

Research Article

Maria Di Maro*, Martina Di Bratto, Sabrina Mennella, Antonio Origlia, and
Francesco Cutugno

Argumentation in recommender dialogue agents (ARDA): An unexpected journey from Pragmatics to conversational agents

<https://doi.org/10.1515/opli-2025-0052>

received July 15, 2024; accepted February 22, 2025

Abstract: This article introduces argumentation in recommender dialogue agents (ARDA), a novel theoretical and practical framework for designing advanced argumentative dialogue systems. Grounded in principles of pragmatics and argumentation theory, ARDA integrates linguistic theory and graph-based representations to model pragmatic dialogue acts such as clarification requests, explanations, and argumentation. The framework bridges the gap between theory and implementation by providing graph-based computational models that translate these formalised concepts into functional dialogue system components. The contributions of this article are twofold: (1) providing an overarching view of the theoretical approaches applied in our prior studies and (2) offering computational models that more effectively represent linguistic phenomena within dialogue systems. Furthermore, the article explores the application of ARDA in the context of movie recommendation systems, providing previously collected results that illustrate how these models enable natural, persuasive, and logically coherent interactions between humans and intelligent agents.

Keywords: argumentation-based dialogue, linguistically-motivated AI, pragmatics

1 Introduction

The advent of large language models (LLMs) has caused a change in how linguistic theories are utilised by scholars working on dialogue systems, e.g. technological systems specifically designed to enable natural and interactive communication between humans and machines. While these theories were previously considered a starting point for developing dialogue system architectures, they now seem to have become regularities detected *ex post* in statistical models trained on texts. Surface-level regularities, such as syntax, can be efficiently modelled statistically: texts produced by LLMs are indeed very convincing and fluent. However, linguistics, particularly in the field of pragmatics, also describes the motivations behind the performance of a speech act. Machine learning, being exposed only to the final manifestation of a complex cognitive process (e.g. the text), models textual regularities but not the reasons for their production. This implies that, in the

* **Corresponding author: Maria Di Maro**, Department of Electrical Engineering and Information Technology, University of Naples Federico II, Interdepartmental Center URBAN/ECO, Naples, Italy, e-mail: maria.dimaro2@unina.it

Martina Di Bratto: Department of Humanities, University of Naples Federico II, Naples, Italy, e-mail: martina.dibratto@unina.it

Sabrina Mennella: Department of Humanities, University of Catania, Interdepartmental Research Center URBAN/ECO, Catania, Italy, e-mail: sabrina.mennella@phd.unict.it

Antonio Origlia: Department of Electrical Engineering and Information Technology, University of Naples Federico II, Interdepartmental Center URBAN/ECO, Naples, Italy, e-mail: antonio.origlia@unina.it

Francesco Cutugno: Department of Electrical Engineering and Information Technology, University of Naples Federico II, Interdepartmental Center URBAN/ECO, Naples, Italy, e-mail: cutugno@unina.it

development of artificial intelligence (AI) approaches, linguistics should be regarded as a *cause* for their design rather than an *effect* they generate. In this way, machines use language to pursue interpretable internal goals rather than merely mimicking patterns from the training set.

The previous research we conducted dealt with the realisation of linguistically motivated dialogue systems where LLMs are present but with a role limited to the management of surface level aspects of communication: being *generative* approaches, we constrain them to the task of natural language *generation*. Conversely, linguistic theories, especially pragmatic ones, are used to design AI modules specialised in the decision-making process. In the past, we presented practical implementations exploring different aspects of dialogue management. A deep presentation of the linguistic background motivating the technological application, however, was beyond the goals of those papers. These previous experimental results are summarised throughout this work for each relevant part of the presented theoretical framework. Our novel contribution consists in all the details of the framework, which was developed as the motivation behind those past applications.

Concerning argumentation-based dialogues in particular, the influential work by Prakken (2018b) has lamented a proliferation of *ad-hoc* technological applications lacking a general theory highlighting the connections between the different pragmatic phenomena. By abstracting the common principles guiding our view about the implementation of argumentative dialogue systems, we believe we contribute in overcoming this problem. By discussing lessons learned, in this article, we provide a relevant starting point for other researchers interested in developing explainable-by-design AI whose behaviour is motivated by transparent pragmatic goals.

The motivation for our approach, which will be explored in detail throughout this work, contrasts with the current trend concerning the use of generative AI, which has both practical and theoretical limitations with respect to natural language processing. The theoretical framework presented here is characterised by the following main contributions:

- C1 A linguistically motivated methodological approach that spans from linguistic research to the development of argumentative dialogue systems; we call this encompassing approach *argumentation in recommender dialogue agents* (ARDA).
- C2 A graph-based interpretation of the methodology for corpus-linguistics data analysis; we call this first specific sub-part of ARDA *Linguistically Oriented Resources and Insights as Expressive Networks* (LORIEN).
- C3 A technological architecture for implementing embodied conversational agents using the graphical representation of linguistic concepts; we call this second relevant subpart of ARDA *Modeling Operative Representations for Dialogue Orchestration Research* (MORDOR).

In this view, we aim to provide a comprehensive overview of argumentative dialogue and its fundamental components and, more specifically, our argumentation-based dialogue model for recommender systems.

The article is structured as follows: in Section 2, we will first analyse the concept of dialogue and how this is closely connected to the idea of argumentation in order to emphasise how important it is to consider it when dealing with conversation, both from a purely linguistic and implementation point of view. Then we will summarise the motivations and the general structure of the proposed framework. In Section 3, we will establish the theoretical foundations of argumentation theory, focusing specifically on the recommendation task. In Section 4, we will focus on the linguistic aspects comprised in the framework to define a methodology for argumentation-based dialogue systems, as far as both theory representation (C1) (Section 4.1) and graph-based pragmatic applications (C2) (Section 4.2) are concerned. In Section 5, we will describe a computational model using the graph-formalised theoretical findings for evaluation purposes (C3). Finally, in Section 6, a previous work that led to the definition of this framework is summarised.

2 ARDA

Motivating the design of technological applications with linguistics theory rather than looking for linguistics phenomena in technological models constitutes our approach to answering **C1**. Therefore, we provide in this

section the linguistic basics concerning argumentation-based dialogue, and we explain how the multidisciplinary interaction of the two fields leads to the technological one reflecting the concepts coming from linguistics. While this type of cooperation has been explored in the past, our approach is based on the shared use of graph-based representation. This is made possible by the recent availability of graph-based tools supporting at the same time linguistic analysis and dialogue system developments. By grounding our exploration in linguistic principles and supported by advanced graph-based tools, we aim, therefore, to develop dialogue systems that move beyond transactional exchanges, capturing the richness of human–human communication, encompassing all the phenomena that this complex process entails.

Dialogue is defined as “a serious exchange of opinion, esp. among people or groups that disagree.”¹ From this definition, the presence of a *disagreement*, intended as an *argument* or a situation in which people do not have the same opinion, is necessary for a dialogue to take place. Similarly, in the study by Walton (1995), *dialogue* is viewed as an exchange aimed at resolving differences or achieving mutual goals through argumentation, explanation, and clarification. On these premises, it can be argued whether modern dialogue systems engage users in actual dialogues. The most common architectures, in fact, are often evaluated on their capability to interpret users’ intents and react to these by providing the correct service. In these cases, the exchange of information is often exhausted after very few interactions and is mostly aimed at understanding the users’ position rather than negotiating a *Common Ground* (Clark 1996) (Section 4.1). This is because dialogue systems designed as interfaces for service providers are typically capable of requesting additional information from users when critical data is missing for fulfilling a service request. However, beyond this basic functionality, their capabilities remain quite limited. Even in cases of unclear input, these systems often rely on a narrow set of strategies, such as basic clarification requests (CRs), which represent only a small subset of the diverse strategies employed in human–human communication to address different types of conversational problems (Section 4.2.1).

While there has been significant progress in the development of argumentative dialogue systems (e.g. Hunter et al. (2019); Hadoux et al. (2023); Castagna et al. (2024); Ruiz-Dolz et al. (2024)), many of these, especially those not explicitly designed for argumentation, lack robust argumentation skills. More specifically, Hinton and Wagemans (2023) pointed out that, even for GPT-3, generating persuasive, well-reasoned argumentation is far more challenging than producing coherent language, necessitating methods for verifying the correctness and plausibility of the generated arguments. Furthermore, as Prakken (2018b) observes, there are many studies in the field of argumentation-based dialogue focusing on different aspects of argumentation, but there is still no general framework to unify these diverse approaches. This highlights a critical gap in the development of dialogue systems that can effectively mimic the nuanced and multi-faceted nature of human argumentation.

In the study by Van Eemeren and Grootendorst (2003), argumentation is described as a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions justifying or refuting the proposition expressed in the standpoint. The art of speaking by engaging people in prolonged interaction has been explored from multiple points of view since ancient times. Without intending to cover the whole historical and philosophical issues, it is useful to consider an overview of how the topic has been dealt with in the field of philosophy. A difference is historically made among rhetoric, debate, and dialectic. For political reasons, both the Greeks and the Romans concentrated on *rhetoric* which refers to the ability of a single person to address a group or class of people and align them with their view. Conversely, *debate* involves one-to-one exchange of ideas aimed at convincing the attending audience to align with the different positions expressed. In this case, the speaker does not directly address people in the audience but indirectly tries to influence them by exchanging ideas with the opposing person. *Dialectic* is heavily based on one-to-one exchange of ideas, and it can be interpreted either in a competitive or in a collaborative way. This is a critical distinction because, in both rhetoric and debate, a collaborative approach cannot be found, either because communication is mostly one directional or because the goal of the exchange is not to find an agreement but, rather, to stress the differences between the involved

¹ Cambridge Dictionary Online: <https://dictionary.cambridge.org/> [last visited on 04/07/2024].

participants. The principles governing dialectic, on the contrary, are of interest for dialogue systems, which are often intended to collaborate with a human user rather than to oppose them.

Dialectic means ‘related to dialogue,’ stressing the close connections between the two. Since the time of the ancient Greek philosophers, dialectic has been explored through the presentation of dialogues, where sequences of statements between two participants were used to illustrate the process of two parties converging towards a shared view about a chosen topic. Socrates, by defining a process of topic investigation through posing questions and detecting counterfactual evidence, poses one of the basis of argumentations: error detection and resolution of problems. Plato interprets dialectic as the process through which it is possible to shift from *sensibles* to *intelligibles*, thus describing an iterative abstraction process leading from the material world to its abstract representation. Plato offers a view of dialectic that puts it very close to formal logic because of the underlying goal of finding an absolute truth. Aristotle’s description of dialectic emphasises the significance of positions expressed by a single individual, in contrast to the positions held by a multitude of people. This is a critical point of dialectic as rhetoric “[...] will not consider what is plausible to an individual, such as Socrates or Hippias, but what is so to such-and-such people” (Evans 1977, p. 76). Although dialectics has since then always been an important topic in philosophical studies, its use in modern philosophy has been strongly shaped by Hegel, who adopts a dialectical approach to describe how the tension between a concept and its opposite creates higher level concepts. In the case of Hegel’s, dialectic is not directly used to describe communication but as a means to present a form of logic that could go beyond the assumption that, given some premises and a conclusion from those premises, should the conclusion be found false, its premises should be discarded too, which dates back to Plato. Schopenhauer, in *The art of always being right* (Schopenhauer 2004), ironically stresses the distance between formal logic and dialectic by listing a series of strategies that are not aimed to verify the truth of the statements put forward by the opposing speaker. They are rather aimed at, for example, swaying the opponent, eliciting strong negative emotions to induce them in error, or putting the arguments in socially despicable categories. In Schopenhauer’s view about dialectic, the competitive stance is fundamental to consider the occurrence of dialogue strategies that are not aimed at solving the matter at hand.

Similarly, we can affirm that dialogue itself emerges through various kinds of conflict, which are navigated and managed through the principles of dialectic. Consequently, the fundamental ability required for a dialogue system to apply dialectic is the capacity to engage in argumentation. This includes, among other things, the ability to detect counter-arguments to statements proposed by the user as *shared knowledge* or to collaboratively find solutions to problems. Dialogue management, however, varies depending on whether the context is competitive or collaborative. In competitive scenarios, strategies aimed at provoking or deceiving the interlocutor may be employed. In a collaborative context, on the other hand, strategies not focused on resolving the problem or achieving shared views about the topic lose their relevance. As dialogue systems are typically intended to collaborate with human users to find a common understanding about the problems at hand, modelling logical capabilities to enable conflict detection, along with managing social interactions to effectively support the exchange of positions, represents a key focus for improving these artefacts.

In recent years, dialogue systems have gained significant attention and become an integral part of our daily interactions with technology. These systems allow users to perform a wide range of tasks, obtain information, and receive personalised recommendations. To fully leverage their potential, dialogue systems must integrate theoretical insights that enhance their ability to manage complex interactions, such as those involving argumentation, which becomes therefore a pivotal aspect to investigate. This growing significance highlights the need for dialogue systems to be grounded in solid theoretical frameworks, which is where the collaboration between humanities and technology becomes crucial. The motivation behind the framework we present here lies in the general principle of such a collaboration. Implementing technological artefacts based on theoretically motivated backgrounds, on the one hand, supports explainability and controllability in AI systems. On the other hand, this process is not one directional. The observations reported about the performance of these artefacts, especially in their limitations, provide humanities researchers with information to further improve the theoretical hypotheses about dialogue management. This virtuous circle, represented in Figure 1, benefits both areas and supports both theoretical advancements in linguistics and the development of human-centred technology.



Figure 1: Virtuous collaboration between humanities and technological research areas.

While the rise of GPT-based models has nowadays sparked significant interest in conversational AI, these models' convincing natural language outputs often lead users to mistakenly attribute intelligence to them. However, they struggle with coherence in prolonged interactions and in avoiding fabricated information. Furthermore, despite their capability to generate content in a correct linguistic form, they do not really care about the truthfulness of the utterances. In this sense, it can be affirmed that such models produce utterances towards whose truthfulness the *speaker* has total indifference, mostly because they lack intentionality in doing so. These characteristics, as shown by Hicks et al. (2024), match at least one definition of *bullshit*.

To address these issues, we conversely propose a framework that integrates linguistic theory into conversational AI, moving beyond purely statistical representations of language use. Key concepts such as intentionality and illocutionary force are crucial for providing technological systems with an initial approximation of a *raison d'exprimer* or a reason to communicate. This concept aligns with John Langshaw Austin's notion of the speech act, where "to say something is to do something" (Austin 1962, p. 12). This perspective is similar to Judea Pearl's emphasis on the importance of *doing* over merely *observing* in AI (Pearl and Mackenzie 2018). In contrast, neural approaches to dialogue management only model surface locutive capabilities (the act of producing the utterance) without considering illocutionary force (the reason for producing the utterance). These systems generate language not to influence the world (perlocutionary act) but as a trained response to stimuli. Our approach recommends using generative AI exclusively for language generation, while relying on symbolic AI for reasoning and planning. This aims to ensure genuine communicative intention in dialogues and advance conversational AI by incorporating explicit symbolic models and an orchestration layer for real-time interaction.

Summarising, concerning our main contributions for this work, **C1** articulates a methodological procedure including linguistics theoretical formulations, their formalisation in graph structures, and the use of these structures to implement technological artefacts designed to test and improve the theoretical foundations. Given our previous research, which we will use to support our proposal, we will concentrate on theoretical aspects of linguistics regarding the use of knowledge for pragmatic purposes, such as Common Ground and Common Sense, especially for argumentation-based dialogue (Section 4.1). Points **C2** and **C3** deepen the discussion about the methodological procedure adopted when dealing with linguistic aspects and technological aspects, respectively.

For **C2**, we formalise a graph-based methodology to enhance corpus-based analysis by cross-referencing domain specialised knowledge bases and dialogical corpora and representing linguistic aspects (i.e. CRs, argumentation, explanation) in a way that supports graph data analysis, applicable to linguistics research. Automatically enriching data and capturing latent information is a research approach that has become popular in recent times, with the advent of graph databases and the extensive application of graph analytics. Graphs provide human-readable, multiple-view representations of data supporting cross-referencing among different sources. For linguistic purposes, for example, they allow analysing, at the same time, the linguistic

forms emerging from collected dialogues and the specific usage of domain items for interaction management purposes. At the same time, being a well-known mathematical object, graphs can be smoothly re-used to implement conversational systems (Section 4.2).

