GraphDial:
Graph-based Neural Models for Dialogue Management

1 Excellence

1.1 State of the art, knowledge needs and project objectives

Spoken language is a natural form of communication for human beings. Since early childhood, we have all learned how to understand and produce speech in order to interact with one another, and a large part of our waking life is spent in social interactions mediated through natural language. In many ways, the human brain is "wired" for spoken dialogue. Although the potential of speech for human-computer interfaces was long neglected, this is fortunately changing rapidly due to the rising importance of virtual assistants (such as Siri, Cortana or Google Home), ambient computing devices (such as Amazon's Echo), in-car voice control and human–robot interaction (HRI).

These examples of *spoken dialogue systems* incorporate several processing steps, as illustrated in Figure 1. The core of the dialogue system is the *dialogue state*, whose aim is to represent all aspects of the interaction that are relevant for the system to decide what do or say at a given time. The process of updating the dialogue state with new observations (e.g. new user utterances) is called *dialogue state tracking*, while the subsequent process of selecting the next system action (responding to the user or executing a non-verbal action) is called *action selection*.

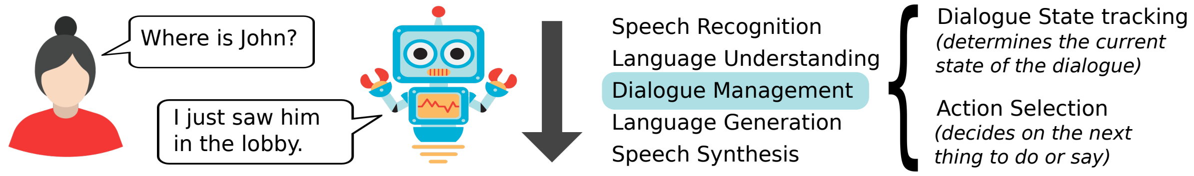

Figure 1: Simple view of a spoken dialogue system, focusing on the role of dialogue management.

These steps are often implemented with data-driven models trained on dialogue corpora using supervised or reinforcement learning. In particular, deep neural networks have become increasingly popular due to their ability to be trained end-to-end and infer latent representations of the dialogue, leading to enhanced performance across a wide range of dialogue tasks. Compared to handcrafted methods, neural models have been shown to be relatively robust to noise, uncertainty, and (context- and user-dependent) variations in conversational patterns (Gao et al., 2019).

However, current neural models of dialogue do suffer from a number of shortcomings:

- The first limitation comes from their representation of the *dialogue state*. In most approaches, this dialogue state is expressed as a predefined set of state variables[1] or as a fixed-length vector of latent features learned from context-response pairs. In both cases, the dialogue state is constrained to a flat, fixed-size representation, making it difficult to capture rich conversational contexts that may include varying numbers of entities to track.

- Another important challenge is that neural dialogue models are typically dependent on large amounts of dialogue data in order to learn useful dialogue behaviours. This requirement is problematic for application domains where data is scarce and expensive to obtain.

The **GraphDial** project will investigate an alternative approach to dialogue management based on the use of (probabilistic) *graphs* as core representation for the dialogue state. Graphs are well suited to capture rich interaction settings encompassing multiple entities and relations. Figure 2 illustrates

[1] For instance, an automated flight booking system will have a list of predefined slots (each with a range of possible values) that the system must fill by interacting with the human user, such as "Departure", "Arrival", "Date", etc.

