The Need for Grounding in LLM-based Dialogue Systems

Kristiina Jokinen
AI Research Center, National Institute of Advanced Industrial Science and Engineering
Tokyo, Japan
kristiina.jokinen@aist.go.jp

Abstract

Grounding is a pertinent part of the design of LLM-based dialogue systems. Although research on grounding has a long tradition, the paradigm shift caused by LLMs has brought the concept onto the foreground, in particular in the context of cognitive robotics. To avoid generation of irrelevant or false information, the system needs to ground its utterances into real-world events, and to avoid the statistical parrot effect, the system needs to construct shared understanding of the dialogue context and of the partner's intents. Grounding and construction of the shared context enables cooperation between the participants, and thus supports trustworthy interaction. This paper discusses grounding using neural LLM technology. It aims to bridge neural and symbolic computing on the cognitive architecture level, so as to contribute to a better understanding of how conversational reasoning and collaboration can be linked to LLM implementations to support trustworthy and flexible interaction.

Keywords: grounding, spoken dialogue systems, large language models, Theory of Mind, conversational AI, knowledge graphs, language-capable robots

1. Introduction

One of the main challenges in cognitive robotics is language-based communication which should be natural as well as grounded in the context in which the dialogue takes place. As pointed out by Wilcock and Jokinen (2023), among others, the main problem of ChatGPT-type interaction is that the models have no understanding of the real world: sentences are generated as strings of words, but they are not grounded in real world experience and they do not convey feelings or a genuine intention to communicate. A robot may assist humans to manipulate objects or navigate in the environment, so the meaning of the utterances must be linked to a true representation of relevant events, objects and actions. Also the lack of trustworthy information and tendency to hallucinate undermine the reliability of LLMs for applications especially in the health and eldercare domains, because of the model's outdated information and unknown data sources, as well as the "long-tail" problem, i.e., problems learning low-frequency facts (Kandpal et al., 2022). Recently also semantic inconsistency of ChatGPT has been studied (Jang and Lukasiewicz, 2023) with the conclusion that inconsistency issues undermine its reliability and cannot simply be resolved by prompt design and data augmentation.

The contributions of this paper deal with research areas of cognitive robotics and conversational AI. We study the linking of neural and symbolic processing from the point of view of conversational AI and support the view that grounding (actually more than one type of grounding) is needed in LLM-based dialogue systems which aim to be of value for human users by providing cognitively plausible dialogue behaviour. We draft a model that uses conversational AI and knowledge graphs for the purpose of building shared understanding of the dialogue situation, combining neural technologies for symbol-level interaction and creating common ground, and also discuss how grounding can be used to leverage both reliable information exchange and smooth interaction for robot dialogues.

The paper is structured as follows. Section 2 summarizes previous and related work. We discuss the grounding models in Section 3. The knowledge graph technologies used in our models are briefly presented in Section 4. We conclude with discussion on future directions in Section 5.

2. Previous and Related Work

We give an overview of our general framework of Constructive Dialogue Model in Subsection 2.1, and summarize related work in grounding in Subsection 2.2

2.1. Constructive Dialogue Model

Context-aware dialogue research (Jokinen, 2018) emphasizes that an intelligent agent needs to be aware of its context in order to support natural and attentive dialogues. An important characteristic of the agent is the ability to communicate in a manner which is well-timed concerning the partner's attention and appropriately formulated concerning the partner's intentions. Such behavior creates common ground to achieve goals, seek information, and create social bonds, i.e. dialogue partners construct conversation together in their conversational interaction.

Bridging Neurons and Symbols for Natural Language Processing and Knowledge Graphs Reasoning @ LREC-Coling 2024, pages 45–52

In cognitive robotics (Cangelosi and Asada, 2022), robots should communicate with humans in a socially correct way, and their ability to recognize the user's spoken and multimodal utterances must be combined with their own speech, gesturing and multimodal behaviour. Consequently, human-robot interactions resemble interactive situations between two agents. However, our claim is not that the robot agent is conscious about its acts or that it understands the meaning of linguistic symbols in the same way as humans; rather, we put forward the view that human-robot interactions are perceived as natural and intentional, if the robot agent's operation and interaction are based on similar capabilities (affordances) as those used in human-human interactions.