As far as **C3** is concerned, given the theoretical foundations for argumentation dialogue management in the form of graph structures, technology provides a useful way to test these theories, detect their weaknesses, and iteratively improve them. From a mathematical point of view, graphs are a very well-known object. Therefore, linguistics theory described in the form of graphs can be transposed in the technological domain seamlessly. As a matter of fact, graph structures, expressed as patterns, take the role of formal definitions of linguistic phenomena or contextual situations. At the same time, graph-based models for decision-making are able to compensate the probabilistic nature of supervised modules trained on the collected corpora. Modelling dialogue management strategies for a machine to follow, implementing the theoretical aspects formalised through linguistic research in graph terms, enables the investigation and definition of *abstract* principles for argumentation-based dialogue management. This aspect has been missing in favour of the development of specific applications, as highlighted by Prakken (2018a) (Section 5).

In Section 3, we will start with a theoretical review of argumentation and argumentation-based dialogue before delving into the aspects that are part of our model.

3 Background theory

Formal and computational argumentation is originally studied in dialogue systems in relation to dialectic. In this area of studies, two main research threads are found: argumentation-based inference and argumentation-based dialogue. Argumentation-based inference concentrates on establishing what conclusions can be reached starting from a possibly incomplete or inconsistent set of information. From a philosophical perspective, models for argumentation-based inference are closer to Hegel's view of dialectic as they are instruments to investigate statements from a strictly logical point of view and they do not involve multiple participants. From a historical perspective, the work presented in Dung (1995) marks the introduction of abstract argumentation frameworks in AI, while the work from Pollock (1987) first established the basis for formal argumentation-based inference (see Prakken (2018a) for a comprehensive history of both argumentation-based inference and dialogue). In the study by Pollock (1987), inference rules are divided into *deductive* and *defeasible* reasons. An argument can be attacked on the basis of its defeasible reasons either by attacking the conclusion of a defeasible inference by means of a conflicting conclusion or by attacking the inference itself without offering alternative solutions. Being based on arguments and *attack* relationships between arguments, inference graphs are used to graphically represent the structure over which conclusions can be drawn about posed statements.

While argumentation-based inference is a formal method for a single entity to decide on the truth of an argument and, therefore, does not consider the problems arising from dialogues among interlocutors, argumentation-based dialogue addresses phenomena that depend on the dynamic exchange of information, which can vary according to turns and participants. In such cases, information is distributed among different agents, who may or may not be willing to share it at different points in time due to individual strategies and goals. This presents challenges both from the perspective of communication protocols, which aim to ensure fairness and efficiency, and from the perspective of behaviour. This dynamic exchange of information, influenced by individual goals and strategies, necessitates a structured approach to understanding the different types of dialogues that can arise in argumentation-based contexts. Adopting a goal-oriented perspective, dialogues have been classified in Walton (1984) and Walton and Krabbe (1995) as follows:

- **Persuasion:** aimed at solving a difference of opinion;
- **Negotiation:** aimed at solving a conflict of interest by reaching a deal;
- **Information seeking:** aimed at information exchange;
- **Deliberation:** aimed at reaching a decision or at establishing a course of action;
- **Inquiry:** aimed at growth of knowledge and agreement *per se*;
- **Quarrel:** aimed at winning a verbal fight or a contest.

These categories are not, however, meant to be absolute, as multiple goals may be present during a single dialogue, and as shifting from one type of dialogue to the other over the course of the interaction is also possible. Persuasion dialogues appear to have been the most studied, in the literature, and have been implemented as a form of intelligent tutoring (Yuan et al. 2004) of procedural justice (Prakken 2008). Formal definitions of logic protocols for argumentation-based dialogue are found in the situation calculus (Brewka 2001), in the event calculus (Bodenstaff et al. 2006), and in C++ (Artikis et al. 2007).

Classic approaches to argumentation-based dialogue adopt the same setting that has been successfully used for argumentation-based inference: i.e. inference rules are derived to establish a course of action that is deterministic given a system configuration. Structural relationships among *claims* and various kinds of *replies* are established in a formal protocol dedicated to establishing whether a speech act is legal or not. This allows us to provide a formal description of situations when a dialogue terminates or, in the case of competitive settings, is *won*. Since persuasion is the most studied situation in argumentation-based dialogues, a typical example of formal communication language is the one described in Prakken (2005). In this type of setting, a *claim* provided by an agent *A* is supported by *data*, constituting an argument that can be explicitly put forward as a reply to a *why* move made by an agent *B*, which explicitly requests the speaker to explain the reasons why a statement should be accepted. Claims can be *attacked* by counter-arguments, which are other claims aimed at proving previous statements as false. *Conceding* and *retracting* moves, respectively, declare the acceptance of a statement or a change of attitude towards it, from commitment to non-commitment. Note that this does not imply a change of *belief*, as it is usually specified that the publicly declared position of an agent may not reflect what the agent actually believes.

An interesting result is found in the framework of deliberation dialogues, where collaboration is assumed in the task of finding an optimal solution to a problem for which none of the involved agents has a solution yet. In the case of a two-agent system adhering strictly to the communication protocol, forming their claims on the basis of their knowledge bases and adopting a collaborative attitude, Black and Atkinson (2010) demonstrated that the agreed solution is always acceptable to both parties. The usefulness of argumentation in dialogue systems designed for deliberation was, instead, demonstrated by Kok et al. (2010).

The problem that characterises argumentation-based dialogue with respect to argumentation-based inference is the presence of different agents in the setting. This introduces multiple, not necessarily aligned, knowledge bases and, possibly, different/conflicting goals in the pursuit of a solution to a problem. There are attempts to deal with the partial knowledge each agent has concerning the others' goals and knowledge using rule-based systems: Dunne and Bench-Capon (2006) examines the consequences of having *suspensions of hidden agenda* in the case of negotiation based dialogues while, in Kok (2013), the strategic usefulness of reinforcing an agent's own claims versus the usefulness of undermining the other agents' claims is considered. These approaches, however, have been recently surpassed by more flexible, probabilistic approaches, modelling opponents in terms of probability distributions over their possible beliefs and goals and using these to compute the utility of each legal dialogue move depending on their own goals and beliefs (e.g. Hadjinikolis et al. (2013); Rienstra et al. (2013)). Moreover, other works put forward the need to model the *degree* or strength of an agent's belief towards a statement, modelled as the probability of the statement being true, rather than assuming it to either be or not be true (Hunter and Thimm 2016, 2017). Some recent studies concentrating on the analysis of argumentative structures have been presented (Budzynska and Reed 2011, Visser et al. 2020, Hautli-Janisz et al. 2022), as a part of the inference anchory theory. These works present annotation methods to analyse the relationship between speech acts linked by argumentation dynamics. While these works use a graph-based representation for this network, relationships are established between speech acts only. The approach we propose here aims at describing relationships between utterances and the relevant knowledge domain in a more descriptive rather than interpretative way, leveraging on network analysis methodologies to make conversation dynamics emerge. Ideally, recurrent patterns emerging from this kind of graph can be linked to complex linguistic phenomena, like argumentation, and constitute a formal definition of these.

In this view, the emergence of communication strategies may depend on three *pragmatic* factors:

- **Beliefs:** collected knowledge (i.e. *Communal Common Ground* in Section 4.1.1) organised into a stable structure (i.e. a graph);

- **Contextual observation:** evidence from the perceived world (i.e. *Personal Common Ground* in Section 4.1.1) that may or may not contradict current beliefs;
- **Goals:** conscious, unconscious, and pseudo-goals changing the subjective view about the relationship between observed reality and current beliefs.

As we will see in Section 4, an argumentation-based dialogue system should be provided with the representation of the aforementioned types of knowledge. Most importantly, such goal-oriented systems are designed to make use of contextual information upon which beliefs are constructed and to assist users in accomplishing specific tasks or goals. Furthermore, they can apply different linguistic strategies (i.e. argumentation, explanation, etc.) to exhibit various types of intentionality. This capability to select a strategy to pursue a specific goal implies a form of reasoning that is lacking in other types of systems, such as in GPT-based ones. As a matter of fact, despite excelling in human-like text response generation in a conversational manner, this type of behaviour is not intentional and does not subtend the *intelligent* pursuit of an objective.

The way the interaction with other agents is shaped, at this point, it would be controlled by the affective appraisal given by the different combinations of the three aforementioned factors involved in the evaluation of the current situation. Further inquiries about a topic, for example, would be dictated by the unpleasant feeling given by the detection of inconsistent beliefs, as people naturally attempt to eliminate or reduce them (Elliot and Devine 1994).

For ARDA, we opted for the recommendation task as it offers an ideal setting to examine the theoretical model that underpins its computational implementation. The recommendation task is, indeed, particularly suitable for this study due to its inner dialogical structure and the defined goal it encompasses. This task revolves around a clear dialogical pattern that involves two distinct phases, exploration and exploitation (E&E). These phases can be viewed as two intertwined types of dialogues: exploration refers here to the system gathering information about unfamiliar information; conversely, during the exploitation phase, the system capitalises on the best-known option (Gao et al. 2021). Throughout such interactions, the goals at play continually evolve in accordance with the current dialogue state. The primary objective of a recommendation dialogue is to achieve agreement between participants during the exploitation phase, with a focus on selecting a specific item supported by compelling arguments. Meanwhile, in the exploration phase, the secondary goal revolves around the need to identify and select these supporting arguments while establishing a Common Ground (Clark 1996).

In the following sections, we summarise our theoretical approach to tackle different kinds of problems related to the development of dialogue systems. This is intended to propose a unified theoretical framework and apply computational approach to the framework of argumentation-based dialogue aimed at surpassing dialogue systems equipped with machine learning techniques only to conduct dialogue without relying on pragmatic aspects.

4 Linguistically oriented resources and insights as expressive networks (LORIEN)

LORIEN focuses on defining a methodology that uses graphs to analyse, represent, and extract semantic and pragmatic information from dialogue data. Since pragmatics primarily deals with phenomena that are contextually determined, context representation – both situational and communicative – becomes pivotal. For instance, knowledge graphs are a popular choice for representing domain information in technological applications. Precisely, managing argumentation-based dialogue requires an efficient and flexible tool, like graphs, to handle the inherent complexity and dynamism of the knowledge involved. Graphs are particularly suitable to represent and visualise the relationships between various arguments, counterarguments, and supporting evidence (Lei et al. 2020, Deng et al. 2021, Marro et al. 2022). This is crucial because it allows simple tracking of how different pieces of information interconnect, which is essential in understanding and evaluating arguments. In addition, graphs can accommodate rapidly changing information by allowing nodes and

edges to be added, removed, or modified, making them highly adaptable to the evolving nature of dialogue. This flexibility ensures that the knowledge base remains up to date and accurately reflects the state of the interaction, facilitating more effective organisation, retrieval, and querying of information (Di Maro et al. 2021, Origlia et al. 2022). Modern approaches to knowledge representation are often based on the concept of graphs and implemented through graph databases. In this work, we make use of Neo4j (Webber 2012) for our examples. Neo4j is an open source graph database manager that has been developed over the last 16 years and applied to a high number of tasks related to data representation. The underpinning of the knowledge graphs presented in this work is found in Linked Open Data (LOD).

Dialogue-based applications are often based on knowledge graphs, as they are useful to represent and study the relationship between how people make use of the domain during human–human dialogues. This supports modelling how a machine should mimic their behaviour. Explainable-by-design systems do not rely only on machine learning to make these strategies emerge so it is important to cross-reference recorded dialogues with the background knowledge concepts they use. One of the case studies we adopt to explore how linguistics research methods can be directly linked to dialogue systems development is the movie recommendation task. This has been deeply explored, in the past, and there are a lot of available resources on which to test data collection, annotation, and analysis. In our case, our ongoing work aims at formalising a methodology to organise both data and linguistic knowledge as graphs connecting multiple resources, to extract more information from the combination of the resources.

For the presented framework, we start by importing common knowledge from LOD sources, such as Wikidata² (Mora-Cantallos et al. 2019). Alternative sources are, then, cross-referenced. In the movie recommendation domain, for example, we extract movies, people that worked in those movies, the genres they belong to, the awards they won and the people they were awarded to. This sub-graph represents what, in linguistics, is referred to as the *Communal Common Ground (CCG)*, whose details will be provided in Section 4.1.1. Briefly, it represents encyclopaedic knowledge about the domain that can be considered objective and generally available. Next, we import human–human dialogues, representing each turn as a node and linking them through chains of relationships. Also, we link the utterances to the elements of the CCG they refer to. Since dialogues collected from a corpus of actual human–machine interactions cannot be considered common domain knowledge, they represent the *Personal Experience (PE)*, the system has about dialogue management in the domain of interest. The dialogues sub-graph, once again, corresponds to a linguistic concept to support theoretical explainability of the model and, later, of the application behaviour. For instance, Origlia et al. (2022) applied a graph-based methodology involving a multi-source graph that integrates both domain and dialogue information to analyse argumentative strategies in recommendation dialogues. First, the multi-source graph was enriched with additional insights extracted from the data it contained. Specifically, we used the following: (i) Plot similarities to calculate content-based similarities between movies, which are useful for verifying whether recommendations could be grounded in content-related parameters. (ii) PageRank to resolve ambiguities in references to people and movies (e.g. homonyms such as *Chris Evans*) and to determine the relative importance of nodes in the graph. (iii) Node embeddings to compactly represent nodes, capturing both structural and semantic information. With the graph thus structured, the dialogue analysis yielded the following observations: (a) Patterns in recommendation dialogues exhibited a strong bias towards recent movies, underscoring the significance of temporal context in argument selection. (b) Argumentation strategies employed by recommenders, as analysed through the graph, demonstrated how leveraging graph data can enhance the overall effectiveness of recommendations. (c) The use of path finding queries on data containing both knowledge representation and dialogical interactions highlights how speakers make use of the knowledge throughout the dialogue in the form of RING-like patterns. These points help researchers understand why people talk about specific domain items in specific moments, supporting the investigation of a *raison d’exprimer*.

In the next sections, we will describe, on the one hand, the importance of specific sets of knowledge explicitly and implicitly used to convey meaning in dialogue and how to represent them (Section 4.1). On the other hand, the use of such representations will be applied to provide for pragmatic skills related to the ability

² <https://www.wikidata.org/>.

to ask for information, to explain, and to argument (Section 4.2). These are the foundations of LORIEN and its theory-based graph representations enabling dialogue systems to exhibit pragmatic argumentative skills.

4.1 Representing knowledge for conversation

When they communicate, speakers make use of different types of knowledge. First, they need *linguistic competence* in order to communicate. This is what Coseriu (1985) called *Linguistic Knowledge* and, before that, De Saussure (2004) called *langue*, which encompasses the items of knowledge that enable a speaker to make effective use of word-signs. This includes the use of the lexicon and syntactic structures. In this sense, we can say that this type of knowledge includes, on the one hand, semantic knowledge – i.e. the knowledge about the literal meaning of words and sentences, as well as knowledge of facts and concepts, which helps in understanding the content and conveying information accurately – and, on the other hand, syntactic knowledge, which pertains to the rules and structures that govern sentence formation and grammar. As speakers of a specific language, we also have *Metalinguistic Knowledge*, which is the knowledge about language itself, including its structure, usage, and how it functions. This enables speakers to reflect on and discuss language explicitly.

Beyond the purely linguistic competence that enables us to use a language correctly, when we communicate, language itself is strictly connected to the context of use. In this sense, we need additional skills to understand how the language is used according to the context and to understand how we can achieve our communicative goals. In this sense, we talk about *communicative competence*. Hymes (1972) described the communicative competence of speakers “as the ability to know when to speak, when not, what to talk or not talk about, with whom, when, where, and in what manner.” This *Pragmatic Knowledge* is related to speaker performance and the individual and contextual ability to use the *parole* (De Saussure 2004) in communication. Strictly connected to the pragmatic knowledge, there is the ability to use and identify *speech acts* (Austin 1962), namely, the ability to understand the intentions behind utterances (e.g. requests, promises, suggestions). The understanding of such intentions also relies on the identification of *implicatures*, i.e. implied meanings that go beyond literal interpretation, and *presuppositions*, i.e. assumptions made by speakers based on shared knowledge. Such shared knowledge, moreover, is part of another important set of information to consider in dialogue, that is Common Ground, which we will refer to in Section 4.1.1 (Clark 1996). Other important sets of knowledge are represented by *Contextual Information* that uses situational cues (e.g. time, place, social roles) to interpret meaning, *Domain Knowledge* that describes the information pertaining a specific field, and *Discourse Knowledge* that involves understanding how utterances relate to each other within a larger conversation or discourse (e.g. coherence and cohesion).