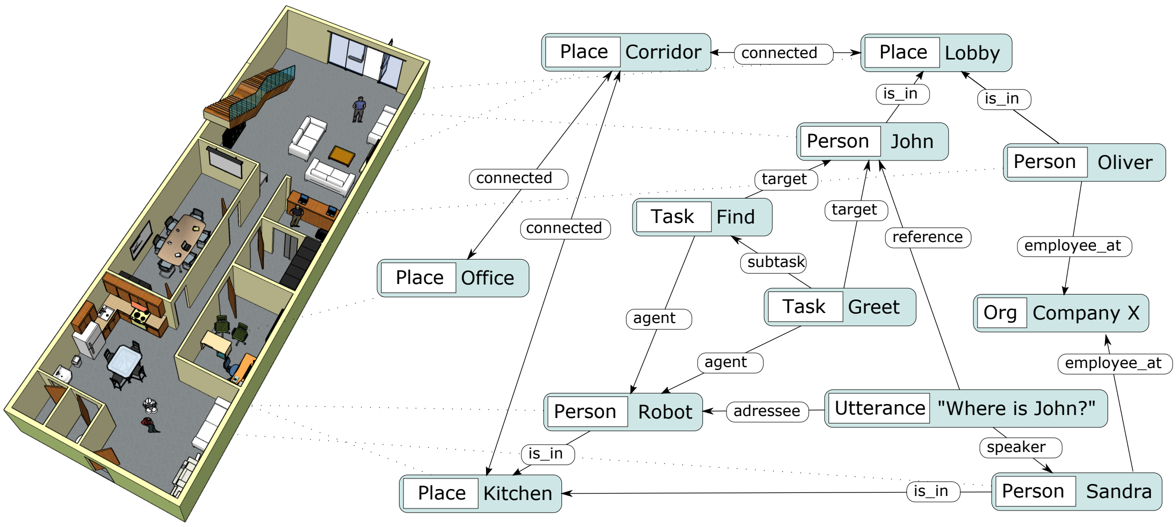

Figure 2: Fragment of graph representation in a human–robot interaction setting. The graph connects various concrete or abstract entities such as places, persons, tasks and utterances. A full graph would also include attributes for each node (for instance physical coordinates) along with probabilities for each entity, relation and attribute (to account for various uncertainties).

this relational structure in the context of human–robot interaction. Such complex structures are difficult to express in a fixed-size representation, as the entities and relations are not known in advance and may change over time. Another key advantage of graph representations is that they facilitate the use of *relational abstractions* covering large portions of the state space in a compact and human-readable manner (Getoor and Taskar, 2007; Battaglia et al., 2018). For instance, the logical constraint $\forall x, y, z : Person(x) \wedge is\_in(x, y) \wedge separate(y, z) \rightarrow \neg\, is\_in(x, z)$ states that a person cannot be present in two separate places at the same time. Such abstraction mechanisms are particularly important when estimating machine learning models from limited amounts of data (Lison, 2015).

Another central research topic for the **GraphDial** project is the use of *weak supervision* to train neural models for dialogue state tracking and action selection. Weak supervision is an emerging paradigm designed to provide machine learning models with indirect training data extracted from (possibly noisy) heuristic labelling functions or domain knowledge. The use of weak supervision is particularly attractive for dialogue management due to the difficulty of obtaining direct annotated data in many dialogue domains. The project will investigate a range of weak supervision signals, including grounding responses, heuristic rules and global constraints on the graph structure.

To achieve these objectives, the **GraphDial** project will feature strong international collaborations with leading researchers in the field of spoken dialogue systems, statistical relational learning, graph neural networks and human–robot interaction. The project will also be in contact with two Norwegian companies (Convertelligence and No Isolation) involved in the development of conversational AI technology to facilitate the dissemination of the project's research results.

1.2 Novelty and ambition

Background

The field of dialogue management is increasingly reliant on data-driven approaches, which are typically based on convolutional or recurrent neural architectures (Gao et al., 2019). The bulk of the research on dialogue management concentrates on so-called "slot-filling applications", where the system objective is to fill a predefined list of slots by interacting with a human user.

In parallel with these task-oriented systems, another line of research focuses on social chatbots able to sustain open-ended conversations with human users. The purpose of these chatbots is to respond to user inputs in a conversationally appropriate manner, without a predefined goal to fulfill. These systems are often trained end-to-end based on (context,response) pairs extracted from large dialogue corpora such as movie subtitles (Vinyals and Le, 2015; Lison and Bibauw, 2017). The dialogue state corresponds in this case to a latent vector derived from the context (which includes the user inputs along with the recent dialogue history). Some neural models also allow for the integration of relevant facts to generate more informative responses (Ghazvininejad et al., 2018).

These data-driven approaches have led to important improvements in the robustness and scalability of dialogue systems. However, in terms of their conceptualisation of the dialogue state, these approaches reflect a step backward compared to earlier, rule-based techniques. Indeed, in many of these earlier techniques, dialogue is explicitly framed as a collaborative activity in which the interlocutors work together to coordinate their actions, maintain a shared conversational context, resolve open issues and satisfy social obligations (Jokinen, 2009). This formalisation require the use of complex state representations encompassing the beliefs, desires and intentions of each conversational partner.. However, although they can account for fine-grained aspects of the dialogue context, these rule-based methods fail to capture partially observed variables in a principled manner.

