Dialogue and Dialogue Systems

Liesbeth Degand, Philippe Muller

Introduction to the Special Issue on Dialogue and Dialogue Systems

Liesbeth Degand* — **Philippe Muller****

* *Institute for Language and Communication, University of Louvain*
** *IRIT, University of Toulouse*

1. An expanding field with new questions

Recent progress in the field of NLP impacts all of its subfields and extends its domain of applications. Central to these developments is automated dialogue, either with chatbots (scripted, conversational, cognitive) or personal assistants, ubiquitous and widely distributed as services via smartphones or commercial websites.

At the same time, new written media continue to grow and most of them involve some sort of interaction: chats, forums, emails, microblogging, or collaborative instant messaging services. This progress and the generalization of natural language for interaction also make way for novel approaches taking into account multiple modalities (images, video) and the situation in which dialogues take place. In this context, all aspects of conversation analysis and dialogue system development are concerned, whether communication is oral or written, task-oriented or open.

The prevalence of neural network based approaches in NLP further influences approaches to dialogue, bringing into focus different matters, such as automatic generation of diverse and natural responses, although these approaches sometimes minimize the role of comprehension, and make it quite challenging to integrate linguistic and extra-linguistic context. New approaches also bring new problems that pervade machine-learning in general: black box models are not easy to explain, and blurring the lines between humans and machines may generate ethical quandaries—such as when conversational agents reproduce the biases and prejudices present in their training data—issues that generate a lot of concern and dedicated workshops.[1] Paucity

1. Such as the *Safety for Conversational AI Workshop*, `safetyforconvai.splashthat.com`.

of annotated data for supervised methods is another issue for machine learning approaches which motivates the collection of auxiliary data, semi-automatically annotated data, or artificial data far removed from real-world use cases. This can generate a gap between theory and practice that needs to be addressed.

For this special issue, the *TAL journal* invited contributions on all aspects of research related to the analysis of written or transcribed conversations, to the development and evaluation of dialogue systems, to data collection for all interaction modalities, and to ethical and social issues pertaining to dialogue and its applications. While not all of these aspects have been addressed in the contributions to this special issue, we will quickly review in this introduction the more important lines of research in Natural Language Processing and Computational Linguistics, both from the point of view of natural language conversation and dialogue analysis, and from that of the perspective of dialogue systems that interact with users. We then present this issue's accepted papers.

2. Analysis of dialogues

Computational analysis of linguistic interaction is not a new topic, as it went hand in hand with progress in speech recognition, which gave rise to interactive applications, and a demand for a more complex understanding of conversations. Another shift has taken place with the increase of written interactions that came with the widespread use of the internet: forums, chat rooms, microblogging, and emails are all manifestations of more or less synchronous dialogues involving two or more participants. This has provided a boost in available data as the volume of written conversations largely surpasses available speech data, and they are also much easier to process. This is then reflected in the growth of annotated corpora, for instance the Ubuntu IRC corpus (Kummerfeld *et al.*, 2019) which includes relations between utterances in a multiparty technical chat discussion, or the STAC corpus (Asher *et al.*, 2016), consisting of chat negotiation dialogues with annotated discourse and conversation structures. In parallel, a rising number of (corpus) linguistic studies have engaged in trying to uncover the linguistic specificities of "online discourse". Thus, according to Baron (2010), a "persistent question intriguing Internet researchers has been whether the stylistic features of CMC [computer-mediated communication] are more like those of informal speech or paradigmatic writing". In this strand of research, interactive text-based computer-mediated conversations have been shown to share many characteristics with informal spoken conversations, especially in terms of conversational and discourse mechanisms such as turn-taking, grounding, and coherence marking.

While the bridge between linguistic descriptive (corpus-based) research and computational work is not yet fully crossed, some of these linguistic findings have found their way to NLP research applications: identifying speech acts (Mohiuddin *et al.*, 2019), analyzing the structure of interactions (Shi *et al.*, 2019; Badene *et al.*, 2019), understanding the flow of information in context, among others. Other questions might be more relevant for oral data (speaker recognition for instance), or

show up in different ways in various types of interactions (disfluencies, monitoring, feedback). Written-text oriented models or annotation standards tend now to be more concerned with integrating oral phenomena, as is shown by the new Universal dependencies model for syntax (Sanguinetti *et al.*, 2020). New interesting problems also appear: with more speakers/writers involved, overlapping threads of conversation complicate understanding (Kummerfeld *et al.*, 2019).