Finally, as part of a society, speakers also need *Socio-Cultural Knowledge* which refers to the awareness of social norms, cultural practices, and conventions that influence language use (Gumperz 1979). This knowledge helps in understanding appropriate ways to communicate based on cultural context and social expectations. In contrast, *Common Sense* consists of practical, everyday reasoning and judgements shared widely within a community, often based on immediate perceptions and intuitive understanding rather than formal learning. While Socio-Cultural Knowledge is more structured and context-specific, Common Sense is more general and universally applied (Section 4.1.2). The two are related in that Socio-Cultural Knowledge can inform and enhance Common Sense by providing a deeper contextual framework, helping individuals navigate and interpret everyday situations more effectively within a specific cultural context.

In conversation, speakers draw on these various types of knowledge simultaneously to convey messages effectively, understand each other, and navigate the complexities of communication. For the development of ARDA, we specifically focused on some of the aforementioned sets of knowledge belonging to the communicative competence, specifically Domain Knowledge, Common Ground, and Common Sense. These are represented as knowledge graphs to extract information and to use their representational structure to identify specific patterns on which to study and apply pragmatic strategies (Section 4.2).

4.1.1 Common Ground

Stalnaker (2002) defined the notion of Common Ground as the sum of interlocutors' mutual, common, or joint beliefs and knowledge. Since Grice (1975), the importance of cooperation in a successful conversation has been pointed out. In Grice (1989, p. 65), the term of Common Ground was introduced as related to communicative processes. In fact, participants in a conversation must have grounded knowledge to understand each other. The process of *grounding* takes place in dialogue when the interlocutors update their Common Ground by accumulating information in the perceived Common Ground. In Clark and Schaefer (1989), the classical model of grounding is illustrated: dialogue participants reach their mutual belief by checking the mutual understanding. This is accomplished through *contributions*, corresponding to the communicative actions collected through dialogue. Contributions can be divided into *presentation phase* and *acceptance phase*. During the presentation phase, the utterance is presented, whereas in the acceptance phase, the utterance is accepted by the interlocutor as understood. The utterance acceptance or refusal is signalled via diverse types of feedback. The refusal, for instance, can depend on different aspects, such as acoustic, semantic, or intentional misunderstanding. According to Allwood et al. (1992, p. 4–5), feedback is indeed a linguistic mechanism which enables interlocutors to exchange information about four different basic communicative functions: (i) contact (i.e. feedback expressing the will and/or ability to continue the interaction), (ii) perception (i.e. feedback referring to the will and/or ability to perceive the message), (iii) understanding (i.e. feedback about the will and/or ability to understand the message), (iv) attitudinal reactions (i.e. feedback referring to the will and/or ability to react and respond appropriately). In Section 4.2.1, these functions will be adopted to identify specific level of analysis for the use of grounding-related corrective feedback, such as CRs.

Common Ground, as acknowledged in Clark (2015), can be of four main types: personal, local, communal, and specialised. In this work, we focus on Personal Communal Ground (PCG) and CCG. PCG is established collecting information over time through communicative exchanges with an interlocutor, and it can be considered as a record of shared experiences with that person. This specific set of information can also be considered, as in this work, as part of what builds the PE of an interlocutor, useful in the future steps of the current interaction or for future exchanges. CCG, conversely, refers to the amount of information shared with people belonging to the same community, such as general knowledge, knowledge about social background, education, religion, nationality, and language(s).

To represent such knowledge, we use graph structures that allow us to investigate how linguistic concepts can be used to describe with deeper details the recurring structures forming between dialogues (PE) and domain knowledge (CCG). As already mentioned, in LORIEN, we propose the use of graphs to represent different types of Common Ground and their inter-dependency in the extraction of information for argumentation purposes (Figure 4). Although PE is represented by dialogues stored in the graph, CCG is represented by a more general knowledge, or in other words by information coming from different resources like LOD. To be more precise, the CCG encompasses entities and their relationships within a specific field of knowledge. For our case study, we choose the movie industry. Here, the information concerning, for instance, the fact that the director *Peter Jackson* directed the movie *The Lord of the Rings* can be connected by a relationship like [DIRECTED] (Figure 2).

Conversely, the PCG is represented by the beliefs created from the information collected during the dialogue and stored in the graph. These are, furthermore, connected with specific relationships to particular entities belonging to the CCG which are referred to during the interaction. For instance, for the movie domain application, the fact that a seeker stated that they like the actor *Ian McKellen* is linked to the corresponding



Figure 2: An example of domain knowledge information.

node entity of the actor which is in turn related to other nodes (i.e. other movies he is starring in). The PCG, indeed, keeps track of the interaction progression with a specific user, allowing for the identification of personalised strategies for future engagements and recommendation purposes (e.g. “User X is interested in fantasy movies; the recommender suggests that the seeker might therefore watch *The Lord of the Rings*”), as shown in Figure 3. This sub-graph maintains a provisional status. It is eventually incorporated into the PE, which contains information regarding entities and their relationships obtained from diverse interactions. As such, the PE sub-graph supplements the information contained within the CCG, for other interactions or various phases of interaction (Di Bratto et al. 2021). Since it is never possible to state that a certain user has created a certain beliefs, in LORIEN beliefs are not explicitly shown in a graph. They are going to be explicitly represented when the system knowledge is taken into account (Section 5). Beliefs representing the PCG can be constructed based on the information of the CCG (e.g. *I believe that Ian McKellen starred in The Lord of the Rings which is a Fantasy movie*) and on the information shared during the interaction (e.g. *I believe the user likes Ian McKellen and Fantasy movies*), from which the recommender can extract their recommendation (e.g. *Then I recommend you watch The Lord of the Rings*). In Figure 4, the relationship between the aforementioned sets of knowledge is illustrated.

4.1.2 Common sense

Common sense is a multifaceted and complex construct that underpins a vast spectrum of intelligent activities, including natural language processing, planning, and learning (Davis 2014), typically acquired through everyday experiences and interactions with the world. It embodies the ability to infer conclusions, i.e. to deduce implications from what is already known (memory stock) and from information given (input flow) (Bauer 2024). Since it denotes inherently evident truths, it does not require explicit justification, leading to its implicit use in communication, both written and oral (Grice 1975). It only surfaces explicitly during ambiguous situations or when the speaker necessitates clarification (Nguyen et al. 2022). In fact, it is commonly assumed that one’s own understanding of concepts is generally shared by others, especially among individuals deemed to possess rationality, constituting the Common Sense Knowledge (CSK) (Rosenfeld 2011). CSK refers to the basic level of practical knowledge and reasoning, encompassing (i) information regarding events unfolding in

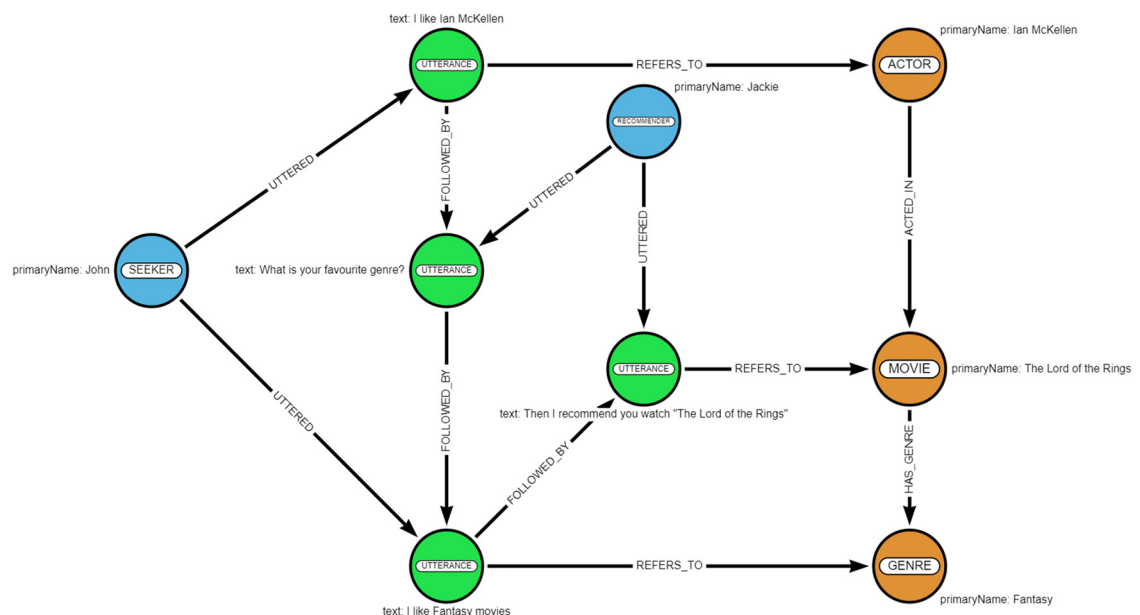


Figure 3: An extract of the graph structure, representing the domain information of the CCG (in brown) and the dialogue exchange (in green) between interlocutors (in blue). Given this structure, it is possible to estimate the PCG of the participants.

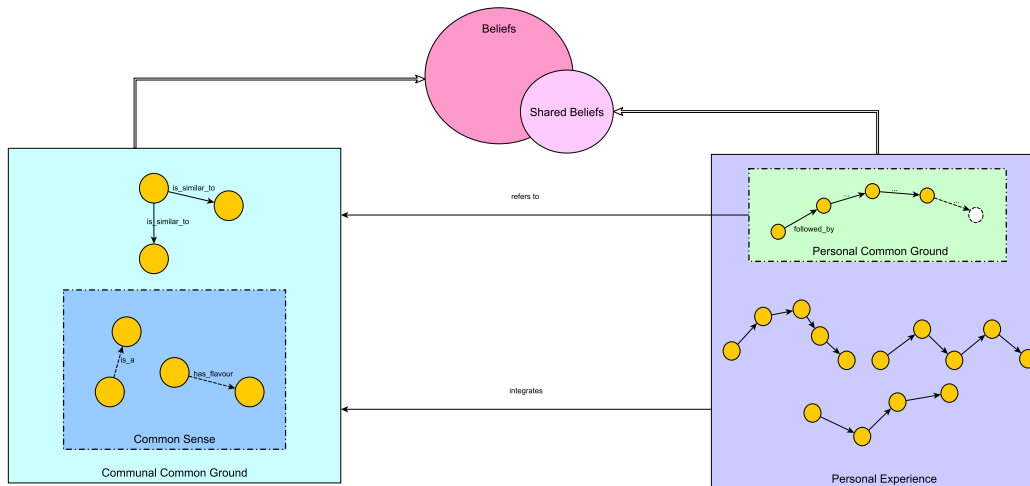


Figure 4: Knowledge graph partitioned into CCG, PCG, PE, and beliefs.

time, (ii) the consequences of actions undertaken by the individual and others, (iii) the characteristics of physical objects, (iv) perception, (v) properties, and (vi) interrelationships. For instance, the fact that an oval object is composed of a *yolk*, a *white* and a *shell*, allows it to be recognised as an *egg*, along with the effects of boiling and dropping. This knowledge is based on relationships between words, concepts, sentences, and thoughts and enables people to communicate with each other and deal with problems that affect everyday life (Cambria et al. 2009). Since CSK has been proven to be difficult to define, several researchers have attempted to limit the scope of investigation by identifying the most representative features that can provide a comprehensive description of this type of knowledge. The case study conducted by Zang et al. (2013), proposes six characteristics:

- Share: A group of people own and share the CSK;
- Fundamentality: People have a good understanding of the CSK and tend to take it for granted;
- Implicitness: Most of the time, people tend not to mention the CSK explicitly, as it is shared knowledge;
- Large-Scale: The CSK contains massive, large-scale information;
- Open-Domain: The CSK is broad in nature and covers all aspects of everyday life rather than a specific domain;
- Default: The CSK includes predefined assumptions about typical cases of everyday life; therefore, most of them may not always be correct.

These features provide a foundation for outlining a comprehensive framework that can illustrate the CSK. Although the concept of Common Sense may share some similarities with the concept of CCG, they are essentially distinct concepts. As it was mentioned in Section 4.1.1, CCG refers to a set of knowledge shared among individuals belonging to the same community (Clark 1996). CCG can be based on, but not limited to, the CSK: while CCG involves a connection between an individual and others within a shared community, in other words, it is partner-specific, CSK pertains to an individual's interaction with the world at large as it is typically shared unconsciously and implicitly. CCG involves an agreement among speakers, establishing a set of shared beliefs, involving active agreement among members of a group. This process helps define the identity and boundaries of the group as well as establish a common language (MacWhinney and O'Grady 2015). In contrast, the establishment of a shared agreement is unnecessary for CSK, as it is presumed to be already universally shared among the speakers. Despite the presence of shared understandings and assumptions in both CCG and CSK, the two concepts can be differentiated based on the process of agreement required to establish them.

Given that CSK has been considered advantageous to systems (McCarthy 1984), providing a comprehensive representation of it was needed. Despite efforts to represent and organise CSK for computational purposes numerous challenges were encountered, in terms of scalability and creation of comprehensive and accurate datasets (Lenat 1995; Sap et al. 2019), and of the possible integration of diverse sources managing

inconsistencies and gaps among them (Zhou et al. 2021). In our view, one possible interesting aspect of Common Sense lies in *considering it as a process rather than a representation*. Indeed, it is difficult, if not impossible, to be able to encapsulate all human knowledge in a resource (Brooks 1991), while it would be more interesting to define a method of deducing and constructing this knowledge from collected and organised data, similarly to what happens with experience. More specifically, as represented by the dotted arrows in Figure 4, Common Sense is not necessarily represented by relationships established in a knowledge base but can be derived and reconstructed using probabilistic approaches. Common sense, then, becomes any kind of graph-based representation that a speaker believes to exist in the knowledge possessed by the interlocutor with a high probability. Although modern state-of-the-art LLMs have made remarkable strides in encapsulating vast amounts of human knowledge, drawing from extensive datasets that cover diverse domains their capabilities are still limited, particularly when it comes to hallucinations, producing plausible-sounding but inaccurate or nonsensical information. This highlights that while LLMs approach a form of generalised knowledge representation, they still fall short of perfect accuracy and reliability. Moreover, despite being able to drawing from vast amount of data, they have been shown to lack formal reasoning capabilities (Mirzadeh et al. 2024).

In our framework, Common Sense is, therefore, extracted from the graph structure. For instance, the fact that a movie must have a director who directed it in order to exist is derived from the probability of the relationship linking the movie and director nodes. However, the specific knowledge, namely, the fact that a director directed a movie, represents a piece of information which is grounded or eligible to be part of the Common Ground, the type of relation between the uttered concepts and the type of concept themselves represent what is defined as CSK, as it is usually not verbalised but given for granted (Mennella et al. 2023). Common Sense can, consequently, be applied in conversation to overcome ambiguous situations and to request needed information where it is not yet clear or made explicit. For illustrative purposes, let us consider the situation where Common Sense is able to guide the speaker's reasoning in asking who the director of a specific movie is instead of asking whether or not the movie has a director. This ability to leverage CSK in conversations helps streamline communication and enhances the overall efficiency of information exchange.

4.2 Graph-based pragmatics

Having defined some specific types of knowledge used in dialogical exchange, we now focus on some typical pragmatic strategies of conversation that make use of these knowledge sets. The technique exemplified in this Section concerns point **C2** of our proposal. In fact, considering the representation of knowledge in the form of a graph, we will describe the use of these strategies by exploiting the graph structure in extracting useful information and patterns. More specifically, we provide here a selection of pragmatic skills in the form of graph patterns, such as the capability of identifying communicative problems and elaborating appropriate CRs, of explaining decisions and reasoning, and of selecting the most plausible arguments.