The Constructive Dialogue Model (CDM) is a conceptual and operational framework which regards conversational interactions as cooperative activities through which the participants build common ground (for more information see (Jokinen, 1996, 2009)). The CDM architecture takes into account the multidimensional and intertwined nature of human-agent interaction from a dynamic systems theory perspective. Dynamic systems theory perceives human development as a connectionist process of self-organization and emergence: systems can generate novelty through their own activity, which consists of many decentralized and local interactions that occur in real time. In systemic approaches, communication is understood as the emergent product of multiple activities in the participants' cognitive neuroarchitectures, and it can be viewed as a constant but regulated change within a complex dynamic system, formed by the intertwined activities.

In CDM, participants aim to achieve their communicative goals by conveying information about their intentions and tasks. They are engaged in the exchange of new information which includes feedback about their understanding, attitude, emotions, and willingness to interact. Their individual acts create a new (cognitive) state and together the participants generate conversation as a joint action. The dynamic development of conversation enables the participants to construct mutual understanding (although not necessarily agreement about the tasks and intentions), whereas various enablements of communication constrain and regulate the interaction, such as the need to be in contact, to perceive various partner actions as communicative signals, to be able to understand the partner's message, and to be able to produce one's own reaction. Reaction encodes new information which changes the system state and causes the agents to organise their reasoning with respect to the new state.

One of the main challenges for CDM is how to update one's knowledge in order to align with the partner to construct shared context and react appropriately. The process of grounding is used to establish links between new and old information, and to determine optimal communicative action for the construction of shared knowledge. Grounding is manifested by the signals that indicate the agents' cooperation and their attention to the partner's needs: verbal acknowledgement and relevant continuation of the conversation is accompanied by non-verbal feedback. Several studies deal with multimodal feedback-giving processes, expressed by eye-gaze, facial expressions, head nods, hand gestures, body movement, and a wide range of vocalisations such as laughter etc. For instance, Mori et al. (2022) studied nods in human conversations, and proposed a model which includes a component for updating the partner's internal cognitive state (such as knowledge, understanding and emotional stance), on the basis of which the agent can decide on the appropriate feedback. The model focuses on the type of nod, but takes into account also a whole repertoire of possible feedback expression (verbal, gesturing, body posture). Models for such expressive interaction are important in many cognitive robotics applications, where task completion is not enough but more comprehensive and affective interaction is desired.

2.2. Related Work

Rather than focusing solely on task completion as the basis of the efficient communication, linguistic grounding research has focussed on naturalness of interactions and measuring user engagement through the participants' multimodal activity. Such approaches concern cooperative dialogue management (Clark and Wilkes-Gibbs, 1986; Clark and Brennan, 1991; Allwood et al., 1992; Traum and Allen, 1992; Traum and Heeman, 1996), and more recently (Kawano et al., 2021; Udagawa and Aizawa, 2021). Some recent research on linguistic grounding has also been conducted by Axelsson and Skantze (2023a,b) using the Furhat robot as an interface robot. In this work, general knowledge graph entities and links are marked by temporary labels, and the memory space has to be updated every time time a dialogue session starts.

Jokinen et al. (2024) explore how to predict grounding and shared knowledge in dialogues, given the listener feedback and a suitable prompt design with examples. In particular, they investigate if LLMs can be used to construct shared knowledge in an interactive context of an information seeker and an information provider conversing over a particular topic. The information provider's knowledge is structured in a table format, while the seeker queries the information until understood and satisfied with the result. Three types of conversational grounding are assumed: explicit (expressed by explicit feedback signals that show the partner's understanding), implicit (expressed by moving on in the dialogue to other topics without explicit linguistic signals to show the partner's understanding), and clarification (expressed by a clarification question to ask further information). The results are positive and demonstrate the LLMs capability to dynamically build structured knowledge, but further studies are needed to distinguish between the implicit and clarification types of grounding, to include multimodal feedback, and to fine-grain the analysis of situations where misunderstandings occur.

In a series of papers (Wilcock and Jokinen, 2022a,c; Jokinen and Wilcock, 2024), Jokinen and Wilcock have extensively studied cooperative and uncooperative robot behaviour, also using Furhat robots. They propose a solution with knowledge graphs for grounding to control LLMs in dialogue modelling and to alleviate the LLM's tendency to produce false information.