The current (industrial) applications are numerous: for instance, a lot of companies provide chat interaction for Customer Relationship Management instead of telephone services, and are interested in the information they can thus gain about their customers in order to improve their interaction with them. Another example is the analysis of meetings and producing the minutes, which was first undertaken with the AMI corpus (Carletta, 2007) and the ICSI corpus (Janin *et al.*, 2003), and is now being revisited, also with progress made in automated summarization (Li *et al.*, 2019).

On a technical level, neural networks have enabled powerful intermediate, learnable representations at each level of a conversation: tokens, utterances and speech acts, and sequences of such, which pervade the predictive models applied to conversation aspects (Kumar *et al.*, 2018).

3. Dialogue systems

Of course, theoretical advances in dialogue models impact dialogue systems, but interactive dialogue systems include more than just the modelling of interactions. They generally consist of three main components: (1) understanding user input (either from speech or written text); (2) managing the interaction: keeping track of the dialogue state and planning actions; and (3) generating a linguistic output form such as text or speech to interact with the user.

The natural language understanding component is more directly tied to progress in dialogue modelling, but within dialogue systems it is usually tailored to a specific application, focusing on a particular part of the input or a classification of the goal of a speech act (sometimes called *intents*), with the goal of extracting predefined pieces of information relevant to the intent (*slot filling*), a task akin to semantic parsing in more general NLP, cf. the survey of recent work on these aspects by Louvan and Magnini (2020). There is also a lot of interest in so-called open-domain systems, which would have the ability to naturally interact with human participants and converse on any subject, but they raise a lot more issues that only partly concern applied systems: background knowledge, and emotional engagement (Huang *et al.*, 2020).

At the level of dialogue management, one can distinguish two tasks that together define the behaviour of the system and the way it responds to users: (a) dialog state tracking (DST), and (b) determining an optimal dialog policy (what is the best move in a given context). DST covers all models and representations of speaker and agent beliefs, goals, and the state of conversation (questions under discussion, common ground, commitments), see a survey by Williams *et al.* (2016). It is the subject of

an ongoing series of annual challenges[2], with varying subtasks. Recent approaches involve neural networks for their flexibility with respect to integrating different levels of representations (Mrkšić *et al.*, 2017). Dialog policy optimization is the planning and/or selection of dialogue actions, and their types and content, before generation of the linguistic output. Technical approaches have been based on reinforcement learning, first with partially observable Markov decision problems (Young *et al.*, 2013), superseded now by deep RL approaches, for instance Li *et al.* (2017).

These components can also be integrated into a single architecture, especially in neural models that try to enforce an end-to-end architecture, from user language input to system output, with intermediate latent representations for dialogue states and for the policy to follow, all supervised by the end result of the interaction with respect to the task. In this case the tasks are simple and focused, for example in the case of interactive question answering (Dhingra *et al.*, 2017; Wen *et al.*, 2017). Answering a user is then either based on dedicated Natural Language Generation (NLG) approaches, which in general is the problem of taking a formal representation to produce a well-formed linguistic output. As mentioned above, this can also be integrated in a general architecture in which it is the final output. In fact, recent work tends to develop end-to-end architectures, where the only input is the user's last utterance(s), in so-called sequence-to-sequence neural models. These are often trained on existing dialogue corpora or social media exchanges, where there is no clear "task" or objective (Zhao *et al.*, 2017; Chan *et al.*, 2019). The recent survey by Dusek *et al.* (2020) underlines the difficulty of evaluating end-to-end NLG systems; and while seq2seq models give excellent results on some metrics (naturalness), they also can lack in semantic fidelity and diversity in their outputs.

Evaluation is an important issue in dialogue systems in general, and a complicated one, as systems involve different components and different objectives, as shown by Deriu *et al.* (2021). There is thus a variety of automatic metrics and human evaluation procedures specific to the different subtasks mentioned above.