4.2.1 CRs

Clarification is a fundamental part of the grounding process introduced in Section 4.1. CRs are pragmatic tools used by interlocutors to ensure mutual understanding (Ginzburg and Macura 2005), when, for example, a speaker did not (fully) understand or is uncertain about what was previously said or meant with an utterance (Gabsdil 2003). According to Purver (2004, 2006), interlocutors make use of CRs, or anaphoric feedback, when there is a problem in processing the previous utterance. As Clark (1996) points out, to pursue the goal of succeeding in their joint activity of communicating, interlocutors must ensure that what is being communicated is also correctly understood. To do this, several strategies can be exploited, such as the use of linguistic and paralinguistic feedback (Traum et al. 1999), among which we also find CRs. Among scholars who have classified different types of CRs, Purver (2004) classifies them according to the surface form and the compromised element on which the request is built. This classification, although very detailed, does not include

information such as the causes and problems that trigger the need for such requests. In the study by Rodriguez and Schlangen (2004), the notion of a problem causing the instantiation of a CR is, on the other hand, extensively explored. Here, different types of problems, such as acoustic or lexical problems, are considered determinant for the use of a given functionally and formally different CR.

Based on these studies and on the analysis of a dialogic corpus, a hierarchical classification of CRs is proposed. Starting from Allwood et al. (1992) (see also Section 4.2), four basic communicative functions were defined, corresponding to the communicative levels of contact, perception, comprehension, and intention (Maro 2021). On each of these levels, one or more *problems* may occur, triggered by specific linguistic and/or informational issues (*triggers*). Moreover, CRs can also occur in different *forms*: open questions (WQ), alternatives (AQ), polar positive (BroadPQ, NarrowPQ), polar negative (specifically, low negative polar questions LNPQ and high negative polar questions HNPQ), and declarative sentences. Each formulation can convey a specific *function* and refer to a problematic item in the previous utterance (*compromised item*). Taking the level of Understanding as an example (Table 1), we identify five different problems:

- *Lexical Understanding*: presence of lexical items that are unknown or ambiguous to the learner, who then asks for clarification.
- *Reference Reconstruction*: uncertainties occur in the resolution of an anaphora or extra-linguistic reference; it may refer to a nominal phrase, a deictic element, or an action.
- *Syntactic Understanding*: ambiguities may be caused by the different meanings associated with particular syntactic structures.
- *Logical Understanding*: occurs when the cause–effect relationship is unclear.
- *Information Processing*: the information received is not sufficient for the entire understanding of the message, or the information entered into the Common Ground must be verified by confirmation or clarification, due to the presence of a possible inconsistency.

To verify the validity of the proposed classification, the Bielefeld Speech and Gestures Alignment Corpus (SaGA) (Lucking et al. 2010, 2012), a German-language corpus of 25 multimodal dialogues of interlocutors engaged in a spatial communication task, was analysed. The CRs were annotated in ELAN (Brugman et al. 2004) regarding the labels of problem, trigger, form, function, and compromised item. For CRs related to Common Ground (Table 1), original bias information (Ladd 1981) and contextual evidence (Buring and Gunlogson 2000) were also taken into account. These were important, as they were useful for later comparison of

Table 1: Part of the CR classification concerning *Understanding* problems we focus on in this work (Maro 2021); PQ refers to polar question, WQ to WH-questions, AQ to alternative questions, HNPQ to high negative polar questions, and LNPQ to low negative polar questions

Communication level	Problem	Trigger	Form	Function	Compromised item
Understanding	Lexical understanding	Unknown_meaning	WQ	Explanation	Clause
		Meaning_Ambiguity	AQ	Metalinguistic function	Constituents
			BroadPQ	Disambiguation	External
	Reference reconstruction	NP_Reference	NarrowPQ	Confirmation	Presupposition
		Deictic_Reference	TagPQ	Interactional	
		Action_Reference	LNPQ		
		Elliptical_Ambiguity	HNPQ		
	Syntactic understanding	Analytical_Ambiguity	TagNPQ	(Imperative)	
		Attachment_Ambiguity	Declarative		
		Coordination_Ambiguity			
	Logical understanding	Elliptical_Ambiguity			
		Cause_Effect			
	Information processing	Missing_Information			
		Common Ground			

For further details, refer to Di Maro et al. (2021).

bias-evidence conflicts (here related to the function expressed by Common Ground CRs) with other work available for the German language (Domaneschi et al. 2017).

Annotation of the corpus was carried out by two annotators, a linguist and a computer science student. The annotation levels considered for agreement calculation are Trigger, Form, and Compromised Item. Bias was not considered because of difficulty of interpretation. For this aspect, further experiments were carried out in Di Maro et al. (2021). According to the agreement scores between annotators, calculated with Cohen's Kappa (Cohen 1960), the agreement is substantial (0.7). In the 25 available dialogues of the SaGA corpus, 201 CRs were annotated. Their distribution is summarised in Figure 5.

The identified CRs all belong to the Understanding level. In fact, since for the corpus collection the spatial task is straightforward, the communication channel was clearly open and the perception was constantly checked through the use of back-channels. Furthermore, no intention-related problems were found, as the goal of the interaction was clear to the participants. Some comprehension triggers, especially syntactic ones, were not found since the task was not ambiguous overall. On the other hand, Miss_Inf and CoG CRs are the most frequent classes. This is related to the nature of the task, in which it was essential to check the completeness and consistency of the information received. Both classes occurred in different syntactic forms. While Common Ground are mostly elaborated as polar questions (more specifically, high negative polar questions), Miss_Inf are also formulated as alternative and open-ended questions (Maro 2021; Di Maro et al. 2021).

This fine-grained classification and its hierarchical functioning are implemented in ARDA, as communication problems are managed in the form of behaviour trees (BTs) (Section 5.1), so that the system knows which pragmatic-dialogical task has to be performed and in which order to ensure mutual understanding with the human interlocutor and to reach its communicative goal. For instance, in our movie domain case study, the inconsistency could occur and be detected in the form of a graph path connecting a positive belief (i.e. *like*) related to a genre node, such as *horror*, and a new collected negative belief related to a movie node, i.e. *Hereditary*. This result in what we call Common Ground Inconsistency, defined as the incompatibility between the listener belief and the new evidence provided by the speaker (Di Maro et al. 2021). In our example, this inconsistency might cause a questioning of the previously constructed belief and the uttering of a corresponding appropriate CR (i.e. *Didn't you like horror movies?*). This leads to the formation of argument patterns for reaching an agreement in the conversation (i.e. deliberation and information sharing). For further transparency and robustness of the system, CRs can also be used in combination with explanations, as it will be shown in the next section.

Besides the aforementioned work on CRs, recent work explored various types and applications of CRs. For example, in the study by Chiyah-Garcia et al. (2023), a dataset was adopted to train multi-modal systems to manage referential ambiguities (Chiyah-Garcia et al. 2023). Conversely, in the study by Higashinaka et al.

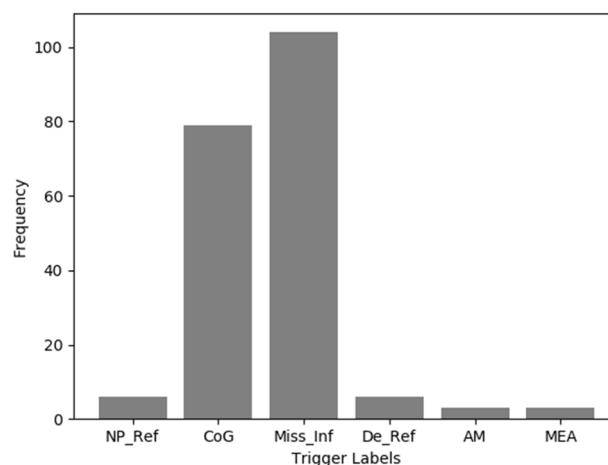


Figure 5: CR distribution (Understanding class) in the SaGA corpus: NP_Ref stands for noun phrase reference, CoG for common ground, Miss_Inf for missing information, AM for ambiguity, and MEA for meaning.

(2021), a precise and compelling taxonomy of errors triggering clarification was illustrated. However, our objective here is not only to provide a comprehensive overview of different types of ambiguous, unclear, or incoherent situations but also to develop a formal method to manage them, on the one hand, and to describe and generate appropriate question forms accordingly, on the other hand. In addition, in the study by Deng et al. (2023), LLMs were also used to evaluate proactive dialogues, in which strategies like clarifications are needed. Nevertheless, the authors pointed out that LLM-based dialogue systems barely ask clarification questions when encountering ambiguous queries. For this reason, the use of such a hierarchy in our framework has the objective to address this issue by defining, based on linguistic theories and analysis, what the authors call (i) a *Clarification Need Prediction*, identifying the need for clarification in the current turn, that we base on graph configurations mapped onto branches of a BT (Section 5); and (ii) a *Clarification Question Generation*, which generates an appropriate clarifying question if needed, as described by the theory and corpus analysis.

4.2.2 Explanatory acts

From a linguistic point of view, explanations are not a piece of cake to define. In fact, multiple linguistic moves can have an explanatory function depending on the linguistic context, the communicative situation, the linguistic form, and the content of the act (Sbisa 1987). Generally, we can define an explanation as a sequence of linguistic acts that exhibit a relationship between *explanans* and *explanandum*. While the *explanans* represents the explanatory schema (i.e. *This child is immortal*), the *explanandum* is what is to be explained. The *explanans* can be formed from presupposed elements that are part of Common Ground (i.e. *Arwen's father is an elf*), or from deduced facts based on Common Sense (i.e. *An elf father has an immortal child*).

In ARDA, explanations are used to reveal the internal state of the system, specifically addressing two main aspects: (i) explanations of the problem and (ii) explanations of the system's behaviour.

For the first case, explanations of the problem can be explicitly or implicitly expressed (i.e. explanation vs CR) and are based on Common Sense and/or Common Ground principles. These explanations aim to disclose the internal beliefs structure of the system because of issues detected in it, making it understandable based on shared knowledge and logical reasoning. Common Sense and Common Ground complement and overlap in conversation and are of fundamental importance, in both achieving communicative purposes and in generating explanations. In the study by Di Maro et al. (2021), for example, the collection of information in the Common Ground graph makes it possible to identify possible inconsistencies among the information received (Di Maro et al. 2021). These inconsistencies are, in turn, based on pre-conditions and post-conditions, described as properties of the graph action nodes, which can be then deduced on the basis of reasoning represented by Common Sense. In this context, where the user produces an utterance such as *Melt the butter* followed by another utterance such as *Cut the butter*, the inconsistency is highlighted through the use of a CR in the form of negative polar question (i.e. *Shouldn't I have melted the butter?*). The conflict arises therefore between: (i) a piece of information that is part of Common Ground, such as *Melt the butter*, whose post-condition is *Ingredient becomes Liquid*, and (ii) a piece of information that is intended to become part of Common Ground, such as *Cut the butter*, whose pre-condition is *Ingredient is Solid*. This is motivated on the base of the world facts that are part of Common Sense, for instance:

- (a) if one melts something solid, it becomes liquid,
- (b) one cannot cut something that is liquid.

The adoption of this information also makes it possible to generate an explanation concerning the impossibility of fulfilment of the uttered action, as in *I can't. Butter is liquid*.

In the second case, explanations of the system's behaviour, following user's requests, can focus on the system desirability and goal orientation. These explanations disclose the internal configuration of the system that led it to act in a certain way. These speech acts link actions to the system's goals and highlight how the behaviour might be desirable for the system, thus providing a clear rationale for the system's decisions, as in Stange et al. (2019). This is fundamentally different from asking an LLM for explanation because the answer will not disclose the internal state of the statistical model but rather provide an estimate of what most of the people in the training set would say to *justify* those actions (Saba 2023).

Some scholars have proposed other possible classifications for explanations. In the study by Von Wright (1997), for example, four different types of explanations are defined as follows:

- Causal explanation, in which the *explanans* set forth a sufficient or necessary condition of the state of affairs or event to be explained. It is typical in the physical sciences.
- Teleological explanation, in which the *explanans* set forth the goal of the behaviour to be explained. It is typical of action explanation.
- Quasi-causal explanation, in which the *explanans* set forth some circumstance that is related to the state of affairs, event, or behaviour to be explained. It can be found in historical explanations.
- Quasi-teleological explanation, in which the *explanans* expose some effect with respect to which the state of affairs or event to be explained is a necessary condition. It is typical of the biological sciences.

In our perspective, Common Ground-based explanations can be considered as *Quasi-causal*, in that the circumstances regarding shared information are considered (i.e. *I cannot cut the butter because you told me to melt it*). On the other hand, *Causal* explanations can be mapped on the explanation based on Common Sense reasoning, as for pre-conditions and post-conditions compatibility, since necessary conditions are taken into account. As far as *Quasi-teleological* explanations are concerned, they are closer to question-answering utterances as the illocutionary strength is weaker than argumentation. *Teleological* explanations, on the other hand, can be mapped on disclosure speech acts about the system's behaviour, where the higher strength of the illocutionary force makes it more similar to justification in argumentation (Walton 2005; Canary and Seibold 2010). This type of explanation is widely used in AI where such speech acts are used to explain decisions (Miller 2019). The weight of the goal in the decision about how to structure the explanation introduces argumentative features. A total objective disclosure statement is a pure explanation, while an explanation depending on the pursuit of personal goals becomes more argumentative and therefore with a higher illocutionary strength (Table 2). Refer to Section 5.1 and Figure 11 for further details.

Specifically, we believe that the difference between argumentation and explanation lies in their position on a continuum where:

- On the one side, we have speech acts guided by pseudo-goals (Miceli and Castelfranchi 2014), like the need to check mutual understanding and to check for consistency, that show a lower level of illocutionary force, as for Common Ground inconsistencies and explanations.
- On the other side of the continuum, we find argumentation, for which goals proper (Miceli and Castelfranchi 2014) are set. These can be divided between interlocutor goals that need to be accommodated by the speaker, and personal goals belonging to the speaker themselves. Argumentative acts are, for example, used to reach the latter.

In other words, while explanatory acts have a clarifying function that enable the recipient to better understand something, with argumentation acts the proponent proposes reasons for the recipient to come to accept or refute a certain thesis (Asterhan and Schwarz 2009). Furthermore, argumentation can be used to support explanations and explanations can be used to argument. Since the difference between one level of intentionality and another is not always clear from a perception perspective, we decided to model this difference in production. In fact, according to the type of goal the system want to perform, we will have one act or the other. This is modelled in our BTs, as it will be described in Section 5.1. Specifically, for our movie recommendation domain case study, we make use of the distinction between explanation or argumentation to choose when to explain a retrieved inconsistency or an

Table 2: Types of explanation classes mapped on speech acts ordered by increasing illocutionary strength

Explanation class	Speech act
Quasi-causal	Common ground based
Causal	Common sense based
Quasi-teleological	Question-answering
Teleological	Justification

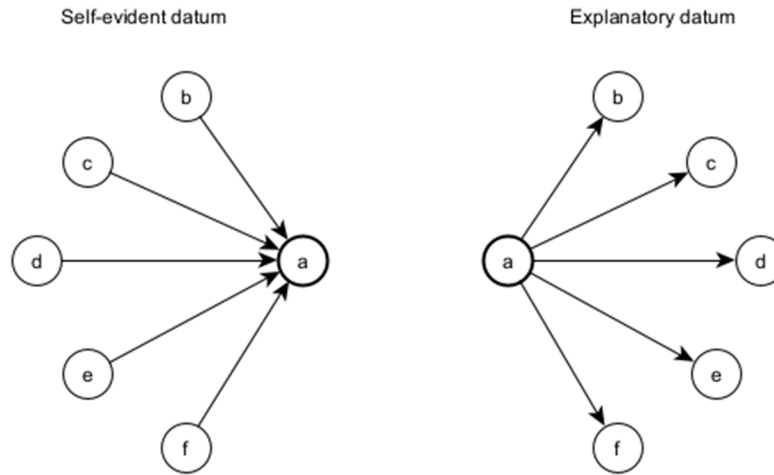


Figure 6: Cognitive model presented in Paglieri (2004) to represent relevant data.

ambiguous situation, whereas argumentation is used to *justify* the selection of a specific argument for the recommendation. Further details on this second aspect are provided in the next section.

