-o-matic>



(a) The View of Person 1

(b) The View of Person 2

Figure 1: View of the chat interface for different participants. Each participant is presented with a different view of the situated scene.

existing *Semant-o-matic* tool which we used for crowd-sourced collection of both linguistic (Rajestari et al., 2021) and situated data (Dobnik and Åstbom, 2017). The additional functionality includes the ability for participants to find each other online, engage in dialogue and presenting them with different scene views as shown in Figure 1.

## 2.1 Database

To make the tool more efficient and easier for researchers to analyse the data, we use a SQL database for storing all data. The table format is easy to understand and provides an organised and structural way to represent information (Davies, 2005). Additional features can be easily added in the future and the collected data can be searched and modified using SQL queries and filters which allow reference to data across columns and tables. In addition to conversations non-linguistic information such as answers from participants about their background, time during which the conversation took place and timings between turns can also be recorded.

## 2.2 Connecting with Mechanical Turk

The chat tool can be used openly by participants who can sign up for an account and then initiate a conversation with one of other users who are currently online. Participants can be attracted through target advertising of the task. Once initiated, the conversation is persistent and the participants can continue at any time later while keeping their participant roles (see Figure 1). We are currently testing the tool with AMT where we are using the same method as with the existing *Semant-o-matic* tool described in (Rajestari et al., 2021). A limited number of hits is published frequently to ensure that workers find our task. When signing up for the

task, participants are issued with a unique id which becomes their username on *Chat-o-matic*. Participants are paid as the tasks are published per each turn after these are checked for quality.

## 2.3 The task

While *Semant-o-matic chat* can be setup for any situated dialogue task, it is currently setup for the data collection of the Cups corpus described in (Dobnik and Silfversparre, 2021; Dobnik et al., 2020) which was previously collected with targeted participants using the DiET chat tool. The goal of this task is to study referring over longer stretches of situated dialogue where each participants has a different view of a slightly divergent scene: some cups are missing for each but these can be seen by the other. We use typing indicators to keep user attention and provide real-time feedback that replicates the feelings and cadence of an in-person conversation. This also helps users to better manage their turn-taking. Participants start with a blank view and the view of Person 1 is assigned to the one making the first turn.

## 3 Conclusion and future development

Our ongoing efforts are focused on testing the tool on AMT. We are also developing functionality that will help us to overview dialogues in a web interface (without the need to directly issue SQL queries) and automatically calculate the amount of payment for the reviewed dialogues. The code of the tool will be released as open source. We intend to extend the tool to new tasks of situated dialogue and adding functionality such as recording of key-strokes and corrections through JavaScript and usage of dynamic multi-modal data such as sound and video.

## References

- Ali Aruqi. 2021. [Embodied question answering in robotic environment: Automatic generation of a synthetic question-answer data-set](#). Masters in language technology (mlt), 30 hec, Department of Philosophy, Linguistics and Theory of Science (FLOV), University of Gothenburg, Gothenburg, Sweden. Supervisor: Simon Dobnik and Nikolai Ilinykh, examiner: Staffan Larsson, opponent: Catherine Vilorja.
- Sonia Chernova, Nick DePalma, and Cynthia Breazeal. 2011. [Crowd-sourcing real-world human-robot dialogue and teamwork through online multiplayer games](#). *AI Magazine*, 32(4):100–111.
- Abhishek Das, Samyak Datta, Georgia Gkioxari, Stefan Lee, Devi Parikh, and Dhruv Batra. 2018. [Embodied question answering](#). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2054–2063.
- Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José MF Moura, Devi Parikh, and Dhruv Batra. 2017. [Visual dialog](#). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 326–335.
- Mark Davies. 2005. [The advantage of using relational databases for large corpora: Speed, advanced queries, and unlimited annotation](#). *International Journal of Corpus Linguistics*, 10(3):307–334.
- Simon Dobnik. 2009. [Teaching mobile robots to use spatial words](#). Ph.D. thesis, University of Oxford: Faculty of Linguistics, Philology and Phonetics and The Queen’s College, Oxford, United Kingdom.
- Simon Dobnik and Amelie Åstbom. 2017. [\(Perceptual\) grounding as interaction](#). In *Proceedings of Saardial – Semdial 2017: The 21st Workshop on the Semantics and Pragmatics of Dialogue*, pages 17–26, Saarbrücken, Germany.
- Simon Dobnik, John D. Kelleher, and Christine Howes. 2020. [Local alignment of frame of reference assignment in English and Swedish dialogue](#). In *Spatial Cognition XII: Proceedings of the 12th International Conference, Spatial Cognition 2020, Riga, Latvia*, pages 251–267, Cham, Switzerland. Springer International Publishing.
- Simon Dobnik and Vera Silfversparre. 2021. [The red cup on the left: Reference, coreference and attention in visual dialogue](#). In *Proceedings of PotsDial - Semdial 2021: The 25th Workshop on the Semantics and Pragmatics of Dialogue*, Proceedings (SemDial), pages 50–60, Potsdam, Germany.
- Patrick G. T. Healey, Matthew Purver, James King, Jonathan Ginzburg, and Greg J. Mills. 2003. [Experimenting with clarification in dialogue](#). In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, volume 25, pages 539–544, Boston, MA.
- Paul Hitlin. 2016. [Research in the crowdsourcing age: A case study](#). Technical report, Pew Research Center, Pew Research Center.
- Panos Ipeirotis. 2011. [Crowdsourcing using Mechanical Turk: quality management and scalability](#). Talk, Stern School of Business, New York University.
- Ramesh Manuvinaurike and David DeVault. 2015. [Pair me up: A web framework for crowd-sourced spoken dialogue collection](#). In *Natural Language Dialog Systems and Intelligent Assistants*, pages 189–201. Springer.
- Maryam Rajestari, Simon Dobnik, Robin Cooper, and Aram Karimi. 2021. [Very necessary: the meaning of non-gradable modal adjectives in discourse contexts](#). In *Selected contributions from the Eighth Swedish Language Technology Conference (SLTC-2020), 25-27 November 2020*, volume 184 of *NEALT Proceedings Series, No. XX*, pages 50–58, Linköping, Sweden. Northern European Association for Language Technology (NEALT), Linköping University Electronic Press: Linköping Electronic Conference Proceedings.
- David Schlangen, Tim Diekmann, Nikolai Ilinykh, and Sina Zarriß. 2018. [slurk – a lightweight interaction server for dialogue experiments and data collection](#). In *Proceedings of the 22nd Workshop on the Semantics and Pragmatics of Dialogue - Poster Abstracts*, Aix-en-Provence, France. SEMDIAL.
- Laura Stoia, Darla Magdalena Shockley, Donna K. Byron, and Eric Fosler-Lussier. 2008. [SCARE: a situated corpus with annotated referring expressions](#). In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08)*, pages 650–653, Marrakech, Morocco. European Language Resources Association (ELRA).