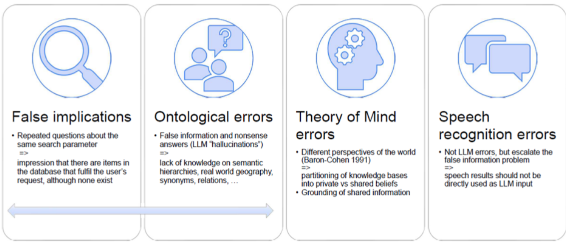

Figure 1: Different errors in LLM-based dialogues.

Their recent work (Wilcock and Jokinen, 2023) compares LLM-based dialogue systems with knowledge graph-based systems from the point of view of errors that occur in testing. They distinguish false implications, ontological errors, and Theory of Mind errors, as shown in Figure 1. The figure also includes speech recognition errors but these are not discussed here as their solutions are not directly included in the knowledge-base reasoning.

False implications are errors where the user is led into making assumptions that are not true, while ontological errors result from a lack of knowledge of the semantics and structure of the world. They can be remedied by adding semantic metadata such as taxonomies and geographical locations to the knowledge graphs, and by using more flexible searches. Theory of Mind errors occur when participants have different perspectives of the situation, and are caused by lack of grounding.

3. Grounding Models

We distinguish between Theory of Mind grounding (in Subsection 3.1) and knowledge grounding (in Subsection 3.2).

3.1. Theory of Mind and Grounding

As mentioned, Wilcock and Jokinen (2023) point out that while the other interaction errors may be resolved by the RAG approach and its developments, Theory of Mind (ToM) errors occur when the participants have different knowledge of the situation and its solution requires modelling of the partner's mental state.

According to Theory of Mind (ToM) (Baron-Cohen, 1991), the development of human cognition requires the understanding of other minds having different content than one's own: another person's mind is related to their perspective of the world which is not necessarily the same as one's own. In cognitive robotics, ToM is used as a basis for the studies to construct a shared knowledge and mutual understanding of the context of the physical world, which are also the main issues in cooperative dialogue modelling.

LLM-based interactions lack shared understanding of the partner's worldview, experience, emotions, and environment. Figure 2 exemplifies a common situation in human-robot interactions. In the user's mind the referent of *the last one* is the recently listed item *To the Herbs*, whereas the system regards the phrase *the last one* to refer to *Pesche Doro*, the last one in its list of database items. The mismatch leads to confusion but can be resolved by a clarification question. The error is not seen as a mistake but as a lack of relevant knowledge, and its recovery thus becomes a matter of constructing appropriate shared knowledge. We call this *conversational grounding*, as it is based on the conversational context of what the partners have been talking about and how they interpret the language referents in the current context.

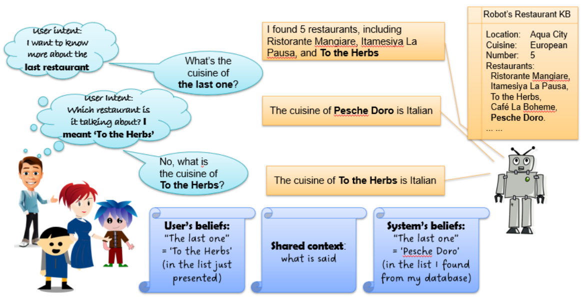

Figure 2: Conversational grounding (ToM error).

Another type of grounding is exemplified in Figure 3. In order to act in the real world and cooperate with humans, the robot agent must have knowledge of the environment and how language concepts are linked to the entities in the environment where the interaction takes place. For instance, in object manipulation and navigation tasks where the robot collaborates with humans, computer vision technology needs to be combined with LLMs to give the robot a sense of the environment and the skill to talk about it. We call this *visual grounding*, which has been long studied in robotics (cf. (Harnad, 1990)), where it refers to the grounding of utterances into the perceptions of the world. Simultaneous visual and conversational grounding allows the agent to assess the relevance and truth of the partner's utterances with respect to the current environment and to generate an appropriate response within the shared knowledge, e.g. asking a clarification question to recognize the correct referent mentioned by the user.