4.2.3 The pursuit of goals and *Plausible arguments*

A further important aspect of dialogue is the pursuit of a goal and the illocutionary force with which a speaker produces a given speech act. This distinction is one we aim to highlight, particularly in relation to generative models that lack this capability. This illocutionary completion also parallels what occurs in argumentation. As already mentioned, argumentation is a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by putting forward a constellation of propositions (i.e. illocutionary complex speech act) justifying or refuting the proposition expressed in the standpoint (Van Eemeren and Grootendorst 2003). The selection of the most appropriate items and features is pivotal for the achievement of the conversational goal. An example of argumentation dialogue where the selection of items and features is particularly clear is the recommendation dialogue. The Recommendation task, introduced in Section 3, tends to present a pattern structured in two phases, Exploration and Exploitation (E&E), intended as two types of dialogues embedded into each other (Gao et al. 2021). While Exploration involves the collection of information about specific features (e.g. genre, actors), Exploitation concerns the selection of a specific item (e.g. movie) associated with its possible justifications and/or motivations. The recovering of relevant information from the user is carried out during the exploration phase, so it is fundamental to adopt a strategy for the retrieving of the right feedback on the data candidate as possible system beliefs. According to Paglieri and Castelfranchi (2004b), and Paglieri (2004, 2005), data are selected as beliefs on the basis of their properties, i.e. the possible cognitive reasons to believe such data. These properties are *credibility*, *importance*, *relevance*, and *(un-)likeability*, described as follows:

- **Credibility:** a measure of the number and values of all supporting data, contrasted with all conflicting data, down to external and internal sources;
- **Importance:** a measure of the epistemic connectivity of the datum, i.e. the number and values of the data that the agent will have to revise, should he revise that single one;
- **Relevance:** a measure of the pragmatic utility of the datum, i.e. the number and values of the (pursued) goals that depends on that datum;
- **(Un-)Likeability:** a measure of the motivational appeal of the datum, i.e. the number and values of the (pursued) goals that are directly fulfilled by that datum.

The authors pointed out that credibility, importance, and likeability determine the outcomes of belief selection, i.e. whether a candidate data is to be believed or not, and with which strength, while relevance is

crucial in pre-selecting the sub-set of active data. The authors refer to that process as *focusing* that rules the way data are considered to be useful and/or appropriate in the agent's mind.

Based on the relevant data collected through the Exploration phase, the Exploitation one use this information to select the *plausible* item to recommend. *Plausibility* can be defined as the degree of connectivity or effectiveness of an argument within the dialogue. It assesses how relevant and believable a new argument is, based on the quality of the new data and also on its connection with data already available to that user. Specifically, a plausible argument is one that appears to be well-supported, logical, and consistent with the available information and Common Ground (Paglieri and Castelfranchi 2004a). In fact, a crucial factor in determining whether a new piece of information will be accepted or rejected as belief is the degree of connectivity of the new datum in the user's background knowledge, defining, thus, its level of plausibility. According to Paglieri and Castelfranchi (2004a)'s data networks, there are two case of argumentation through plausibility: (i) self-evident data having a large number of data connections that support them; (ii) explanatory data which, in turn, are connected to many other data to support them, as shown in Figure 6. Again, graphs play a pivotal role in structuring and representing such cognitive model, enabling the extraction of plausible arguments by visualising the relationships between data points and supporting the decision-making process in argumentation-based systems. In Section 5.2, this graph-based cognitive model will be associated with mathematical descriptors for a representation and management of argumentation-based dialogue systems.

5 Modelling operative representations for dialogue orchestration research (MORDOR)

While the advent of machine learning has provided a way to get past the limitations of rule-based reasoners, it appears that the last decade has brought us to the opposite extreme, with the main research trends using data-driven statistical models for all AI-related tasks. This led to an explainability problem and has also caused an increased interest towards explanations modelling in dialogue systems (for a full survey, see Vassiliades et al. (2021)). MORDOR attempts to take the best of each world by avoiding a one-wins-all approach to model usage and proposing a general architecture for dialogue management where perception and behaviour synthesis are managed with neural approaches and deeper layers, dedicated to reasoning and decision making, are instead implemented using graphical models like native graph databases and probabilistic graphical models (PGMs). Graph databases, in MORDOR, are used as a modern version of traditional inference engines, using performance-oriented path search capabilities in complex networks to implement deduction and data selection skills. PGM, on the other hand, are used to implement decision making processes, being able to take into account uncertainties coming from machine learning estimates and causal relationships between variables. To

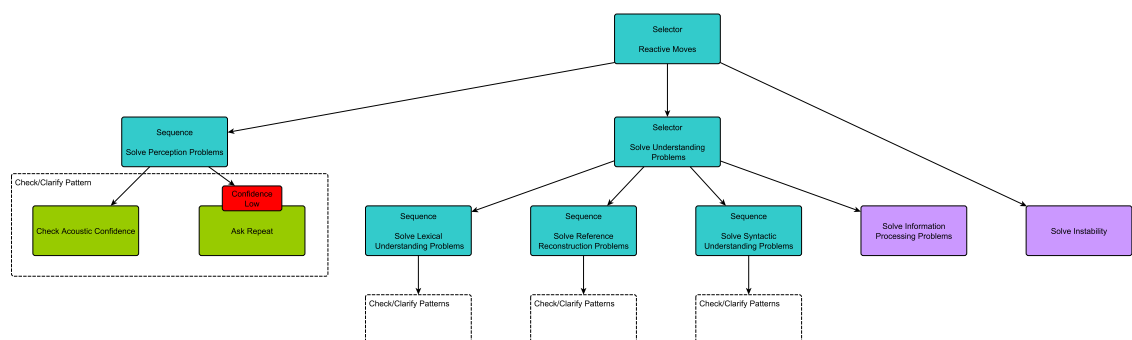


Figure 7: The BT for reactive moves. Checks, first, interpretability problems and generates a CR following the priorities described in Maro (2021). The check/clarification pattern is simplified, in the Figure, for readability purposes. A dedicated subtree handles information processing problems. If there are no interpretability problems, the subtree handling Instability is activated.

organise tasks and priorities, MORDOR makes use of PGM, a traditional AI approach to behaviour selection, and a decision-making engine, based on Bayesian Networks and network analysis algorithms. Starting from the graph-based formalisation of linguistics observations coming from LORIEN, we define a technological architecture to implement embodied conversational agents, as specified by **C3**.

In the rest of the following sections, we summarise the architectural organisation of a MORDOR dialogue manager designed to host theories developed through LORIEN observations.

5.1 Behaviour trees

At the highest level, behaviours that can be expressed through dialogue moves are organised in a tree-like structure defining priority relationships. This model and the rules to interpret it are called BTs and they have been extensively applied in robotics, AI and particularly in the game industry (Iovino et al. 2022). In MORDOR, BTs organise dialog strategies following the priority order found through LORIEN.

Following the CR model described in Section 4.2.1, a hierarchy of potential communication problems exists that must be checked and resolved, if detected, before considering other dialogue moves. BTs can efficiently represent the CR hierarchy as a mathematical model implementing the mechanism described by the linguistic theory. In general, a sequence of checks whose order is motivated by the hierarchy is activated and, if one of these succeeds, the corresponding clarification strategy is performed, ignoring the rest of the available behaviours. For instance, a particularly interesting communication problem concerns conflicts between previous beliefs from the system and incoming evidence from the user, i.e. Common Ground inconsistencies. Linguistics research highlights that specific question forms must be used to efficiently communicate the problem, and the transaction system, made available by graph databases, allows us to implement a reasoning mechanism that allows us to verify that beliefs implied by the user do not conflict with existing beliefs. Specifically, if no higher priority problem is detected, the system can temporarily accept the beliefs implied by the last user utterance by opening a transaction and update the belief graph without committing the changes. Then, conflicting patterns can be searched for in the temporary graph: this includes checking that a belief and its negation do not exist at the same time, in the graph, concerning the same subject and predicate. If a conflicting pattern is found, the corresponding CR is generated and the transaction is rolled back. This effectively implements a hypothesising mechanism that allows the system to reason about what would happen if it was to accept the belief implied by the user.

If no communication problems are detected, the system can move to intentional moves. In this case, the priority is given to solving open issues like questions asked by the user. If no open issues are detected, the system can produce a dialogue move to direct the dialogue towards the desired goal, like for example having the user accept a proposed stance, like a recommendation. Deciding which move to perform, when communication problems are not present, depends on the *reason* why the system is producing linguistic moves. The concepts of *illocutionary force* and *intentionality* that drive Explanation models, described in Section 4.2.2, are foundational to implement a *reason* to act, in a dialog system, that goes beyond the reactive behaviour of GPT models. Recovering the concept of goal-oriented behaviour from traditional AI, we propose that goals can be expressed as graph patterns that the system aims at creating, in collaboration with the human user. To create these patterns, the system is provided with linguistic capabilities it must select and configure in such a way as to maximise their *utility*. This approach lies between rule-based and statistical approaches, as it needs to balance the uncertainty of the information coming from probabilistic estimates with factual knowledge hosted in the graph database. Following an approach similar to the one presented in Opendial (Lison and Kennington 2016), we adopt the use of PGM to implement goal-oriented decision mechanisms for dialogue management.

Asking questions, in this framework, corresponds to collecting evidence about the variables in PGMs, which represents the implementation of the theoretical foundation provided by our Argumentation model, as described in Section 4.2.3. By computing the nodes' entropy, which represents the uncertainty of the information contained in the node, and considering the general goals of the system, dialogue management strategies can evaluate the next move. Technological limitations do not allow us to create PGMs that are as large as the

entire database as they would be unmanageable. It is, however, possible to extract *relevant* subgraphs from the database to dynamically assemble PGMs to take decisions informed only by relevant data and their causal structure. Classic part of relationships represented in the database, for example, can be used to form causal relationship between variables in a PGM. Differently from traditional approaches, modern ways to represent data in the form of graph provide a strong connection between decision systems and large data hosting, approximating, respectively, the role of *working memory* and *long-term memory*. Selecting a subset of data on which to reason may be done by both rule-based queries or by using embedding-based similarities, to take into account nuances in the organisation of the dataset of interest. This extraction operation may be related to a kind of *Artificial Instinct* and makes use of authority and hub scores to prioritise the extraction of nodes, when selecting a relevant subset. In general, Graph Data Science techniques, as anticipated in Section 4.2.3, can support this step by making latent knowledge emerge from the graph structure using statistical measures.

When producing dialog moves, it may be necessary to present inferential statements, obtained by extracting paths over nodes in the graph, to support the positions expressed by the system. Depending on the illocutionary force of the statements, the motivations behind their expression, they may represent argumentations, if they are meant to persuade the interlocutor into accepting an unsettled claim, or they may represent explanations, if they are meant to let the interlocutor understand a point. In terms of a BT, the illocutionary force of a statement is represented by the position, in the BT, of the task that generates the statement. In the current version of MORDOR, inferential statements supporting the main one are explanations if they are produced to support an answer to a question. If they are generated during exploitation, the actual recommendation phase, or during the exploration phase, where more information about the user is collected, they are considered argumentations.

Moving through the behaviour hierarchy defined in such a way requires a formal definition of dialogue states activating different system behaviours and producing adequate utterances. LORIEN analysis describes these dialogue configurations through Common Ground representations, possibly integrated by Common Sense reasoning, as described in Section 4.1. The same configurations describe patterns that are used to activate adequate behaviour generation strategies. Graph configurations, together with BTs organisation, produces a clear sequence of formal checks performed on graph structures activating specific dialogue management behaviours, as follows:

- **Interpretability:** a graph is uninterpretable if any of the graph patterns describing each of the foreseen communication problems is activated. In these cases, a CR is produced (Figure 7);

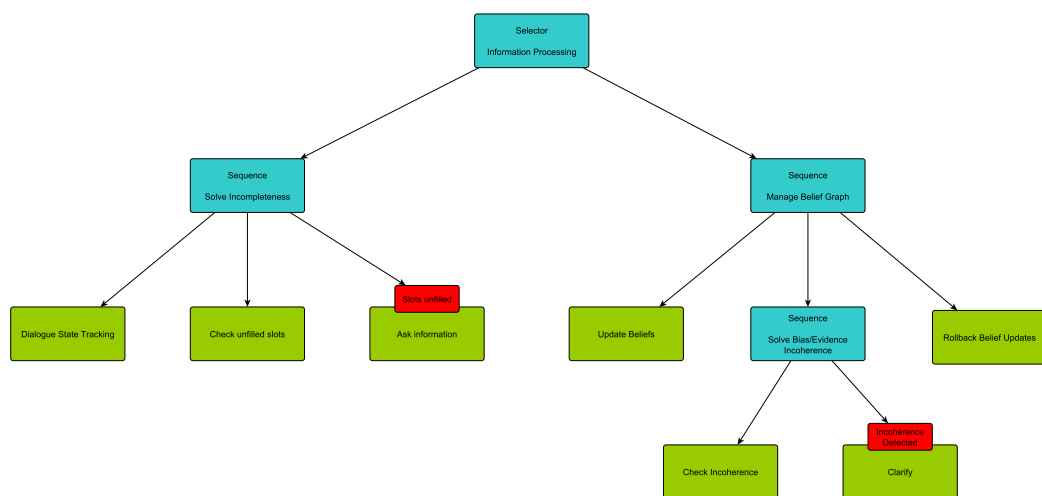


Figure 8: The BT for information processing problems. First, incompleteness is checked and, if necessary, solved by, first, attempting to apply dialogue state tracking or by generating an information request. If the graph is complete, the belief graph is updated coherently with the user utterance. Then, incoherence in the belief graph is checked. If an incoherence is found, a CR is generated and the updates are rolled back.

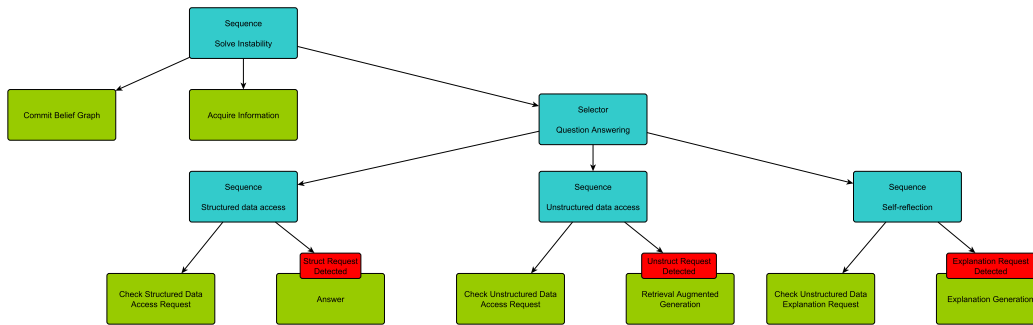


Figure 9: The BT to manage Instability. If the graph is coherent, changes to the belief graph are committed and information contained in the user utterance (e.g. the user's name) is saved in the graph. The system, then, checks if the user asked a question and, in this case, it activates either the strategy dedicated to catalographic data extraction or more advanced Question Answering strategies, for example retrieval augmented generation.

- **Completeness:** an interpretable graph is incomplete if information needed to respond to the user intent is missing. In these cases, a request for information is produced (left side of Figure 8);
- **Coherence:** a complete graph is incoherent if logical conflicts are found in the *belief graph*. In these cases, the adequate disambiguation question is produced (right side of Figure 8);
- **Stability:** a coherent graph is unstable if there are open issues, like unanswered questions. In these cases, a question answering strategy is activated (Figure 9);
- **Desirability:** a stable graph is undesirable if it does not exhibit the goal pattern. In these cases, the most *useful* dialog move to create the goal pattern is produced, like an exploration or exploitation move (Figure 10).

