Lucas Druart

Orange
2 Avevnue de Belle Fontaine
35510 Cesson-Sévigné
France

lucas1.druart@orange.com
lucas.druart@alumni.univ-avignon.fr
https://lucasdruart.github.io/

1 Research interests

Task-Oriented Dialogue (TOD) systems provide interactive assistance to a user in order to accomplish a specific task such as making a reservation at a restaurant or booking a room in a hotel.

Speech presents itself as a natural interface for TOD systems. A typical approach to implement them is to use a modular architecture (Gao et al., 2018). A core component of such dialogue systems is Spoken Language Understanding (SLU) whose goal is to extract the relevant information from the user's utterances. While spoken dialogue was the focus of earlier work (Williams et al., 2013; Henderson et al., 2014), recent work has focused on text inputs with no regard for the specificities of spoken language (Wu et al., 2019; Heck et al., 2020; Feng et al., 2021). However, this approach fails to account for the differences between written and spoken language (Faruqui and Hakkani-Tür, 2022) such as disfluencies.

My research focuses on **Spoken Language Understanding** in the context of **Task-Oriented Dialogue**. More specifically I am interested in the two following research directions:

- **Annotation** schema for **spoken** TODs,
- Propagation of **dialogue history** for **contextually coherent** predictions.

1.1 Annotation schema for spoken TODs

Chat TODs corpora benefit from a wide diversity of semantic annotation schema which have different levels of precision. The Slot-Value scheme is probably the most commonly used one, such as for the Dialogue State Tracking (DST) annotations of Multi-Woz (Budzianowski et al., 2018). However this scheme lacks grounding (*e.g.* two mentions of the same entity are seen as two separate values) which is a fundamental aspect of human-human interactions (Benotti and Blackburn, 2021). It also does not provide dynamic links between the mentioned entities[1] which can be essential to co-reference resolution. For instance, when a user booking a hotel room refers to the previously mentioned hotel by its address such as "I would like to book the one in Prague.", the link between a hotel and its address becomes essential.

More recent schema such as Dialogue-AMR (Bonial et al., 2020) and Dialogue Meaning Representation (DMR) (Hu et al., 2022) address these shortcomings by relying on the same mechanisms as Abstract Meaning Representation (AMR) (Banarescu et al., 2013). Users of spoken dialogue system (SDS) tend to refer to previously mentioned entities by their characteristics which change over the course of the dialogue. The hierarchical relations defined in DMR track those relations thus enabling direct disambiguation. Dialogue-AMR further maps it to specific robotic controls.

I believe such rich annotation scheme will help address spoken TODs specificities and I am currently working on such a scheme for the MEDIA spoken TOD dataset (Devillers et al., 2004).

1.2 Dialogue history propagation

For TOD systems to help users accomplish complex tasks, such as choosing the most relevant hotel to a user requirements, it must take into account the information provided in previous turns. Dialogue history is thus crucial information for contextually accurate and consistent predictions. However it remains unclear how to propagate such context in a spoken dialogue understanding model's predictions.

During the recent Speech Aware Dialogue Systems Technology Challenge (Soltau et al., 2022) all proposed systems aggregated the dialogue history once transcribed, including our system (Jacqmin et al., 2023) which ranked first. End-to-End models, which benefit from joint-optimization, require more sophisticated mechanism to limit the input size.

[1] Note that definition of slot types often imply some static relations between the slots. For instance in Multi-Woz DST annotations slots are grouped by domain.

Proceedings of the 19th Annual Meeting of the Young Reseachers' Roundtable on Spoken Dialogue Systems, pages 37–39
September 11–12, 2023.

I am currently exploring different fusion strategies between a textual semantic context and audio extracted features (*e.g.* two cross attention modules each attending to an encoder, modality fusion before the decoder).

2 SDS research

The Spoken Dialogue System field is moving at an incredibly fast pace and the gap between research and deployment is narrowing. Therefore I believe future research will have to rely on more realistic datasets.

- Such datasets should provide a database of entities given that all errors do not lead to a wrong entity matching.
- The very interactive nature of SDS implies that a misunderstanding at a given turn can change how the next turns unfold. I believe future datasets should be dynamic and provide several continuations at each turn. This will enable researchers to measure which misunderstandings lead to poorer dialogue trajectories.
- Some chat corpora have been vocalized to benefit from the large quantity of data of such corpora. However SDS research also requires natural speech datasets to take into account the specific interactions (*e.g.* confirmations, repetitions, turn taking) of spoken dialogues.

Finally evaluating the impact of SDS components on the completion of the targeted task seems to be a promising and mandatory research path.