The variety of approaches and applications can sometimes make generalization difficult from one domain to another, between tasks and contexts of use. This is also reflected in the richness and diversity of available data useful for designing systems, and a good view is given in Serban *et al.* (2018). The prevalence of data-driven models also means there is less *a priori* control on the behaviour of the systems, which can lead to undesirable outcomes and raise ethical questions (Henderson *et al.*, 2018).

2. `https://dstc9.dstc.community/past-challenges`.

4. Papers

4.1. *Dialogue management with linear logic: the role of metavariables in questions and clarifications*

The paper by Maraev, Bernardy and Ginzburg focuses on the dialogue management component of dialogue systems: they are concerned by the modelling of dialogue states, consisting here in several elements: a set of *questions under discussion*, recording unresolved interactions (mostly questions waiting for an answer), the history of speech acts and their content, as interpreted by the system. The manager is also supposed to take care of an *agenda* of planned moves by the system. They focus on a type of interaction that is common in information-seeking conversations: question and answers, including embedded sequences of questions when *clarification questions* occur.

The model they present uses formal representations for these elements, and proposes to characterize updates of the dialogue state and its agenda as proof derivations in a linear logic, in which dialogue possible operations are linear logic formulas: processing a question, processing a potential answer, generating a clarification question when no unambiguous answer exists in the system database.

This gives an interesting formal framework to understand these dialogue moves, and gives a blueprint for a rule-based dialogue manager, as they implemented a prototype of the dialogue acts covered in the paper.

While rule-based formal approaches to dialog management were once very popular, they now tend to be in the shadow of probabilistic approaches, among them mostly neural approaches. There is nonetheless a growing awareness that the two kinds of approaches can benefit each other: empirical methods help achieving better coverage and robustness, while injecting knowledge in systems relying on supervised learning helps diminish the demands on huge amounts of data (Lison, 2015; Williams *et al.*, 2017).

4.2. *Situated meaning in multimodal dialogue: human-robot and human-computer interactions*

The paper by Pustejovsky & Krishnaswamy focuses on the important topic of “situated" conversation, where interaction between humans or between humans and a system takes into account the specific context of the interaction, including user location, and activities that connect users to each other and to their environment (e.g. a common task or a game), and non-linguistic interactions. A model of this kind of conversation must address different issues, most importantly linguistic references to the context (deixis), reasoning about the environment, tying the perceptions of agents and their actions to the conversation (Hunter *et al.*, 2018).

In order to study these sorts of interactions, the paper presents a system which provides a simulated physical environment for agent interactions and a few scripted

tasks involving the manipulation of objects. The system demonstrates the kind of knowledge that is needed to explain conversation moves and references to the situation in which the conversation takes place. To this end, it proposes an ontology for physical objects that makes explicit how they might be manipulated and discussed.

An important topic of the paper is how to address paralinguistic conversation, with gestures associated with speech acts: interpreting the other speaker's gestures jointly with the linguistic message. They thus provide a formal representation for the semantics of speech acts and accompanying gestures.

The interest of such a platform and associated models is two-fold: it provides a simulation environment to record situated conversations with a trace of the situation: knowledge of the environment, the geometry of objects and agents, and their gestures superimposed on the conversation content. Such a platform could prove useful in collecting rich conversational data. Furthermore, depending on the ease with which it might be configured, the platform could provide an environment and testable model for situated conversational agents.

4.3. *Comparaison linguistique et neuro-physiologique de conversations humain-humain et humain-robot*

The paper by Hallart, Maes, Spatola, Prévot and Chaminade addresses questions about the behaviour of speakers in a conversation, comparing a context involving an automated dialog agent (or perceived as such, in a Wizard-of-Oz experimental setup) and a more natural context with two humans. By observing various parameters, linguistic and conversation patterns, but also cognitive aspects through fMRI scans, experiments show how human behave differently when facing a perceived robot with limited linguistic capabilities, and how they adapt their linguistic behaviour to the agent. For instance, humans facing robots show a stronger lexical *alignment*, i.e. their vocabulary converges more towards the other conversant's vocabulary. In general, language complexity decreases when talking to a perceived robot.