The use of relational representations for dialogue management has been explored in several recent papers. Ramachandran (2015) presented a belief tracker based on stacked relational trees. In Lison (2015), we described an hybrid approach to dialogue management using probabilistic rules. These rules can rely on logical abstractions to update the dialogue state and select the system actions. However, the structure of these rules must be manually specified. Chen et al. (2018) presented a policy learning method based on structured deep reinforcement learning. In their approach, the Q-values of the dialogue policy are estimated using graph neural networks. The graph structure is, however, limited to the policy optimisation process, while the dialogue state is still expressed with traditional slots. Another approach that stands close to this proposal is the Conversational Entity Dialogue Model of Ultes et al. (2018), which is a dialogue model centred around entities (in their case slots to fill) and relations, with a dialogue policy optimised through feudal reinforcement learning. Finally, He et al. (2017) presents a sequence-to-sequence model in which the decoder relies on graph embeddings derived from a knowledge graph (assumed to be fully observable). Their model is then evaluated in a symmetric collaborative dialogue setting. One limitation of this approach is its dependence on relatively large dialogue corpora to reach acceptable performance.

Weak supervision is a recent machine learning framework employed to estimate model parameters using "weaker"forms of supervision than gold standard annotation labels. These weaker supervision signals may for instance come from low-confidence labels (e.g. from crowd workers), handcrafted patterns/rules, out-of-domain classifiers, prior distributions, domain-specific constraints or invariants. In other words, weak supervision provides a principled, model-agnostic approach to inject domain knowledge (extracted from various sources) into deep learning architectures. Several methods can be employed for parameter learning, such as posterior regularisation (Hu et al., 2016) or variational EM (Wang and Poon, 2018). To our knowledge, the only use of weak supervision methods in dialogue modelling is Mallinar et al. (2018), who described an approach to automatically expand intent recognition labels on chat logs using a data programming approach.

Project objectives

The two central objectives of the **GraphDial** project are:

1. The development of new neural approaches to dialogue management (dialogue state tracking and action selection) using *probabilistic graphs* as core representation of the dialogue state.

2. The use of *weak supervision* signals extracted from user feedback, heuristic rules and generic constraints to learn the parameters of these neural models.

These two objectives are complementary, since expressing the dialogue state as a graph facilitates the specification of weak supervision signals based on heuristic rules and structural constraints (and enhance the human-readability and transparency of the resulting dialogue models).

The neural models developed through the project will be evaluated in the context of human–robot interaction tasks, and a secondary objective of **GraphDial** is to advance the state of the art in this domain. However, the project intends to construct generic dialogue models that are applicable to a broad range of application domains beyond human–robot interaction.

1.3 Research questions and hypotheses, theoretical approach and methodology

To fulfil the objectives mentioned above, the **GraphDial** project will first seek to adapt models of dialogue state tracking and action selection to operate on *graph representations*:

Dialogue state tracking with graph operations. Dialogue state tracking is the task of updating the dialogue state upon new observations, such as user utterances or other events relevant to the system (e.g. new persons detected in the visual scene). The project will investigate how dialogue state tracking can be framed in terms of *graph transformations* conditioned on a given input. In particular, the project will build upon Gated Transformers Neural Networks (Johnson, 2017), which is a recent model for mapping sequential inputs to graph transformations such as adding/updating nodes and edges. However, this framework requires strong supervision (with the full graph provided at each timestep) and may be prone to overfitting when applied to long interactions. Alternatively, graphs can also be inferred from sequential inputs using neural relational inference models (Kipf et al., 2018). Due to uncertainties and partial observability, the graph expressing the dialogue state must be probabilistic (Khan et al., 2018).

A related research question concerns the integration of dialogue management with natural language understanding (NLU). Indeed, the output of semantic parsing typically takes the form of trees or directed acyclic graphs over linguistic entities (Kuhlmann and Oepen, 2016), thereby making it easier to connect NLU with graph-based representations of the dialogue state. The project will look at how dialogue state tracking can operate on semantic graphs (thereby viewing dialogue state tracking as a *graph-to-graph* problem) instead of raw utterances.

Action selection with graph neural networks. Once the probabilistic graph representing the current dialogue state is updated, the next step is to determine how the system should respond. This system response may either be verbal or non-verbal (such as the physical action of fetching an object or moving to another room). The response may also consist of several atomic actions (such as a verbal response accompanied with a gesture). The project will rely on graph neural networks to determine which action(s) to perform based on the current state, using structured graph-based dialogue policies such as Chen et al. (2018).