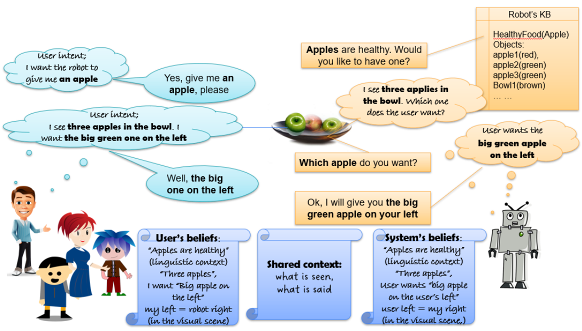

Figure 3: Visual grounding and the real world.

Perspective taking is one of the challenges in current computer vision research (Lemaignan et al., 2011), whereas recent advances in Visual Dialogue Modelling (Wu et al., 2017) combine speech and images in order to allow spoken natural language questions and answers deal with the elements that are recognized in the image.

To address ToM errors, the agent must distinguish private and shared knowledge, and have a goal to build shared knowledge in order to advance the task via communication. We make a distinction between existing static knowledge and dynamic dialogue processing knowledge, but represent both in a knowledge graph. Each user can have a personal knowledgebase which contains their personal information and preferences but can also be extended dynamically in the dialogue, including their view of the dialogue situation. In order to update one's own knowledge and align it with the partner's knowledge, the agent constructs a shared context as part of the knowledge representation. We aim to leverage the knowledgebase approach for updates and reasoning by deploying the typical procedures for searching and updating knowledge graph databases. For instance, communicative actions establish links between the nodes in the graph structure, and these can be dynamically updated as property updates of the entities and the links.

3.2. LLMs, KGs and Grounding

As discussed above, simple application of LLMs enables the robot agents to talk fluently on any topic, but the sentences are basically imitations of what could be said, rather than manifestations of the speaker's intention to convey some information to the partner (giving rise to the phrase "statistical parrot" (Bender et al., 2021)). In the knowledgebase approach, generated sentences are grounded in the knowledgebase, curated by humans to represent true facts of the world. Ontologies and semantic metadata are important tools in providing necessary information about how the world is structured (see Wilcock and Jokinen (2022b)) and we can also use different knowledgebases (document collections, knowledge graphs) which contains relevant information about the domain and dialogue, and can also be said to "represent" the world.

Currently much research is focused on combining Knowledge Graphs (KGs) and LLMs, and a survey of this work is provided by Pan et al. (2023). When KGs are curated by human experts, the data provenance is known and errors of outdated data can be resolved (cf. Wikipedia). For instance, Di Bratto et al. (2021) describe how graph databases can be used as a framework for a understanding the domain during dialogue. They use Internet Movie Database and Wikidata with a reference to personal and common ground concepts. Wilcock and Jokinen (2023) discuss how KGs can be used to provide trustworthy information to the user and how KGs can be augmented with WikiData metadata. Fu et al. (2023) present how KG reasoning and ontologies enable more cooperative responses based on reliable data, and Schneider et al. (2023) describe how to use knowledge graphs and conversational interfaces for exploratory search, bridging the gap between structured and unstructured information retrieval on news articles.

As fluent conversational capability is one of the main advantages of LLMs, current research efforts aim to combine such capability with trustworthy reliable information. The third meaning for "grounding" can hence be found in the LLM and Knowledge Graph literature: it is discussed in the context of knowledgebases providing a reliable starting point for the LLM generation. We call this *knowledge grounding* as it refers to the grounding of linguistic information to the speaker's knowledge and experience of world, stored in knowledgebases and represented in texts, KGs, and cognitive models of the agent's knowledge.

4. Grounding and Knowledge Graph Technologies

In this section we briefly explore how knowledge grounding can be included in dialogue management, using LLMs, retrieval augmentation, and knowledge graphs.

The RAG (Retrieval Augmented Generation) approach (Lewis et al., 2020) is commonly used as a generation model for reliable knowledge inclusion. It provides a solution to problems with false implications and ontological errors. The processing pipeline is divided into language understanding and response generation. First the user input is analysed to extract important concepts and the user intent. The analysis is then used for making a search query to retrieve relevant information from the knowledgebase. Response generation uses the information retrieved from the knowledgebase together with the user query and dialogue history, as input for the LLM-based generation module which then generates a response.