These configurations represent a hierarchy producing system reactions depending on their priority. A linguistically motivated model for this hierarchy provides a first level of explainability for the system moves, as they are produced in response to a specific graph configuration. As previously mentioned in Section 3, key concepts like intentionality and illocutionary force are indeed essential for giving technological systems a reason to communicate, aligning with Austin's idea that speaking is acting. The machine's *reasons* to speak are dictated by increasingly self-oriented goals: solving communication problems has less illocutionary strength than answering questions, which in turn has less illocutionary strength compared to taking stances and asking questions. On this continuum, we position different kinds of speech acts, which are the subject of analysis and modelling presented in LORIEN (Section 4), as shown in Figure 11. More in detail, in Section 4.2.2, different types of explanatory behaviours were mapped on different degrees of illocutionary force, from speech acts

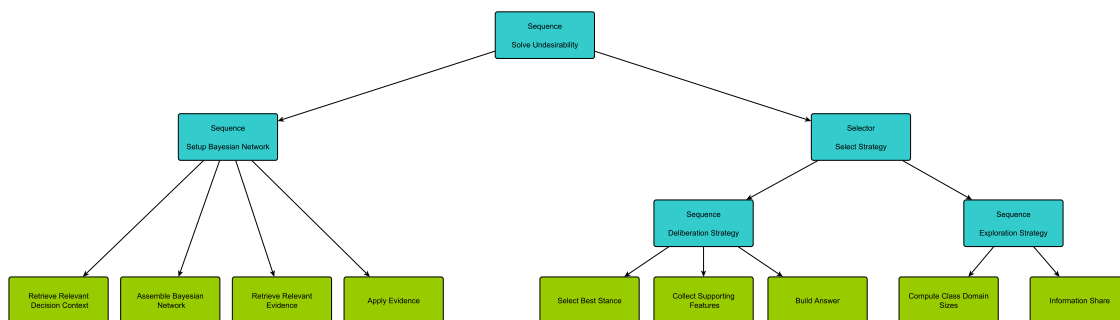


Figure 10: The subtree dedicated to the management of undesirable dialogue state graph configurations. Depending on the collected information, in terms of beliefs the system has about the user, a subgraph relevant for decision making is extracted from the database and the ontological relationships are used to assemble a PGM. The most useful action is, then, selected as either a deliberative or an explorative statement.

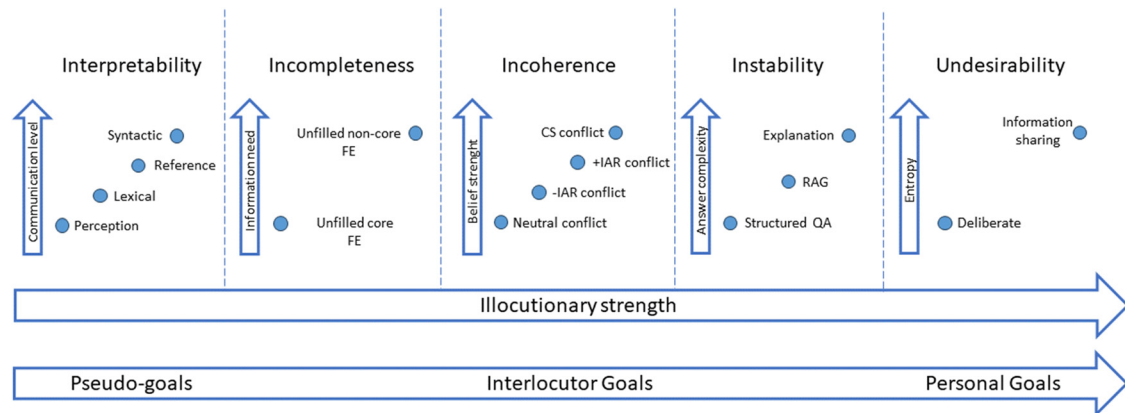


Figure 11: The continuum defined by illocutionary strength, strictly connected to the type of goal, whose corresponding speech acts are aimed to, and mapped onto graph configurations. Different kinds of speech acts are positioned on the continuum depending on secondary axes defined for each graph configuration. For example, in the interpretability graph configuration, acts corresponding to the communication level are ordered by gradual degrees of illocutionary strength in the effort to reach a pseudo-goal.

aimed at solving inconsistencies (i.e. requests or explanations), to question answering and goal-based explanations (i.e. justification), from deliberation (possibly with argumentation) to exploration.

Summarising, the process by which a dialogue system generates linguistic content to interact with a human user follows the perception-action cycle, employing different models at various stages to simulate the cognitive mechanisms involved in dialogue management. As a general principle, purely statistical models are found closer to the physical world, for tasks like speech recognition, intent classification, and utterance generation. Higher functions are simulated using technologies that are closer to symbolic reasoning like graph databases. In between, Bayesian models reconcile the probabilistic nature of perceptual data with the factual nature of general knowledge. This is summarised in Figure 12. Such systems are implemented using the FANTASIA tool (Origlia *et al.* 2019), designed to implement Embodied Conversational Agents, and the tool's architecture is shown in Figure 13.

5.2 Decision-making

It is not possible to answer questions about interventions with passively collected data, regardless of the size of the dataset or the depth of the neural network. The construction of a sufficiently strong and accurate causal model allows us to use observational data from rung one to answer queries from rung two, which concerns

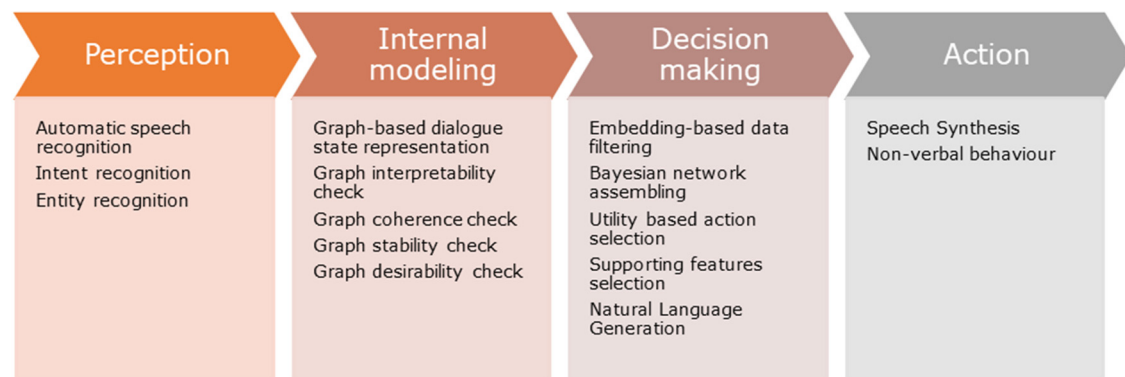


Figure 12: The MORDOR process for dialogue management.

interventions: without a causal model, it would not be possible to go from rung one to rung two of the Ladder of Causation. Causal diagrams are employed to express what is known and are simply dot-and-arrow diagrams that summarise the existing scientific knowledge. The dots represent quantities of interest, called ‘variables,’ and the arrows represent known or suspected causal relationships between those variables, i.e. which variable ‘listens’ to which others. Bayesian Networks are the key to linking causal diagrams to data (Pearl and Mackenzie 2018). BNs are directed acyclic graphs in which the nodes represent the variables of interest, and the links represent the informational or causal dependencies between these variables. The strength of a dependency is represented by conditional probabilities attached to each cluster of parent–child nodes in the network. BNs are an attempt to develop a computational model of human inferential reasoning, i.e. the mechanism by which people integrate data from sources and generate a coherent interpretation of that data. According to Pearl (1985), humans organise their knowledge using the concept of *conditional independence*. Let us consider events A, B, and C. We will say that event A is conditionally independent of event B given C if $P(A|B, C) = P(A|C)$. Knowledge of B does not change the probability of A relative to the probability of C, from which A is not conditionally independent. BNs represent conditional independencies, in particular each node is conditionally dependent on each of its ascendants (parent nodes) and conditionally independent of the others. BNs are the root of most PGMs, which are important for our approach as they represent a method for reasoning under uncertainty using a model that is topologically compatible with the graph-based representation of knowledge. The model is informed using both prior information (data coming from previous experience) and situation-specific information, presented in the form of evidence. Pieces of evidence can be introduced, in the model, both as *soft evidence* – probability distributions over possible values (representing what may be true for the interlocutor) – or as *hard evidence* – certain values for specific variables (representing what is true about the interlocutor). The shape of the probability distributions also provides information about the system’s uncertainty, using entropy. Graph analysis methods, therefore, can be used to support the extraction of *relevant* subsets of information from the knowledge graph to dynamically assemble decision models to handle contingent situations. This approach was pioneered by Lison and Kennington (2016), with the Opendial framework. We build upon this idea by framing PGMs in a larger, linguistically motivated, architecture for dialogue management.

Specifically, we use the cognitive model presented in Section 4.2.3 to select and evaluate the data at our disposal leveraging on a graph database hosting a knowledge base of common facts collected from LOD

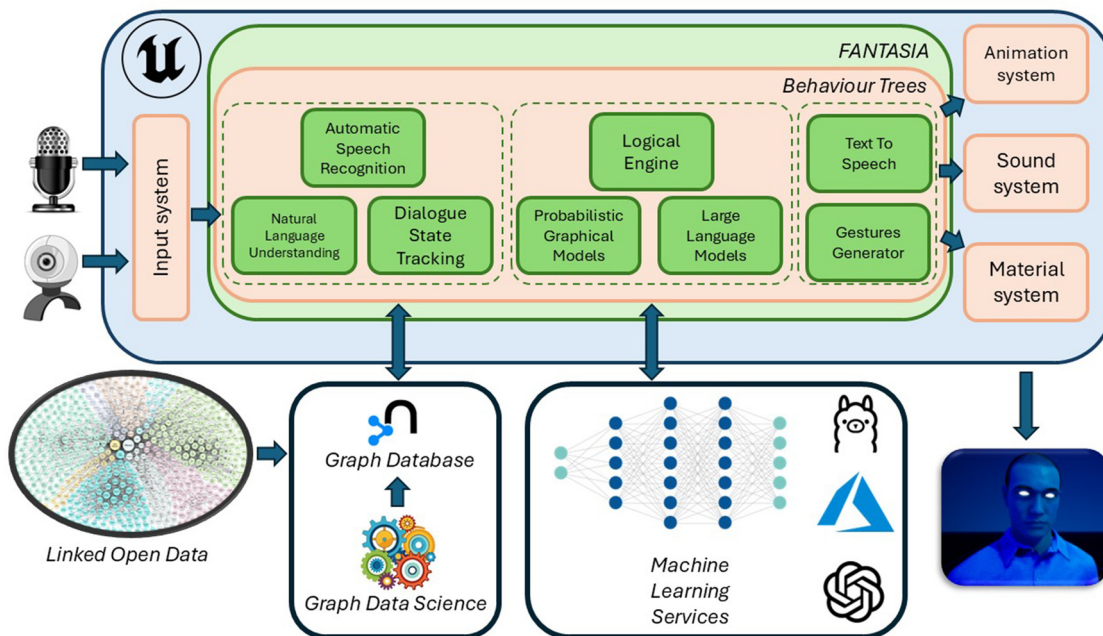
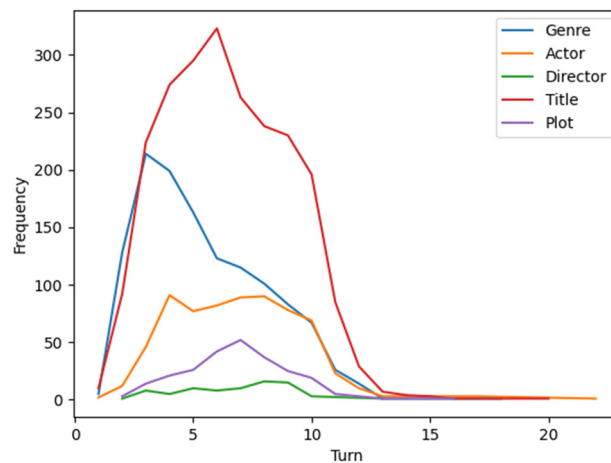


Figure 13: The FANTASIA architecture.

Table 3: Cognitive properties described by Paglieri (2004) mapped on computational scores

Theoretical model	Computational scores
Credibility A measure of the number and values of all supporting data, contrasted with all conflicting data, down to external and internal sources	Authority score The authority score identifies the node with a fundamental role in the graph since a solid number of hub nodes support its validity
Importance A measure of the epistemic connectivity of the datum, i.e. the number and values of the data that the agent will have to revise, should they revise that single one	Hub score The hub score identifies nodes most connected to authoritative nodes.
Relevance A measure of the pragmatic utility of the datum, i.e. the number and values of the (pursued) goals that depends on that datum	Entropy The entropy identifies which relevant and less certain data needs a feedback to continue with the dialogue interaction
(Un-)likeability A measure of the motivational appeal of the datum, i.e. the number and values of the (pursued) goals that are directly fulfilled by that datum.	Hard evidence The system beliefs involvement in the selection of the feature explicates the user appeal towards that kind of data.

sources using the procedure described by Origlia *et al.* (2022) and in Section 4. Graph analysis allows us to find regularities that can be exploited to form new background theories and to support technological approaches built upon them. Among the various measures shown in Table 3, we highlight a network analysis procedure with the HITS algorithm (Kleinberg 1999), which attributes *authority* and *hub* scores to the nodes (Figure 15). These measures are mapped onto the aforementioned cognitive properties of importance and credibility, respectively. They indicate the number of nodes with high hub scores pointing towards the considered node (i.e. data mostly supported by other data and, therefore, considered credible) and, symmetrically, the number of authority nodes that can be reached from the considered node (i.e. important data supporting specific information). As the two measures depend on each other, the algorithm alternatively updates them at each iteration until they converge. This analysis provides indications about how the disambiguation of nodes in the network should be prioritised. In fact, during the Exploration phase, the selection of the right feature to explore is important for the achievement of the pursued goal. For instance, this can be translated, in the movie recommendation domain, into the necessity to know which appropriate feature to select to start the conversation. According to the graph configuration, for example, it has been pointed out that the most authoritative nodes are genre nodes. Therefore, the first feature to be explored, that also help lowering the entropy, is genre.

**Figure 14:** Analysis of placeholders distributions across turns in the INSPIRED Corpus (Hayati *et al.* 2020).

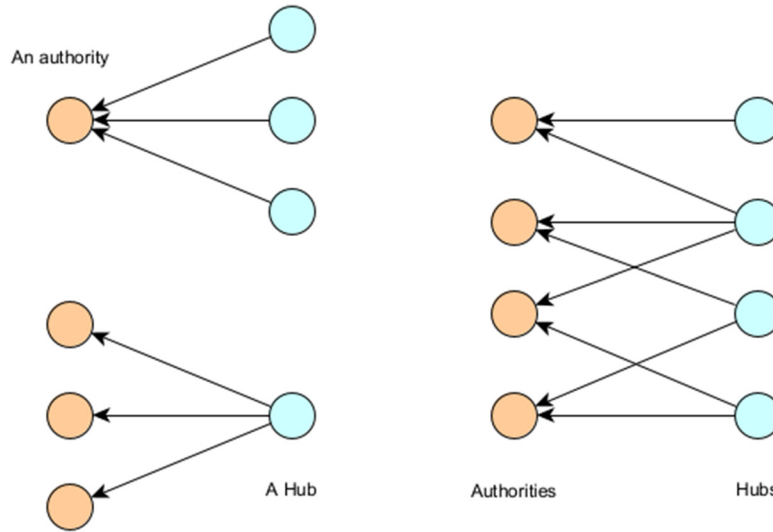


Figure 15: According to Kleinberg (1999), nodes with a high hub score refer to nodes with a high authority score, while nodes with a high authority score are referred to by nodes with a high hub score. These scores provide insight over the relevant parts of the graph structure.

Importantly, this tendency also resulted from our analysis of human–human dialogues of the same type collected in the INSPIRED Corpus (Hayati et al. 2020), confirming the appropriateness of our measures. The result of this analysis is represented in Figure 14. The selection of the further features always depends on the updated graph configuration resulting from the previously acquired information. Finally, as far as relevance and (un-)likeability are concerned, soft evidence (entropy), and hard evidence, respectively, are computed to select items which are pragmatically useful and likeable. In Di Bratto et al. (2024a,b), an argumentation-based dialogue system was developed, grounded in the aforementioned cognitive principles and their graph-dependent mathematical descriptors, to enable the use of plausible arguments. Dialogues simulated by the system, structured in this way, were evaluated by assessors unaware of the nature of these dialogues. The results showed high scores for the naturalness and plausibility of the arguments. These findings contribute to advancing research on argumentation in AI, with future work focusing on deploying the model in various domains and expanding its communication abilities using linguistic theories. This is influenced by the needs of a computational approach to theory evaluation, as detailed in MORDOR (Section 5).