The neural models above require training data for parameter estimation. The manual annotation of dialogues is a difficult and time-consuming endeavour for complex domains such as HRI. The project will therefore rely on *weak supervision* signals from three distinct sources:

Heuristic rules: Following "data programming" frameworks (Hu et al., 2016; Wang and Poon, 2018), dialogue management models can be learned from handcrafted heuristic rules for state tracking and action selection. Crucially, such an approach will be able to exploit the relational structure of the graph to cover a large fraction of the state-action space in a limited number of rules. In practice, heuristic rules for dialogue state tracking will map specific input patterns to graph transformations, while action selection rules will map graphs to appropriate actions.

Graph constraints: Soft and hard constraints on the graph structure can also provide important signals to the learning algorithm. One example of hard constraint (already mentioned in the introduction) is the fact that a person can only find herself at one location at a given time. Various "soft" constraints can also be provided, such as the fact that, in an office environment, the most likely location of an employee is her own office. This soft constraint may be seen as providing a prior probability on the employee location in the absence of other evidence. In weak supervision frameworks such as Wang and Poon (2018), the relative strength of the constraint is represented by a weight value, with a hard constraint corresponding to an infinite weight.

Grounding patterns: A third source of supervision is the human user herself. Human conversations are characterised by a high degree of collaboration between dialogue participants (Garrod and Pickering, 2009), who routinely collaborate in order to coordinate their contributions and ensure mutual understanding. This process of providing signals about how they understand (or fail to understand) each other's contribution is called *grounding*. The design of appropriate grounding strategies (such as confirmation requests: "Do you want me to search for John?") are crucial in spoken dialogue systems, given the frequent misunderstandings that may arise from e.g. speech recognition errors. But grounding can also be exploited to collect indirect data about the performance of dialogue state tracking[2]. Contrary to the manually annotation of training examples, collecting such conversational data is relatively straightforward as long as grounding and clarification strategies are integrated as part of the dialogue system.

Experimental setup

GraphDial will use *human–robot interaction* as application domain to evaluate the dialogue models developed during the project. Human–robot interaction tasks can indeed provide an ideal testbed for the use of graph-based state representations and weak supervision methods:

- HRI must account for a rich external context including entities in both *small-scale space* (i.e. the robot's immediate surroundings) and *large-scale space* (the entire environment accessible to the robot). In addition, HRI systems must also capture abstract entities such as tasks to perform or facts derived from external databases. This context is inherently *dynamic* (e.g. persons may change positions) and *relational* (as entities are connected through various relations).
- HRI is also characterised by high levels of *noise* and *uncertainty*. In particular, the level of acoustic noise is often higher in HRI than in other domains (due to technical factors such as the relatively large distance between human users and the microphones on the robot, as well as the noise introduced by the robot's mechanical motors). In addition, many aspects of the robot's environment – such as detected persons and objects – are only partially observable, making it necessary to adopt a probabilistic representation of the dialogue state.
- Finally, HRI is also a domain where collecting dialogue data is particularly difficult and time-consuming. Contrary to chatbots and phone-based systems where many users may interact with the same system in parallel, physical HRI platforms cannot be easily scaled up. Relying on user simulators to generate synthetic data is also far from trivial, since such simulations also need to capture the external physical environment in addition to the human user. Appropriate abstraction mechanisms (such as the weak supervision methods investigated in this project) are therefore essential to supplement the scarcity of training data.

The project intends to use the Pepper robot (see Figure 3) as platform to collect interaction data and conduct experiments. The Pepper robot is a semi-humanoid, mobile robot manufactured by Softbank robotics and used in numerous universities and research labs. On the hardware level, the robot features 20 degrees of freedom and is able to navigate through indoor environments using an omnidirectional wheel. The robot integrates a host of sensors, including microphones, cameras (2D, 3D and infrared) and navigation sensors (sonars, lasers, inertial). The robot is fully programmable and comes with a high-level API with dedicated perception modules for e.g. simultaneous localisation and mapping (SLAM), navigation, person detection and object recognition.

These hardware and software components will play an important role for the project, as it will allow the system to engage in *situated* dialogues including references in both small-scale and large-scale spaces. The robot also includes speech recognition and speech synthesis modules based on Nuance's speech engines and available in 15 languages, including Norwegian.

[2] Such grounding responses can take a variety of forms. For instance, a common clarification strategy is to ask the user to rephrase utterances that were not understood. If the user intention is understood after this rephrasing and successfully grounded by the system, one can then assume the semantic content of both utterances to be identical and use this clarification response as a weak supervision signal for the first (misunderstood) utterance.

A key factor behind the choice of Pepper as experimental platform lies in its integration in a single architecture of a large set of ready-to-use modules for human–robot interaction tasks. As a consequence, the project will be able to focus on core scientific challenges (the development of new models for data-driven, graph-based dialogue management), while relying on existing software components for input/output processing tasks such as speech recognition and synthesis.