We use Neo4j graph databases (Robinson et al., 2015) in our work. Most recently LLMs have been used to generate Cypher queries that can search knowledge graphs in Neo4j (Bratanic, 2023). Symbolic representation of knowledge can thus be used as a grounding model for LLMs. Neo4j also includes a vector search capability which supports efficient semantic search of KGs by adding an embedding vector to each node. It can be used with LLMs to make semantic searches based on user queries in natural language that do not require exact lexical matches with node labels. It is interesting that the description of this capability refers to "grounding LLM responses", which in this paper is regarded as an example of knowledge grounding, i.e. representing a way how generative models ground their responses into curated knowledge.

This approach has been demonstated in Wilcock and Jokinen (2023) where knowledge graphs are used with robot dialogues in the CityTalk application to talk about restaurants and hotels (Wilcock, 2019). A similar approach is used in Jokinen and Wilcock (2024) but the interaction deals with the Kyoto cooking database (Kiyomaru et al., 2018) which has been converted into a knowledge graph. The graph is stored in a Neo4j graph database, as shown in Figure 4.

All the nodes in the Kyoto Cooking database are labelled with Japanese names. It is thus possible to have multilingual interaction as the graph can be queried in English or Japanese. An example of these mixed-language queries in Figure 5 is from an earlier version of the system where the names of dishes, ingredients, nutrients, and cooking methods in the responses are in Japanese, and the number of responses is limited to 3.

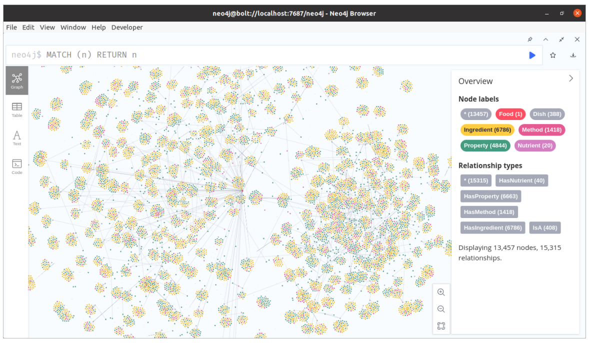

Figure 4: Kyoto Cooking knowledge graph in Neo4j.

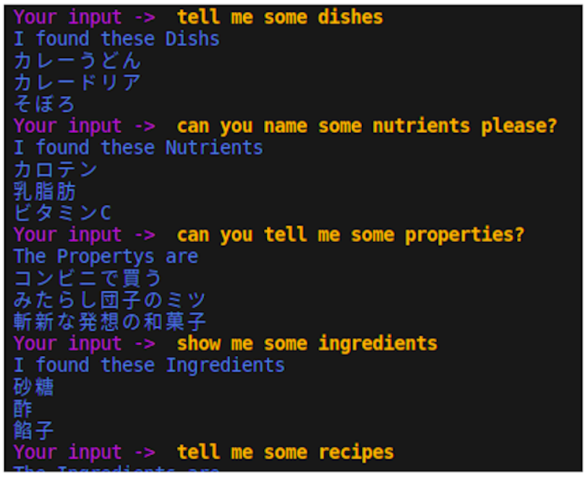

Figure 5: Currently mixed-language responses.

5. Conclusion and Future Work

The paper describes ongoing research on human-robot dialogues where knowledge graphs are used to make the interaction more natural and trustworthy. The paper supports the view that human interaction with robots is quite unlike interactions with text-based systems or with other types of mobile devices, and that conversational robot agents should enable a grounding process in order to create shared context with the human partner, so as to advance technological readiness of cognitive robot applications.

The shared context is constructed through grounding dynamically in the conversation, and it is represented by knowledge graphs. Structured knowledge modelling concerns relevant information of the application domain and of the world, and ultimately of the speaker's own view-point of the real-world events and entities. The paper aims to show the dynamic nature of grounding and the complexity of the construction of shared knowledge between the dialogue partners.