Conversational aspects that are observed include speech time, interaction with the other speaker with feedback moves, specific discourse markers. Linguistic aspects are mostly related to language complexity: lexical, syntactic, and also *descriptive*, related to the use of adjectives and adverbs. The paper proposes or refines quantitative measures of those complexity types. An original aspect of this study is to compare the linguistic observations with fMRI scans to see what regions of the brain are more or less active during a conversation, and if differences can be identified when talking to another human or to a robot. In particular, lexical complexity seems correlated negatively with the activation of certain regions that would indicate cognitive resources (i.e. memory) are less mobilized when talking to a simple robot agent.

While it is only a perspective of the present work, studying linguistic and cognitive patterns of interaction raises potentially interesting lines of research to improve dialog agents and anticipate unforeseen or undesirable human reactions to a system.

Acknowledgements

We wish to thanks the editors-in-chief for their unfailing support, especially Sophie Rosset, the scientific committee for this special issue for their constructive work and reactivity: Nicholas Asher (IRIT, CNRS), Fred Bechet (Aix Marseille University), Christophe Cerisara (LORIA, CNRS), Nancy Chen (A*STAR, Singapore), Laurence Devillers (Paris Sorbonne University, LIMSI), Yannick Esteve (LIUM, University of Le Mans), Raquel Fernández (ILLC, University of Amsterdam), Kerstin Fischer (Hamburg University), Jonathan Ginzburg (Paris University), Casey Kennington (Boise State University), Nicolas Hernandez (LS2N, Nantes University), Julie Hunter (Linagora), Frédéric Landragin (CNRS, ENS), Pierre Lison (Norwegian Computing Center), Laurent Prévot (Aix Marseille University), Lina Maria Rojas Barahona (Orange Labs, Lannion), David Schlangen (Bielefeld University), Noël Nguyen Trong (Aix Marseille University), and finally Kate Thompson for proof-reading and suggestions.

5. References

Asher N., Hunter J., Morey M., Benamara F., Afantenos S. D., "Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus", *in* N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (eds), *Proceedings of the 10th International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016*, European Language Resources Association (ELRA), 2016.

Badene S., Thompson K., Lorré J.-P., Asher N., "Weak Supervision for Learning Discourse Structure", *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Association for Computational Linguistics, Hong Kong, China, p. 2296-2305, November, 2019.

Baron N. S., "Discourse structures in Instant Messaging: The case of utterance breaks", *Language@Internet*, 2010.

Carletta J., "Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus", *Lang. Resour. Evaluation*, vol. 41, n° 2, p. 181-190, 2007.

Chan Z., Li J., Yang X., Chen X., Hu W., Zhao D., Yan R., "Modeling Personalization in Continuous Space for Response Generation via Augmented Wasserstein Autoencoders", *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Association for Computational Linguistics, Hong Kong, China, p. 1931-1940, November, 2019.

Deriu J., Rodrigo Á., Otegi A., Echegoyen G., Rosset S., Agirre E., Cieliebak M., "Survey on evaluation methods for dialogue systems", *Artif. Intell. Rev.*, vol. 54, n° 1, p. 755-810, 2021.

Dhingra B., Li L., Li X., Gao J., Chen Y.-N., Ahmed F., Deng L., "Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access", *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long*

Papers), Association for Computational Linguistics, Vancouver, Canada, p. 484-495, July, 2017.

Dusek O., Novikova J., Rieser V., "Evaluating the state-of-the-art of End-to-End Natural Language Generation: The E2E NLG challenge", *Comput. Speech Lang.*, vol. 59, p. 123-156, 2020.

Henderson P., Sinha K., Angelard-Gontier N., Ke N. R., Fried G., Lowe R., Pineau J., "Ethical Challenges in Data-Driven Dialogue Systems", *in* J. Furman, G. E. Marchant, H. Price, F. Rossi (eds), *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2018, New Orleans, LA, USA, February 02-03, 2018*, ACM, p. 123-129, 2018.

Huang M., Zhu X., Gao J., "Challenges in Building Intelligent Open-domain Dialog Systems", *ACM Trans. Inf. Syst.*, vol. 38, n° 3, p. 21:1-21:32, 2020.

Hunter J., Asher N., Lascarides A., "A formal semantics for situated conversation", *Semantics and Pragmatics*, 2018.