The experimental evaluation of graph-based dialogue management models will be performed using a combination of objective and subjective metrics of dialogue quality. Such metrics can include measures of task success and dialogue efficiency (e.g number of repetitions or grounding failures), along with subjective metrics based on human ratings (Jokinen and McTear, 2010). The evaluation of dialogue systems is a particularly difficult question (Liu et al., 2016), and Task 1.2) will be specifically devoted to the design of appropriate interaction scenarios and corresponding evaluation metrics to assess the performance of graph-based dialogue management models compared to various baselines. The interaction experiments will be conducted in Norwegian. While the bulk of the data collection effort will be conducted at NR, several partners (Open University, AIRC) have access to Pepper robots and will be able to conduct similar experiments on their premises.

Figure 3: The Pepper robot from Softbank Robotics.

2 Impact

2.1 Potential impact of the proposed approach

Conversational AI is one of the fastest growing field within artificial intelligence, with several popular platforms such as Siri, Google Home, Alexa, Cortana and Xiaoice, along with numerous chatbots currently deployed for e-commerce and customer support. Conversational interfaces are also increasingly integrated in home automation systems, IoT platforms (notably in smart speakers), in-car driving assistants, and service robots for e.g. elderly care.

These interfaces typically rely on deep learning architectures to process the user inputs and produce appropriate responses. However, such neural models have difficulties scaling up to rich conversational contexts and depend on large amounts of training data to reach good performance. The **GraphDial** project will address these two challenges by developing new neural models of dialogue that (1) are able to operate on rich, graph-based representations of the dialogue state and (2) can be estimated from indirect data through weak supervision methods. While the **GraphDial** project makes use of human–robot interaction as application domain, the neural dialogue models developed through the project are intended to be as domain-independent as possible.

Finally, the project is also expected to impact the development of language resources for Norwegian, which remain relatively scarce compared to other languages. In particular, the interaction data collected through the HRI experiments will be released under an open-source licence and made available to the research community (after obtaining the due consent of each participant).

2.2 Measures for communication and exploitation

Dissemination efforts will target a range of complementary channels. As is common practice in natural language processing, scholarly dissemination will follow a process of staged publication, starting with early results in specialised workshops, followed up by consolidated experiments in selective conferences and finally exposed in full in journal articles. The project results will be presented in established international venues such as ACL, SIGDIAL, Interspeech, COLING, EMNLP and HRI along with top-tiered journals such as *Computational Linguistics*, *Computer Speech & Language*, *Dialogue and Discourse*, *ACM Transactions on Human–Robot Interaction* and *Interactive Intelligent Systems*.

To ensure the widest visibility of the research results and foster the involvement of all interested parties, the neural dialogue management models and processing tools developed during the project will be released under an open-source licence and made available on a code repository. In particular, the project aims to release a *domain-independent toolkit* allowing developers to build spoken dialogue systems through the specification of weak supervision sources.

The project will also be in regular contact with two Norwegian "stakeholder" companies that are involved in the development of conversational AI technology:

- Convertelligence AS is a conversational technology company based in Oslo and specialising in delivering chatbot solutions. Their core product is *Kindly*, a platform for building and maintaining chatbots which has been adopted by a growing number of Norwegian companies.
- No Isolation AS develops warm technology to defeat unwanted social loneliness. The company produces the AV1 robot for children with long-term illness and KOMP for analogue seniors.

Although these companies will not participate directly to the project's R&D activities, they will be invited to yearly project meetings and other project-related events, along with other companies that may be interested in following up the project. The objective of these regular contacts is to facilitate the transfer of research results into practical technological solutions.

Popular dissemination will also play an important role in the project, as talking robots are ideal platforms for public outreach. A particular focus for **GraphDial** will be to highlight both the *promises* and *limitations* of conversational AI technology. Media coverage tend to oversell the capabilities of AI models while downplaying the large body of unresolved challenges (such as the lack of commonsense reasoning, embodied knowledge and social cognitive skills). Human–robot interaction constitutes an excellent medium to showcase these limitations in an entertaining and accessible manner.

Finally, the project will also take advantage of its international partners to promote the project ideas beyond Norwegian borders. The project will therefore constitute a springboard for a subsequent international scale-up through the submission of an EU project proposal on this topic.

3 Implementation

3.1 Project manager and project group

The core team that will drive the **GraphDial** is composed of the project leader Pierre Lison along with a PhD Research Fellow that will be hired at the Norwegian Computing Centre for this project. In addition to this core team, a number of collaboration partners from both academia and the industry will contribute to the project with their scientific and technological expertise.