Three different types of grounding are distinguished: 1) conversational grounding establishes links from language expressions to the shared dialogue context (i.e. beliefs of what knowledge is shared in the context), 2) visual grounding supports grounding of language expressions to suitable elements in the context taking into account the whole visual scene, and 3) knowledge grounding anchors language expressions into the agent's own knowledge (long-term memory in which the agent's knowledge and experience is stored). Each type has an important role in the communication and in the processing of the partner's communicative signals. They also demonstrate how the symbolic representations can be grounded within the same framework of structured knowledge graphs as vectorized documents and LLMs, thus linking symbolic representations of thoughts and intentions to cognitive processing of neural representations. The grounding models also show how the dynamic communication system can be controlled by communicative enablements, and how the problematic issues of false and irrelevant information can be alleviated to harness the conversational power of LLMs for language-capable robots.

Future work concerns user studies to evaluate appropriateness and success of the dialogues, as well as application of the approach to knowledge bases of various sizes and domains. In grounding research, multimodal aspects of dialogue need to be taken into account, as well as better understanding of the grounding process and its cogntive modelling. Main challenges deal with the construction of structured knowledge bases, their maintenance and updating, sustainability of LLMs, and various ethical aspects (Williams et al., 2023) related to language capable agents.

6. Acknowledgements

The author acknowledges the support of Project JPNP20006 commissioned by the New Energy and Industrial Technology Development Organization (NEDO), Japan

7. Bibliographical References

Jens Allwood, Joakim Nivre, and Elizabeth Ahlsén. 1992. On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics*.

Agnes Axelsson and Gabriel Skantze. 2023a. Do you follow? a fully automated system for adaptive robot presenters. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '23, page 102–111, New York, NY, USA. Association for Computing Machinery.

Agnes Axelsson and Gabriel Skantze. 2023b. Using large language models for zero-shot natural language generation from knowledge graphs. ArXiv:2307.07312.

Simon Baron-Cohen. 1991. Precursors to a theory of mind: Understanding attention in others. In A. Whiten, editor, *Natural theories of mind: Evolution, development and simulation of everyday mindreading*, pages 233–251. Basil Blackwell.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, page 610–623. Association for Computing Machinery.

Tomaz Bratanic. 2023. Generating Cypher queries with ChatGPT 4 on any graph schema. `https://neo4j.com/developer-blog/generating-cypher-queries-with-chatgpt-4-on-any-graph-schema/`.

Angelo Cangelosi and Minoru Asada. 2022. *Cognitive Robotics*. The MIT Press.

Herbert H. Clark and S. A. Brennan. 1991. Grounding in communication. In L.B. Resnick, J.M. Levine, and S.D. Teasley, editors, *Perspectives on socially shared cognition*. APA Books.

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Martina Di Bratto, Maria Di Maro, Antonio Origlia, and Francesco Cutugno. 2021. Dialogue analysis with graph databases: Characterising domain items usage for movie recommendations. In *Proceedings of the Eighth Italian Conference on Computational Linguistics CliC-it*, Milan, Italy.

Yahui Fu, Koji Inoue, Chenhui Chu, and Tatsuya Kawahara. 2023. Reasoning before responding: Integrating commonsense-based causality explanation for empathetic response generation. In *Proceedings of the 24th Meeting of the Special Interest Group on Discourse and Dialogue, SIGDIAL 2023, Prague, Czechia, September 11 - 15, 2023*, pages 645–656. Association for Computational Linguistics.

Stevan Harnad. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1):335–346.

Myeongjun Erik Jang and Thomas Lukasiewicz. 2023. Consistency analysis of ChatGPT. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing,* *EMNLP 2023, Singapore, December 6-10, 2023*. Association for Computational Linguistics.

Kristiina Jokinen. 1996. Cooperative Response Planning in CDM: Reasoning about Communicative Strategies. In *Twente Workshop Series in Language Technology*.

Kristiina Jokinen. 2009. *Constructive Dialogue Modelling: Speech Interaction and Rational Agents*. John Wiley & Sons.

Kristiina Jokinen. 2018. Dialogue models for socially intelligent robots. In *10th International Conference, ICSR 2018, Qingdao, China, November 28 - 30, 2018, Proceedings*, pages 127–138.

Kristiina Jokinen, Phillip Schneider, and Taiga Mori. 2024. Towards Harnessing Large Language Models for Comprehension of Conversational Grounding. In *14th International Workshop on Spoken Dialogue Systems Technology (IWSDS 2024)*, Sapporo, Japan.