3 Suggested topics for discussion

In a broader discussion I believe the following topics might be interesting to discuss:

- While generative models are displaying impressive capacities they may not provide a reliable and consistent behavior. For instance when SDS are connected to APIs, it becomes essential to include some control over the inputs of the APIs. Hence I believe discussing techniques to secure the use of such models in SDS might prove helpful.
- Prompting techniques are being widely adopted, however we have only little understanding of how they work. I believe sharing our experience and knowledge of prompting can provide indications of what seems to happen internally with prompts. For instance, one might wonder if any type of information (*e.g.* structured, audio, image) can be passed as a prompt.
- Finally I believe SDS research should take into account its ethical implications such as the greenhouse gas emission burden of deep learning or the anthropomorphic relation users tend to develop with dialogue systems. Investigating computing wise efficiency and human computer interaction in the context of SDS might help move forward in both directions.

References

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*. Association for Computational Linguistics, Sofia, Bulgaria, pages 178–186. https://aclanthology.org/W13-2322.

Luciana Benotti and Patrick Blackburn. 2021. Grounding as a collaborative process. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Association for Computational Linguistics, Online, pages 515–531. https://doi.org/10.18653/v1/2021.eacl-main.41.

Claire Bonial, Lucia Donatelli, Mitchell Abrams, Stephanie M. Lukin, Stephen Tratz, Matthew Marge, Ron Artstein, David Traum, and Clare Voss. 2020. Dialogue-AMR: Abstract Meaning Representation for dialogue. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pages 684–695. https://aclanthology.org/2020.lrec-1.86.

Pawe\l Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, pages 5016–5026. https://doi.org/10.18653/v1/D18-1547.

Laurence Devillers, Hélène Bonneau-Maynard, Sophie Rosset, Patrick Paroubek, Kevin McTait, Djamel Mostefa, Khalid Choukri, Laurent Charnay, Caroline Bousquet-Vernhettes, Nadine Vigouroux, Frédéric Béchet, Laurent Romary, Jean-Yves Antoine, Jeanne Villaneau, Myriam Vergnes, and Jérôme Goulian. 2004. The french media/evalda project: the evaluation of the understanding capability of spoken language dialogue systems. In *International Conference on Language Resources and Evaluation*.

Manaal Faruqui and Dilek Hakkani-Tür. 2022. Revisiting the boundary between ASR and NLU in the age of conversational dialog systems. *Computational Linguistics* 48(1):221–232. https://doi.org/10.1162/coli_a_00430.

Yue Feng, Yang Wang, and Hang Li. 2021. A Sequence-to-Sequence Approach to Dialogue State Tracking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, pages 1714–1725. https://doi.org/10.18653/v1/2021.acl-long.135.

Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural Approaches to Conversational AI. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, SIGIR '18, pages 1371–1374. https://doi.org/10.1145/3209978.3210183.

Michael Heck, Carel van Niekerk, Nurul Lubis, Christian Geishauser, Hsien-Chin Lin, Marco Moresi, and Milica Gasic. 2020. TripPy: A Triple Copy Strategy for Value Independent Neural Dialog State Tracking. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*. Association for Computational Linguistics, 1st virtual meeting, pages 35–44. https://aclanthology.org/2020.sigdial-1.4.

Matthew Henderson, Blaise Thomson, and Jason D. Williams. 2014. The second dialog state tracking challenge. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*. Association for Computational Linguistics, Philadelphia, PA, U.S.A., pages 263–272. https://doi.org/10.3115/v1/W14-4337.

Xiangkun Hu, Junqi Dai, Hang Yan, Yi Zhang, Qipeng Guo, Xipeng Qiu, and Zheng Zhang. 2022. Dialogue meaning representation for task-oriented dialogue systems. In *Findings of the Association for Computational Linguistics: EMNLP 2022*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, pages 223–237. https://aclanthology.org/2022.findings-emnlp.17.

Léo Jacqmin, Lucas Druart, Valentin Vielzeuf, Lina Maria Rojas-Barahona, Yannick Estève, and Benoît Favre. 2023. Olisia: a cascade system for spoken dialogue state tracking.

Hagen Soltau, Izhak Shafran, Mingqiu Wang, Abhinav Rastogi, Jeffrey Zhao, Ye Jia, Wei Han, Yuan Cao, and Aramys Miranda. 2022. Speech aware dialog system technology challenge (dstc11).

Jason Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. 2013. The Dialog State Tracking Challenge. In *Proceedings of the SIGDIAL 2013 Conference*. Association for Computational Linguistics, Metz, France, pages 404–413. https://aclanthology.org/W13-4065.

Chien-Sheng Wu, Andrea Madotto, Ehsan Hosseini-Asl, Caiming Xiong, Richard Socher, and Pascale Fung. 2019. Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, pages 808–819. https://doi.org/10.18653/v1/P19-1078.

Biographical sketch

Lucas Druart is a PhD student working jointly with Orange Labs and the Speech and Language Team of the University of Avignon. Previously he graduated from a double degree in applied mathematics and computer science, with a focus on Statistics and Machine Learning, from Grenoble-INP Ensimag and University of Grenoble Alpes (UGA). His current research interests include Natural Language Processing, Spoken Language Understanding and Task-Oriented Dialogue.