Project leader

The **Norwegian Computing Centre** (NR) is one of Norway's leading research institutions within statistical modelling, machine learning and ICT. Established in 1952, NR is organised as a private, non-profit foundation that carries out R&D projects for a broad range of commercial and public organisations in Norway and internationally. NR leads the BigInsight centre of excellence for research-based innovation, which aims to produce innovative solutions for the knowledge economy through novel statistical and machine-learning methodologies for extracting actionable knowledge from complex data. NR has about 85 employees, a majority of whom research scientists with a Ph.D.

Pierre Lison is a senior research scientist at NR and will lead the **GraphDial** project. Pierre studied computer science and computational linguistics at the Universities of Louvain (Belgium) and Saarland (Germany), and received in 2014 his PhD in computer science from the University of Oslo. Pierre has an extensive research background in natural language processing and has led several R&D projects in this field. In particular, he led the recently completed DialMT project on dialogue modelling for statistical machine translation and the first phase of the SAFERS project on speech analytics for emergency response services, both funded by the Research Council of Norway. Pierre has also been involved in several EU projects in human–robot interaction while a researcher for the German Research Centre for Artificial Intelligence (DFKI), notably CogX ("Cognitive Systems that Self-understand and Self-extend") and ALIZ-E ("Adaptive Strategies for Sustainable Long-Term Social Interaction").

Pierre has over 40 publications in the fields of cognitive robotics, human–robot interaction, dialogue management, natural language processing and machine learning. He is also the principal developer of OpenDial, an open-source toolkit to develop spoken dialogue systems through probabilistic rules. OpenDial has been deployed in several application domains such as intelligent tutoring systems and in-car driver assistants (Lison and Kennington, 2016). More recently, Pierre has been involved in the release and maintenance of large open-source datasets, notably the OpenSubtitles corpus which contains over 3.7 million subtitles from movies and TV series in 60 languages (Lison et al., 2018). The OpenSubtitles corpus is used by hundreds of institutions around the world for research on dialogue modelling, machine translation and cross-lingual NLP.

Pierre was a runner-up for the E. W. Beth prize for outstanding dissertations in the fields of Logic, Language, and Information. He is also a member of the Young Academy of Norway.

Collaboration partners

- *Filippo M. Bianchi* is a research scientist at the Norwegian Research Centre (NORCE). He received his PhD in Machine Learning in 2016 from Sapienza University of Rome. He then held research positions at Ryerson University (Toronto, Canada), Università della Svizzera Italiana (Lugano, Switzerland) and the Machine Learning group at UiT the Arctic University of Norway (Tromsø). He has an extensive research background in complex networks, graph theory, time-series analysis, and dynamical systems. His research interests focus on deep learning and reservoir computing, mainly applied to sequences and graphs. He authored 40 scientific publications.

- *Kristiina Jokinen* is a Senior Researcher at the Japanese Artificial Intelligence Research Centre (AIRC), which is part to the National Institute of Advanced Industrial Science and Technology (AIST), one of the largest public research organisations in Japan. Prior to moving to Japan, she was Adjunct Professor and Project Director at University of Helsinki, Finland, and Visiting Professor at University of Tartu, Estonia. Her research focuses on human–robot interaction, spoken dialogue systems, and multimodal communication. She has published widely on these topics, including three books. She also led several national and international research projects and served as chair for various conferences related to speech, language and dialogue, including the organisation of IWSDS 2016 in Lapland. She is a Life Member of Clare Hall, University of Cambridge.

- *Martijn van Otterlo* is a Senior Lecturer in the Department of Computer Science of the Open University, The Netherlands. He has held research positions at the Katholieke Universiteit Leuven (Belgium), the Vrije Universiteit Amsterdam and Tilburg University and received his PhD from Twente University (The Netherlands). He published two books on reinforcement learning and decision making using probabilistic logical representations, and also worked on language processing, logical models of dialogue and human–robot interaction. His latest research work focused on probabilistic logical models of decision making for responsible AI.

- *Stefan Ultes* is an R&D Engineer at Mercedes Benz Research & Development, Germany. He received his PhD in Computer Science in 2015 from Ulm University (Germany). He was previously Research Associate at the Spoken Dialogue Systems Group at the University of Cambridge, working with Prof. Steve Young on the EPSRC project "Open Domain Statistical Spoken Dialogue Systems", which led to the publication of the Conversational Entity Dialogue Model (Ultes et al., 2018). Stefan was also a lead developer for the open-source PyDial dialogue toolkit (Ultes et al., 2017).