6 Journey through the ages: rediscovering past approaches

From a linguistic point of view, the study of CRs, as introduced in Section 4.2.1, allowed for preliminary considerations on argumentation strategies in dialogue systems. Indeed, these systems can detect conflicts and employ argumentation strategies, specifically Common Ground Clarification Requests in the form of negative polar questions, to consistently signal them based on previous observations (Di Maro et al. 2021), facilitating the pursuit of agreement. Furthermore, the use of such requests has been proven to enhance the usability and naturalness of dialogue systems (Di Maro et al. 2021).

With the aim of reproducing linguistic observations in dialogue systems, graph databases, the cornerstone of LORIEN, were applied as a linking bridge between linguistics and technology. On the one hand, graphs can indeed be seen as an integrated solution for dialogue state tracking, knowledge representation, and conflict detection, serving as a fundamental building block for dialogue systems with argumentation capabilities, as shown in Di Maro et al. (2021) and Russo et al. (2022). More specifically, in the study by Di Maro et al. (2021), a Conflict Search Graph was designed to represent dialogue history and connect it with domain knowledge, enabling Common Ground stability checks and dialogue state tracking to be represented in the form of graph

queries. On the other hand, graphs have also been proven to be an efficient tool for corpus analysis. In the study by Di Bratto *et al.* (2021), different sets of shared knowledge were represented as graphs to gain a deeper understanding of dialogue phenomena and support a more informed design of dialogue systems. One example is the RING-like pattern, which is used to compute the consistency maintained in the dialogue. RING-like patterns explicitly consider relationships established across both the PE and the CCG in different turns of a dialogue. This pattern is extracted through a query to the database containing the dialogue history and can be used for both disambiguation purposes and to extract the context that led interlocutors to discuss a certain topic, as discussed by Di Bratto *et al.* (2021) and as already mentioned in Section 4.1.

From a technological standpoint, the presented methodology began to take shape with the investigation of graph databases applied to linguistic research, specifically for speech recognition grammars (Di Maro *et al.* 2017). This exploration, leveraging the emerging technology of Neo4j, was later combined with the Bayesian approach to dialogue management provided by Opendial (Lison and Kennington 2016). The concept of ‘utility’ in Opendial was used to dynamically select minimal pairs in a perception test for young children, aiming to detect language acquisition problems (Origlia *et al.* 2018a,b). These works marked the initial proposal of using a graph database to host linguistic knowledge, the Unreal Engine as part of the interactive interface, and Bayesian Networks for the dialogue management (Origlia *et al.* 2017). Simultaneously, building on previous experiences, applications of graph databases to cultural heritage knowledge representation were being explored. These applications were intended to be integrated with techniques for semantically annotating 3D models obtained through photogrammetry and laser scanning (Cera *et al.* 2018). The goal was to ensure coherent behaviour in AIs dealing with reconstructed 3D spaces for presentation purposes (Campi *et al.* 2021). Specifically, pointing gestures were generated coherently with the concepts expressed by a virtual avatar during the description of selected environments in the San Martino Charterhouse (Naples, Italy). Furthermore, the investigation into the role of disfluencies in expert guide presentations, relevant for the social behaviour of Embodied Conversational Agents, started (Origlia *et al.* 2019, Cataldo *et al.* 2019). Through these experiences, graph-based knowledge representation formats and the corresponding methodology for analysis and enrichment were developed, resulting in the graph databases presented by Di Bratto *et al.* (2021), Origlia *et al.* (2022) and (Origlia *et al.* 2021). In addition, the first version of FANTASIA, still built around Opendial and tested in these domains, was introduced (Origlia *et al.* 2019). The system architecture was then re-engineered to better integrate basic technologies. The most significant change was abandoning Opendial in favour of the aGrUM library (Ducamp *et al.* 2020), which, being implemented in C++, is compatible with the plugin system of the Unreal Engine. The aGrUM library provides all the necessary tools to assemble and probe different kinds of probabilistic graphical models, allowing for a more flexible development of dialogue systems compared to Opendial. Most of the connectors developed to provide data to Opendial were also rewritten inside the Unreal Engine to offer a coherent development experience. At this point, the re-engineered version of FANTASIA³ (Origlia *et al.* 2022) represents the implementation of the methodology described in this article, linking graph-based resource analysis procedures based on linguistics with the technological tools needed to implement conversational agents using the Unreal Engine. FANTASIA was used to develop the applications presented by Origlia *et al.* (2022), and it is actively being developed to keep up with the Unreal Engine while adding more functionalities, especially from the aGrUM library. At present, FANTASIA also includes a logical inference engine, based on SWI Prolog, which will be used in future iterations of ARDA.

Other than simple technological advancement, the structure of FANTASIA and the general methodology presented here allow us to investigate an alternative path to the now popular one, involving heavy use of LLMs. This approach is built upon the theoretical model provided by Judea Pearl (Pearl and Mackenzie 2018) regarding the representation of causal relationships in AI for advancements in the field. Probabilistic graphical models, as pioneered by (Lison and Kennington 2016), still have much power to be exploited in the field of dialogue systems, especially considering all the sources of uncertainty highlighted in (Prakken 2018a) concerning argumentation-based dialogue. This topic, in particular, provides a suitable field of investigation, as the amount of research is relatively low when compared to argumentation-based inference, but the

³ Freely available at <https://github.com/antori82/FANTASIA>.

complexity of the challenges is still very high. Also, considering that Pearl's framework is model-based rather than data-driven, an important challenge consists of correctly assembling the graphical models needed to implement reasoning at runtime, following domain-independent principles. The methodological framework we presented and the tools provided by FANTASIA allow us to explore the relationship between linguistic models, numerical measures computed through graph data science, and the graphical structure of PGMs. The general approach to follow when working with ARDA consists of separating domain-specific knowledge, represented in the form of graphs, from generic behavioural models, represented using BTs and PGMs. The role of machine learning remains fundamental but is limited to interactions with the physical world, either to perceive it (listening and understanding) or to interact with it (generating sentences and speaking).

7 Discussion and conclusions

We have presented the methodological framework developed in the last few years of our cross-disciplinary research between linguistics and AI. Our aim was to explore the following main contributions:

- C1 A linguistically motivated methodological approach that spans from linguistic research to the development of argumentative dialogue systems (ARDA);
- C2 A graph-based interpretation of the methodology for corpus-linguistics data analysis (LORIEN);
- C3 A technological architecture for implementing embodied conversational agents using the graph representation of linguistic concepts (MORDOR).

We started with a graph-based view of corpus linguistics and investigated the correspondence between linguistic models for dialogue management and Common Ground representations, utilising mathematical tools linked with graph structures. Subsequently, we explored multiple domains of application for graphs in dialogue management, ranging from simple knowledge representation to conflict detection in Common Ground, and currently, to argument selection for recommendation. These linguistically motivated representation models, combined with Graph Data Science approaches, guide the behaviour of conversational agents using a combination of graph databases, probabilistic graphical models, and real-time interactive 3D technology. Through this report on experiences and lessons learned, and by providing the community with the freely available FANTASIA plugin, we aim to equip researchers interested in Embodied Conversational Agents with the tools needed to explore alternative approaches to dialogue management, involving but not blindly relying on LLMs. We believe this approach can serve as a platform for linguistics and computer science researchers to collaborate in building linguistically motivated conversational AI capable of dealing with very complex scenarios while remaining accessible from a computational point of view.

More in-depth, the recent success of generative AI approaches in managing dialogues has sparked a worldwide discussion concerning the actual process that occurs when a human interacts with such a model through linguistic means. In our model, we make use of generative AI, but we only constrain it to very specific tasks. In particular, we do not use it for dialogue management. First, we highlight three main aspects of generative models used for dialogue management:

- They only capture surface aspects of communication because they are trained to predict the most probable *continuation* of some textual input;
- Their only *intention* or goal is to complete the text: they don't act to pursue any other specific goal, so they have no other *reason* to communicate;
- They operate on a reactive basis: being connectionist models, no matter how complex they are, they are not conceptually different from Braitenberg vehicles (Braitenberg 1986), with all the implications.

On these grounds, we argue that generative AI is a powerful tool for Natural Language Generation only, and it should be integrated as such into a Conversational Agent. A complete Conversational AI should have an explicit module dedicated to decision-making modelling inspired by deeper cognitive functions, producing linguistic output as a means to pursue a *desirable* goal. In this sense, we emphasise that the emergence of

dialogues is impossible if one of the involved parties lacks an actual communicative intention. This foundational point has been crucial in a significant amount of cross-disciplinary research involving linguistics and AI since the early 1990s (Cohen *et al.* 1990). Intentions, in particular, have been presented as the missing element in classic philosophical models dating back to Aristotle, based on beliefs and desires, and still present in modern models, as described in Forgas (1989). Models concerning the philosophy of action explicitly include intentions, along with beliefs and desires (Searle 1983, Bratman 1987), although this component appears to emerge at a later stage of human development, as evidenced by studies on the development of theory of mind in young children (Astington and Gopnik 1991). It should be noted that the perception of intentionality by one of the parties involved in dialogue is not sufficient for an actual communication exchange to occur. Attribution of intentions is, in fact, a known phenomenon in psychology (Malle and Knobe 1997). The opposition between *acting* based on beliefs, desires, intentions, and possibly also skills and awareness (Malle and Knobe 1997), and *reacting* to linguistic stimuli with the most probably appropriate content forms the basis of our stance. In an attempt to reconcile classic AI approaches with modern trained statistical models, we consider dialogues as sequences of cooperative communicative *actions* rather than sequences of coherent language generation *reactions*. On the other hand, the flexibility and local contextualisation capabilities of large language models provide a powerful way to actualise dialogue intentions in linguistic content.

It is important, therefore, not to consider linguistic material as the final output of the communication process but only as a means to an end, so that the underlying *raison d'être* becomes the real goal to pursue. This is evidenced by the amount of research directed towards modelling the relationship between language and real-world measurements and the improvement in the quality of semantic representations when combined with video or audio data; language has no reason to exist *per se*. Causal models (Pearl and Mackenzie 2018), in this sense, find their way into a model of dialogue management as one of the manifestations of higher orders of intelligence capable of modelling *interventions*. In general, we consider language as a tool used to interact and influence the world to intentionally create desirable outcomes, considering cause-and-effect links between linguistic behaviour and evolving representations of reality. LORIEN is concerned with investigating communication strategies in close connection with the pragmatic context, so that the observed communication strategies can be related to the situational needs of the participants in the dialogue. These models are then converted into technological implementations to verify their soundness and further improve them by studying emergent problems and contradictions. We define the set of methodological procedures to transfer LORIEN findings for use in a computational model as MORDOR.

In MORDOR, neural approaches are mainly used to manage tasks that are closer, in the perception–action cycle, to the physical level. Machine learning models, in general, are extremely proficient in generalisation and pattern recognition. However, due to their nature, they are not a good choice for decision-making and symbolic reasoning, which implies the capability to exhibit goal-oriented behaviour. Classic AI was more adept at long-term planning and had the advantage of providing clearly interpretable reasoning to motivate its actions, along with less imposing requirements for computational power and data availability. On the other hand, due to its need for exact definitions for everything, it struggled and eventually failed to provide a valid approach for applications that needed to deal with physical reality, as the world is ‘the best representation of itself,’ as stated by the frame problem.

Funding information: This work was supported by the Supporting Patients with Embodied Conversational Interfaces and Argumentative Language (SPECIAL) project, funded by the University of Naples on the ‘Fondi per la Ricerca di Ateneo’ (FRA) program (CUP: E65F22000050001). Open Access publication was supported by the University of Naples Federico II through the project DIGHT Digital Interventions in Cultural Heritage Technologies.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission. Maria Di Maro and Martina Di Bratto designed the linguistic methodological part of ARDA (LORIEN). Maria Di Maro and Antonio Origlia designed the technological methodological part of ARDA (MORDOR). Maria Di Maro supervised the integration of the two parts. Sabrina Mennella contributed the view concerning Common Sense. Francesco Cutugno supervised the work and revised the final manuscript.

Conflict of interest: The authors state no conflict of interest.

References

- Allwood, Jens, Joakim Nivre, and Elisabeth Ahlsén. 1992. "On the semantics and pragmatics of linguistic feedback." *Journal of semantics* 9 (1): 1–26.
- Artikis, Alexander, Marek Sergot, and Jeremy Pitt. 2007. "An executable specification of a formal argumentation protocol." *Artificial Intelligence* 171 (10–15): 776–804.
- Asterhan, Christa SC, and Baruch B. Schwarz. 2009. "Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog." *Cognitive Science* 33 (3): 374–400.
- Astington, Janet Wilde, and Alison Gopnik. 1991. "Theoretical explanations of children's understanding of the mind." *British Journal of Developmental Psychology* 9 (1): 7–31.
- Austin, John Langshaw. 1962. *How to Do Things with Words*. Cambridge: Clarendon Press.
- Bauer, Martin W. 2024. "AI with common sense: What concept of common sense?." in: *AI and Common Sense*, 13–29. Abingdon: Routledge.
- Black, Elizabeth, and Katie Atkinson. 2010. "Agreeing what to do." in: *International Workshop on Argumentation in Multi-Agent Systems*, 12–30. Springer.
- Bodenstaff, Lianne, Henry Prakken, and Gerard Vreeswijk. 2006. "On formalising dialogue systems for argumentation in the event calculus." in: *Proceedings of the eleventh international workshop on nonmonotonic reasoning*, 374–82. Windermere: Gesellschaft fuer Informatik.
- Braitenberg, Valentino. 1986. *Vehicles: Experiments in synthetic psychology*. Berlin: MIT Press.
- Bratman, Michael. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- Brewka, Gerhard. 2001. "Dynamic argument systems: A formal model of argumentation processes based on situation calculus." *Journal of logic and computation* 11 (2): 257–82.
- Brooks, Rodney. 1991. "Intelligence without representation." *Artificial intelligence* 47 (1–3): 139–59.
- Brugman, Hennie, and Albert Russe. 2004. "Annotating multi-media/multi-modal resources with ELAN." in *LREC*, 2065–8. Lisbon: European Language Resources Association (ELRA) Press.
- Budzynska, Katarzyna, and Chris Reed. 2011. "Speech acts of argumentation: Inference anchors and peripheral cues in dialogue." in: *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*. San Francisco, California, USA.
- Buring, Daniel, and Christine Gunlogson. 2000. "Aren't positive and negative polar questions the same?." *UCSC/UCLA*.
- Cambria, Erik Amir Hussain, Catherine Havasi, and Chris Eckl. 2009. "Common sense computing: From the society of mind to digital intuition and beyond." in: *Biometric ID Management and Multimodal Communication: Joint COST 2101 and 2102 International Conference, BioID_MultiComm 2009, Madrid, Spain, September 16–18, 2009. Proceedings 2*, 252–9. Berlin: Springer.
- Campi, Massimiliano, Valeria Cera, Francesco Cutugno, Antonella DI LUGGO, Domenico Iovane, Antonio Origlia, et al. 2021. "Chrome project: Representation and survey for AI development." *Diségno-open Access*, 173–7. Milano: FrancoAngeli.
- Canary, Daniel J, and David Rand Seibold. 2010. "Origins and development of the conversational argument coding scheme." *Communication Methods and Measures* 4 (1–2): 7–26.
- Castagna, Federico, Nadin Kökciyan, Isabel Sassoon, Simon Parsons, and Elizabeth Sklar. 2024. "Computational argumentation-based chatbots: a survey." *Journal of Artificial Intelligence Research* 80: 1271–310.
- Cataldo, Violetta, Loredana Schettino, Renata Savy, Isabella Poggi, Antonio Origlia, Alessandro Ansani, et al. 2019. "Phonetic and functional features of pauses, and concurrent gestures, in tourist guides speech." *Audio Archives at the Crossroads of Speech Sciences, Digital Humanities and Digital Heritage* 6: 205–31.
- Cera, Valeria, Antonio Origlia, Francesco Cutugno, Massimiliano Campi. 2018. "Semantically annotated 3d material supporting the design of natural user interfaces for architectural heritage." in *AVI-CH*. Conference at AVI-CH 2018 Workshop on Advanced Visual Interfaces for Cultural Heritage, Castiglione della Pescaia, Italy.
- Chiyah-Garcia, Javier, Alessandro Suglia, Arash Eshghi, and Helen Hastie. 2023. "What are you referring to? evaluating the ability of multi-modal dialogue models to process clarificational exchanges." *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Prague: Association of Computational Linguistics.
- Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Clark, Eve V. 2015. "Common ground." in: *The Handbook of Language Emergence*, 328–53. Chichester: Wiley.
- Clark, Herbert H, and Edward F Schaefer. 1989. "Contributing to discourse." *Cognitive science* 13 (2): 259–94.
- Cohen, Jacob. 1960. "A coefficient of agreement for nominal scales." *Educational and psychological measurement* 20 (1): 37–46.
- Cohen, Philip R, Jerry L Morgan, Martha E Pollack. 1990. *Intentions in Communication*. Cambridge, MA: MIT Press.
- Coseriu, Eugenio. 1985. "Linguistic competence: What is it really?" *The Modern Language Review*, 80 (4): xxv–xxxv.
- Davis, Ernest. 2014. *Representations of Commonsense Knowledge*. California: Morgan Kaufmann.
- De Saussure, Ferdinand. 2004. "Course in general linguistics." *Literary theory: An anthology* 2: 59–71.