Kristiina Jokinen and Graham Wilcock. 2024. Exploring a Japanese cooking database. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI 2024)*, pages 578–582, Boulder, Colorado, USA. Association for Computing Machinery.

Nikhil Kandpal, Haikang Deng, Adam Roberts, Eric Wallace, and Colin Raffel. 2022. Large language models struggle to learn long-tail knowledge. ArXiv:2211.08411.

Seiya Kawano, Koichiro Yoshino, David Traum, and Satoshi Nakamura. 2021. Dialogue structure parsing on multi-floor dialogue based on multi-task learning. Presented at Robotdial Workshop.

Kanichi Kiyomaru, Sadao Kurohashi, Mitsuru Endo, and Katsuyoshi Yamagami. 2018. Building a basic cooking knowledge base based on cooking recipes and crowdsourcing (in Japanese). In *24th Annual Conference of the Natural Language Processing Society*, Japan.

Séverin Lemaignan, Raquel Ros, Rachid Alami, and Michael Beetz. 2011. What are you talking about? grounding dialogue in a perspective-aware robotic architecture. In *2011 RO-MAN*, pages 107–112.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, pages 9459–9474, Vancouver, Canada.

Taiga Mori, Kristiina Jokinen, and Yasuharu Den. 2022. Cognitive States and Types of Nods. In *Proceedings of the International LREC Workshop on People in Vision, Language and the Mind (P-VLAM)*, pages 17–25, Marseille, France. European Language Resources Association (ELRA).

Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. 2023. Unifying large language models and knowledge graphs: A roadmap.

Ian Robinson, Jim Webber, and Emil Eifrem. 2015. *Graph Databases (2nd edition)*. O'Reilly Media.

Phillip Schneider, Nils Rehtanz, Kristiina Jokinen, and Florian Matthes. 2023. From data to dialogue: Leveraging the structure of knowledge graphs for conversational exploratory search. In *Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation (PACLIC 2023)*, Hong Kong, China.

David R. Traum and James F. Allen. 1992. A Speech Acts Approach to Grounding in Conversation. In *Proceedings of 2nd International Conference on Spoken Language Processing (ICSLP-92)*, pages 137–140.

David R. Traum and Peter Heeman. 1996. Utterance Units and Grounding in Spoken Dialogue. In *Proceedings of International Conference on Spoken Language Processing (ICSLP-96)*.

Takuma Udagawa and Akiko Aizawa. 2021. Maintaining Common Ground in Dynamic Environments. *Transactions of the Association for Computational Linguistics*, 9:995–1011.

Graham Wilcock. 2019. CityTalk: Robots that talk to tourists and can switch domains during the dialogue. In *9th International Workshop on Spoken Dialogue Systems Technology*, pages 411–417. Springer.

Graham Wilcock and Kristiina Jokinen. 2022a. Conversational AI and knowledge graphs for social robot interaction. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI 2022)*, pages 1090–1094, Sapporo, Japan. Association for Computing Machinery.

Graham Wilcock and Kristiina Jokinen. 2022b. Cooperative and uncooperative behaviour in task-oriented dialogues with social robots. In *31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2022)*, pages 763–768, Napoli, Italy.

Graham Wilcock and Kristiina Jokinen. 2022c. Should robots indicate the trustworthiness of information from knowledge graphs? In *10th International Conference on Affective Computing and*

Intelligent Interaction (ACII 2022) Workshops and Demos, Nara, Japan. IEEE Computer Society.

Graham Wilcock and Kristiina Jokinen. 2023. To Err Is Robotic; to Earn Trust, Divine: Comparing ChatGPT and Knowledge Graphs for HRI. In *32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023)*, pages 1396–1401, Busan, Korea.

Tom Williams, Cynthia Matuszek, Kristiina Jokinen, Raj Korpan, James Pustejovsky, and Brian Scassellati. 2023. Voice in the Machine: Ethical Considerations for Language-Capable Robots. *Communications of the ACM*, 66(8):20–23.

Qi Wu, Peng Wang, Chunhua Shen, Ian Reid, and Anton van den Hengel. 2017. Are you talking to me? reasoned visual dialog generation through adversarial learning. ArXiv:1711.07613.