Janin A., Baron D., Edwards J., Ellis D., Gelbart D., Morgan N., Peskin B., Pfau T., Shriberg E., Stolcke A., Wooters C., "The ICSI Meeting Corpus", *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '03, Hong Kong, April 6-10, 2003*, IEEE, p. 364-367, 2003.

Kumar H., Agarwal A., Dasgupta R., Joshi S., "Dialogue Act Sequence Labeling Using Hierarchical Encoder With CRF", *in* S. A. McIlraith, K. Q. Weinberger (eds), *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, AAAI Press, p. 3440-3447, 2018.

Kummerfeld J. K., Gouravajhala S. R., Peper J. J., Athreya V., Gunasekara C., Ganhotra J., Patel S. S., Polymenakos L. C., Lasecki W., "A Large-Scale Corpus for Conversation Disentanglement", *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Florence, Italy, p. 3846-3856, July, 2019.

Li M., Zhang L., Ji H., Radke R. J., "Keep Meeting Summaries on Topic: Abstractive Multi-Modal Meeting Summarization", *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Florence, Italy, p. 2190-2196, July, 2019.

Li X., Chen Y.-N., Li L., Gao J., Celikyilmaz A., "End-to-End Task-Completion Neural Dialogue Systems", *Proceedings of the 8th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Asian Federation of Natural Language Processing, Taipei, Taiwan, p. 733-743, November, 2017.

Lison P., "A hybrid approach to dialogue management based on probabilistic rules", *Comput. Speech Lang.*, vol. 34, n° 1, p. 232-255, 2015.

Louvan S., Magnini B., "Recent Neural Methods on Slot Filling and Intent Classification for Task-Oriented Dialogue Systems: A Survey", *in* D. Scott, N. Bel, C. Zong (eds), *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, International Committee on Computational Linguistics, p. 480-496, 2020.

Mohiuddin T., Nguyen T.-T., Joty S., "Adaptation of Hierarchical Structured Models for Speech Act Recognition in Asynchronous Conversation", *Proceedings of the 2019 Conference of*

the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (Volume 1: Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, p. 1326-1336, June, 2019.

Mrkšić N., Ó Séaghdha D., Wen T.-H., Thomson B., Young S., "Neural Belief Tracker: Data-Driven Dialogue State Tracking", *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Vancouver, Canada, p. 1777-1788, July, 2017.

Sanguinetti M., Bosco C., Cassidy L., Çetinoğlu Ö., Cignarella A. T., Lynn T., Rehbein I., Ruppenhofer J., Seddah D., Zeldes A., "Treebanking User-Generated Content: A Proposal for a Unified Representation in Universal Dependencies", *Proceedings of the 12th Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, p. 5240-5250, May, 2020.

Serban I. V., Lowe R., Henderson P., Charlin L., Pineau J., "A Survey of Available Corpora For Building Data-Driven Dialogue Systems: The Journal Version", *Dialogue Discourse*, vol. 9, n° 1, p. 1-49, 2018.

Shi W., Zhao T., Yu Z., "Unsupervised Dialog Structure Learning", *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (Volume 1: Long and Short Papers)*, Association for Computational Linguistics, Minneapolis, Minnesota, p. 1797-1807, June, 2019.

Wen T.-H., Vandyke D., Mrkšić N., Gašić M., Rojas-Barahona L. M., Su P.-H., Ultes S., Young S., "A Network-based End-to-End Trainable Task-oriented Dialogue System", *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Valencia, Spain, p. 438-449, April, 2017.

Williams J. D., Asadi K., Zweig G., "Hybrid Code Networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning", *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Vancouver, Canada, p. 665-677, July, 2017.

Williams J. D., Raux A., Henderson M., "The Dialog State Tracking Challenge Series: A Review", *Dialogue Discourse*, vol. 7, n° 3, p. 4-33, 2016.

Young S., Gašić M., Thomson B., Williams J. D., "POMDP-Based Statistical Spoken Dialog Systems: A Review", *Proceedings of the IEEE*, vol. 101, n° 5, p. 1160-1179, 2013.

Zhao T., Zhao R., Eskenazi M., "Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders", *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Vancouver, Canada, p. 654-664, July, 2017.