- *Krisztian Balog* is a full professor at the University of Stavanger (Norway), where he leads the Information Access & Interaction group (IAI). His research interests lie in intelligent information access, including entity retrieval and conversational search. Balog serves as a senior PC member at SIGIR, CIKM and ECIR, as an Associate Editor of the ACM Transactions on Information Systems. He has authored over 100 peer-reviewed publications, including a recent book on Entity-Oriented Search

(Springer, 2018). For his significant and influential contributions to the area of semantic search and evaluation methodology, he was honored with the Karen Spärck Jones Award in 2018.

- *Stephan Oepen* is a full professor at the Language Technology Group (LTG), University of Oslo. He previously worked at DFKI and Saarland University (both Germany), YY Technologies (Mountain View), and Stanford University (both USA). His research revolves around the integration of linguistics and computing, where he has published some ninety peer-reviewed articles and conference papers. He is a co-founder of both the DELPH-IN network and the Nordic Language Processing Laboratory (NLPL), and is also member of the editorial board of Computational Linguistics and of the executive committee of the European Association for Computational Linguistics. His recent research work investigates the use of graph representations for semantic parsing.

3.2 Project organisation and management

The Norwegian Computing Centre (NR) will be responsible for the overall management of the project. The project is planned for a duration of four years and is divided in 4 work packages. The tasks, work packages and their relationships are illustrated in Figures 4 and 5.

The project leader (Pierre Lison) will be responsible for leading the four work packages, coordinate the work with the project partners and supervise the PhD student that will be hired for the purpose of this project. Project partners (besides NR) are mentioned by initials for each WP.

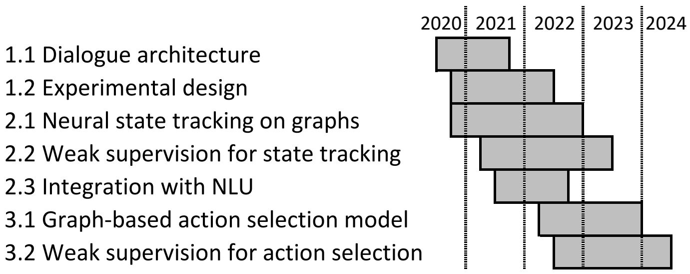

Figure 4: Timeline for the project tasks.

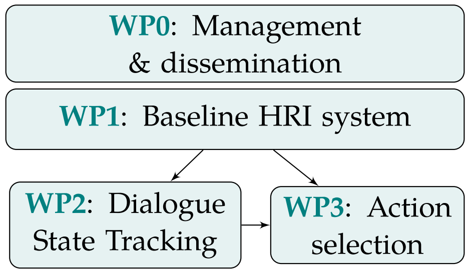

Figure 5: Work packages.

WP0: Project management & dissemination	**Period**: M1-M48
Task 0.1: Project management. Coordination of research activities, PhD supervision, organisation of project-related events, reporting, quality assurance and administrative duties. **Task 0.2: Dissemination**. Research publications, participation to international conferences, public outreach activities, demonstration of prototypes to stakeholders, etc.	

WP1: Baseline system for HRI	**Partners**: KJ, MvO	**Period**: M1-M24
Task 1.1: Dialogue architecture. Integration of core dialogue architecture into Pepper and its onboard modules (person detection, speech recognition & synthesis, navigation) **Task 1.2: Experimental design**. Design of interaction scenarios and evaluation metrics to be used for collecting data and assessing the performance of dialogue management models. Implementation of rule-based baseline system for these interaction scenarios.		

WP2: Dialogue state tracking	**Partners**: FB, KJ, MvO, SU, SO	**Period**: M4-M36
Task 2.1: Neural state tracking on graphs. Tracking models to update the dialogue state (represented as a probabilistic graph) upon new observations through graph transformations. **Task 2.2: Weak supervision for dialogue state tracking**. State tracking models trained with multiple supervision signals (heuristic rules, constraints, grounding/clarification responses). **Task 2.3: NLU integration**. Dialogue state tracking on structured inputs (semantic parses).		

WP3: Action selection	Partners: FB, KJ, MvO, SU, KB	Period: M21-M48
Task 3.1: Graph-based action selection model. Development of action selection models to determine the most appropriate action to perform using graph neural networks.		
Task 3.2: Weak supervision for action selection. Training of action selection model based on a combination of weak supervision signals, including heuristic rules and user feedback.		

In addition to various informal meetings between partners, the project will organise yearly meetings to present ongoing research work and draft future plans. The project partners will be actively encouraged to collaborate across institutions, notably through research visits.

References

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V. F., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gülçehre, Ç., Song, F., Ballard, A. J., Gilmer, J., Dahl, G. E., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., and Pascanu, R. (2018). Relational inductive biases, deep learning, and graph networks. *CoRR*, abs/1806.01261.

Chen, L., Tan, B., Long, S., and Yu, K. (2018). Structured dialogue policy with graph neural networks. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1257–1268.

Gao, J., Galley, M., and Li, L. (2019). Neural approaches to conversational AI. *Foundations and Trends® in Information Retrieval*, 13(2-3):127–298.

Garrod, S. and Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2):292–304.

Getoor, L. and Taskar, B. (2007). *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

Ghazvininejad, M., Brockett, C., Chang, M., Dolan, B., Gao, J., Yih, W., and Galley, M. (2018). A knowledge-grounded neural conversation model. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)*, pages 5110–5117, New Orleans, USA.

He, H., Balakrishnan, A., Eric, M., and Liang, P. (2017). Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1766–1776.

Hu, Z., Ma, X., Liu, Z., Hovy, E., and Xing, E. (2016). Harnessing deep neural networks with logic rules. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2410–2420.

Johnson, D. D. (2017). Learning graphical state transitions. In *Proceedings of the International Conference on Learning Representations (ICLR 2017)*.

Jokinen, K. (2009). *Constructive Dialogue Modelling: Speech Interaction and Rational Agents*. Wiley-Interscience.

Jokinen, K. and McTear, M. (2010). *Spoken Dialogue Systems*. Synthesis lectures on human language technologies. Morgan & Claypool Publishers.

Khan, A., Ye, Y., and Chen, L. (2018). On uncertain graphs. *Synthesis Lectures on Data Management*, 10(1):1–94.

Kipf, T., Fetaya, E., Wang, K.-C., Welling, M., and Zemel, R. (2018). Neural relational inference for interacting systems. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2688–2697, Stockholmsmässan, Stockholm Sweden. PMLR.

Kuhlmann, M. and Oepen, S. (2016). Towards a catalogue of linguistic graph banks. *Computational Linguistics*, 42(4):819–827.

Lison, P. (2015). A hybrid approach to dialogue management based on probabilistic rules. *Computer Speech & Language*, 34(1):232 – 255.

Lison, P. and Bibauw, S. (2017). Not all dialogues are created equal: Instance weighting for neural conversational models. In *Proceedings of the 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2017)*, pages 384–394, Saarbrücken, Germany. ACL.

Lison, P. and Kennington, C. (2016). OpenDial: A toolkit for developing spoken dialogue systems with probabilistic rules. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Demonstrations)*, pages 67–72, Berlin, Germany.

Lison, P., Tiedemann, J., and Kouylekov, M. (2018). OpenSubtitles 2018: Statistical rescoring of sentence alignments in large, noisy parallel corpora. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC-2018)*, Miyazaki, Japan.

Liu, C.-W., Lowe, R., Serban, I., Noseworthy, M., Charlin, L., and Pineau, J. (2016). How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2122–2132, Austin, Texas.

Mallinar, N., Shah, A., Ugrani, R., Gupta, A., Gurusankar, M., Ho, T. K., Liao, Q. V., Zhang, Y., Bellamy, R. K. E., Yates, R., Desmarais, C., and McGregor, B. (2018). Bootstrapping conversational agents with weak supervision. *CoRR*, abs/1812.06176.

Ramachandran, Deepaknd Ratnaparkhi, A. (2015). Belief tracking with stacked relational trees. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 68–76.

Ultes, S., Budzianowski, P., Casanueva, I., Rojas Barahona, L. M., Tseng, B.-H., Wu, Y.-c., Young, S., and Gasic, M. (2018). Addressing objects and their relations: The conversational entity dialogue model. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 273–283.

Ultes, S., Rojas-Barahona, L., Su, P.-H., Vandyke, D., Kim, D., Casanueva, I., Budzianowski, P., Mrkšić, N., Wen, T.-H., Gašić, M., and Young, S. (2017). Pydial: A multi-domain statistical dialogue system toolkit. In *Proceedings of ACL 2017, System Demonstrations*, pages 73–78, Vancouver, Canada. Association for Computational Linguistics.

Vinyals, O. and Le, Q. (2015). A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Wang, H. and Poon, H. (2018). Deep probabilistic logic: A unifying framework for indirect supervision. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1891–1902.