- Deng, Yang, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, and Tat-Seng Chua. 2023. "Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration." arXiv: <http://arXiv.org/abs/arXiv:2305.13626>.
- Deng, Yang, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam. 2021. "Unified conversational recommendation policy learning via graph-based reinforcement learning." in: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1431–41. Canada: Virtual Event.
- Di Bratto, Martina, Maria Di Maro, Antonio Origlia, and Francesco Cutugno. 2021. "Dialogue analysis with graph databases: Characterising domain items usage for movie recommendations." in *CLiC-it*. Torino: Academia University Press.
- Di Bratto, Martina, Maria Di Maro, Antonio Origlia, et al. 2024a. "On the use of plausible arguments in explainable conversational AI." in: *Proc. Interspeech 2024*, 4054–8, ISCA Archive.
- Di Bratto, Martina, Antonio Origlia, Maria Di Maro, and Sabrina Mennella. 2024b. "Linguistics-based dialogue simulations to evaluate argumentative conversational recommender systems." *User Modeling and User-Adapted Interaction* 34, 1–31. Berlin: Springer.
- Di Maro, Maria, Antonio Origlia, and Francesco Cutugno. 2021. "Cutting melted butter? common ground inconsistencies management in dialogue systems using graph databases." *Italian Journal of Computational Linguistics* 7(7-1), 2, 157–190.
- Di Maro, Maria, Antonio Origlia, and Francesco Cutugno. 2021. "Conflict search graph for common ground consistency checks in dialogue systems." in: *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue*.
- Di Maro, Maria, Antonio Origlia, and Francesco Cutugno. 2021. "Polarexpress: Polar question forms expressing bias-evidence conflicts in Italian." *International Journal of linguistics* 13 (4): 14–35.
- Di Maro, Maria, Marco Valentino, Anna Riccio, and Antonio Origlia. 2017. "Graph databases for designing high-performance speech recognition grammars." in: *Proceedings of the 12th International Conference on Computational Semantics (IWCS)*. Short papers.
- Domaneschi, Filippo, Maribel Romero, and Bettina Braun. 2017. "Bias in polar questions: Evidence from English and German production experiments." *Glossa: A Journal of General Linguistics* 2 (1): 26.
- Ducamp, Gaspard, Christophe Gonzales, and Pierre-Henri Wuillemin. 2020. "aGrUM/pyAgrum : a Toolbox to Build Models and Algorithms for Probabilistic in Python." in: *10th International Conference on Probabilistic Graphical Models*, Skørping, Denmark, 138 of Proceedings of Machine Learning Research, 609–12.
- Dung, Phan Minh. 1995. "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games." *Artificial intelligence* 77 (2): 321–57.
- Dunne, Paul E and TJM Bench-Capon. 2006. "Suspicion of hidden agenda in persuasive argument." *Frontiers in Artificial Intelligence and Applications*, 144: 329.
- Elliot, Andrew J, and Patricia G Devine. 1994. "On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort." *Journal of Personality and Social Psychology* 67 (3): 382.
- Evans, John David Gemmill. 1977. *Aristotle's concept of dialectic*. Cambridge: Cambridge University Press.
- Forguson, Lynd. 1989. *Common sense*. London: Routledge.
- Gabsdil, Malte. 2003. "Clarification in spoken dialogue systems." in: *Proceedings of the 2003 AAAI Spring Symposium. Workshop on Natural Language Generation in Spoken and Written Dialogue*, 28–35, Technical report.
- Gao, Chongming, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. 2021. "Advances and challenges in conversational recommender systems: A survey." *AI Open* 2: 100–26.
- Ginzburg, Jonathan, and Zoran Macura. 2005. "The emergence of metacommunicative interaction: some theory, some practice." in: *Symposium on the Emergence and Evolution of Linguistic Communication (EELCa05)*, 35, Citeseer.
- Grice, Herbert Paul. 1975. "Logic and conversation." in: *Speech acts*, 41–58. New York: Academic Press.
- Grice, Paul. 1989. *Studies in the way of words*. Cambridge, Massachusetts: Harvard University.
- Gumperz, John J. 1979. "The retrieval of socio-cultural knowledge in conversation." *Poetics Today* 1 (1/2): 273–86.
- Hadjinikolis, Christos, Yiannis Siantos, Sanjay Modgil, Elizabeth Black, and Peter McBurney. 2013. "Opponent modelling in persuasion dialogues." in *Twenty-Third International Joint Conference on Artificial Intelligence*. Beijing: AAAI Press.
- Hadoux, Emmanuel, Anthony Hunter, and Sylwia Polberg. 2023. "Strategic argumentation dialogues for persuasion: Framework and experiments based on modelling the beliefs and concerns of the persuadee." *Argument & Computation* 14 (2): 109–61.
- Hautli-Janisz, Annette, Zlata Kikteva, Wassiliki Siskou, Kamila Gorska, Ray Becker, and Chris Reed. 2022. "Qt30: A corpus of argument and conflict in broadcast debate." in: *Proceedings of the 13th Language Resources and Evaluation Conference. European Language Resources Association (ELRA)*, 3291–300.
- Hayati, Shirley Anugrah, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, and Zhou Yu. 2020. "Inspired: Toward sociable recommendation dialog systems." arXiv: <http://arXiv.org/abs/arXiv:2009.14306>.
- Hicks, Michael Townsen, James Humphries, and Joe Slater. 2024. "Chatgpt is bullshit." *Ethics and Information Technology* 26 (2): 38.
- Higashinaka, Ryuichiro, Masahiro Araki, Hiroshi Tsukahara, and Masahiro Mizukami. 2021. "Integrated taxonomy of errors in chat-oriented dialogue systems." in: *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 89–98, Singapore and online: Association for Computational Linguistics.
- Hinton, Martin, and Jean HM Wagemans. 2023. "How persuasive is ai-generated argumentation? an analysis of the quality of an argumentative text produced by the gpt-3 AI text generator." *Argument & Computation* 14 (1): 59–74.
- Hunter, Anthony, and Matthias Thimm. 2016. "On partial information and contradictions in probabilistic abstract argumentation." in: *Fifteenth International Conference on the Principles of Knowledge Representation and Reasoning*.
- Hunter, Anthony, and Matthias Thimm. 2017. "Probabilistic reasoning with abstract argumentation frameworks." *Journal of Artificial Intelligence Research* 59: 565–611.

- Hunter, Anthony, Lisa Chalaguine, Tomasz Czernuszenko, Emmanuel Hadoux, and Sylwia Polberg. 2019. "Towards computational persuasion via natural language argumentation dialogues." in: *KI 2019: Advances in Artificial Intelligence: 42nd German Conference on AI, Kassel, Germany, September 23–26, 2019, Proceedings* 42, 18–33. Springer.
- Hymes, Dell. 1972. "On communicative competence." in *Sociolinguistics*, 269–93. Harmondsworth: Penguin.
- Iovino, Matteo, Edvards Scukins, Jonathan Styrd, Petter Ögren, and Christian Smith. 2022. "A survey of behavior trees in robotics and AI." *Robotics and Autonomous Systems* 154:104096.
- Kleinberg, Jon M. 1999. "Authoritative sources in a hyperlinked environment." *Journal of the ACM (JACM)* 46 (5): 604–32.
- Kok, Eric M. 2013. *Exploring the practical benefits of argumentation in multi-agent deliberation*. PhD diss., Utrecht University.
- Kok, Eric M, John-Jules Ch Meyer, Henry Prakken, and Gerard AW Vreeswijk. 2010. "A formal argumentation framework for deliberation dialogues." in: *International Workshop on Argumentation in Multi-Agent Systems*, 31–48. Springer.
- Ladd, D Robert. 1981. "A first look at the semantics and pragmatics of negative questions and tag questions." in: *Papers from the Regional Meeting of the Chicago Linguistic Society* (vol. 17), 164–71.
- Lei, Wenqiang, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020. "Interactive path reasoning on graph for conversational recommendation." in: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2073–83.
- Lenat, Douglas B. 1995. "Cyc: A large-scale investment in knowledge infrastructure." *Communications of the ACM* 38 (11): 33–8.
- Lison, Pierre, and Casey Kennington. 2016. "Opendial: A toolkit for developing spoken dialogue systems with probabilistic rules." in: *Proceedings of ACL-2016 System Demonstrations*, 67–72. Berlin: Association for Computational Linguistics.
- Lücking, Andy, Kirsten Bergmann, Florian Hahn, Stefan Kopp, and Hannes Rieser. 2010. "The Bielefeld speech and gesture alignment corpus (SaGA)." in: *LREC 2010 Workshop: Multimodal Corpora-advances in Capturing, Coding and Analyzing Multimodality*.
- Lücking, Andy, Kirsten Bergman, Florian Hahn, Stefan Kopp, and Hannes Rieser. 2012. "Data-based analysis of speech and gesture: the Bielefeld Speech and Gesture Alignment corpus (SaGA) and its applications." *Journal on Multimodal User Interfaces* 7: 5–18.
- MacWhinney, Brian, and William O'Grady. 2015. *The Handbook of Language Emergence*. New Jersey: John Wiley & Sons.
- Malle, Bertram F, and Joshua Knobe. 1997. "The folk concept of intentionality." *Journal of Experimental Social Psychology* 33 (2): 101–21.
- Maro, Maria Di. 2021. *Shouldn't I use a polar question? Proper Question Forms Disentangling Inconsistencies in Dialogue Systems*. PhD diss., Università degli Studi di Napoli Federico II.
- Marro, Santiago, Elena Cabrio, and Serena Villata. 2022. "Graph embeddings for argumentation quality assessment." in: *EMNLP 2022-Conference on Empirical Methods in Natural Language Processing*.
- McCarthy, John. 1984. "Some expert systems need common sense." *Annals of the New York Academy of Sciences* 426 (1): 129–37.
- Mennella, Sabrina, Maria Di Maro, and Martina Di Bratto. 2023. "Common sense knowledge graph generation for information-gap requests in dialogue systems." in: *Proc. of 16th International Cognitive Linguistics Conference*. Heinrich: Heine University Düsseldorf.
- Miceli, Maria, and Cristiano Castelfranchi. 2014. *Expectancy and Emotion*. Oxford: OUP.
- Miller, Tim. 2019. "Explanation in artificial intelligence: Insights from the social sciences." *Artificial Intelligence* 267: 1–38.
- Mirzadeh, Iman, Keivan Alizadeh, Hooman Shahrokhi, Oncel Tuzel, Samy Bengio, and Mehrdad Farajtabar. 2024. *GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models*, <https://arxiv.org/abs/2410.05229>.
- Mora-Cantalops, Marçal, Salvador Sánchez-Alonso, and Elena Garcíia-Barriocanal. 2019. "A systematic literature review on wikidata." *Data Technologies and Applications* 53 (3): 250–68.
- Nguyen, Tuan-Phong, Simon Razniewski, Julien Romero, and Gerhard Weikum. 2022. "Refined commonsense knowledge from large-scale web contents." *IEEE Transactions on Knowledge and Data Engineering* 35 (8): 8431–47.
- Origlia, Antonio, Federico Altieri, Giorgia Buscato, Alice Morotti, Claudio Zmarich, Antonio Rodà, et al. 2018a. "Evaluating a multi-avatar game for speech therapy applications." in: *Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good*, 190–5.
- Origlia, Antonio, Antonio Rodà, Claudio Zmarich, Piero Cosi, Stefania Nigris, Benedetta Colavolpe, et al. 2018b. "Gamified discrimination tests for speech therapy applications." *Book series Studi AISV* 4: 195–216.
- Origlia, Antonio, Antonio Cosi, Antonio Rodà, and Claudio Zmarich. 2017. "A dialogue-based software architecture for gamified discrimination tests." in *GHITALY@ CHItaly*.
- Origlia, Antonio, Francesco Cutugno, Antonio Rodà, Piero Cosi, and Claudio Zmarich. 2019. "Fantasia: a framework for advanced natural tools and applications in social, interactive approaches." *Multimedia Tools and Applications* 78:13613–48.
- Origlia, Antonio, Martina Di Bratto, Maria Di Maro, and Sabrina Mennella. 2022. "A multi-source graph representation of the movie domain for recommendation dialogues analysis." in: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 1297–306.
- Origlia, Antonio, Silvia Rossi, Sergio Di Martino, Francesco Cutugno, and Maria Laura Chiacchio. 2021. "Multiple-source data collection and processing into a graph database supporting cultural heritage applications." *Journal on Computing and Cultural Heritage (JOCH)* 14 (4): 1–27.
- Origlia, Antonio, Renata Savy, Violetta Cataldo, Loredana Schettino, Alessandro Ansani, Isora Sessa, et al. 2019. "Human, all too human: Towards a disfluent virtual tourist guide." in: *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, 393–9.
- Origlia, Antonio, Martina Di Bratto, Maria Di Maro, Sabrina Mennella. 2022. "Developing embodied conversational agents in the unreal engine: The FANTASIA plugin." in: *Proceedings of the 30th ACM International Conference on Multimedia*, 6950–1.

- Origlia, Antonio, Marco Grazioso, Maria Laura Chiacchio, and Francesco Cutugno. 2022. "3d avatars and semantic models annotations for introductory cultural heritage presentations." in: *Proceedings of the 2022 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage*. CEUR-WS. org.
- Paglieri, Fabio. 2004. "Data-oriented belief revision: Towards a unified theory of epistemic processing." in: *Proceedings of STAIRS*, 179–90.
- Paglieri, Fabio. 2005. "See what you want, believe what you like: Relevance and likeability in belief formation." *Want and Like: Motivational and Emotional Roots of Cognition and Action*, 90. Hatfield: AISB.
- Paglieri, Fabio, and Cristiano Castelfranchi. 2004a. "Argumentation and data-oriented belief revision: On the two-sided nature of epistemic change." in: *CMNA IV: 4th Workshop on Computational Models of Natural Argument*. sn, 5–12.
- Paglieri, Fabio, and Cristiano Castelfranchi. 2004b. "Revising beliefs through arguments: Bridging the gap between argumentation and belief revision in MAS." in: *International Workshop on Argumentation in Multi-Agent Systems*, 78–94. Berlin, Germany: Springer.
- Pearl, J. 1985. "Bayesian networks: A model of self-activated memory for evidential reasoning." *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, 329–34.
- Pearl, Judea, and Dana Mackenzie. 2018. *The book of why: the new science of cause and effect*. New York: Basic Books.
- Pollock, John L. 1987. "Defeasible reasoning." *Cognitive science* 11 (4): 481–518.
- Prakken, Henry. 2005. "Coherence and flexibility in dialogue games for argumentation." *Journal of logic and computation* 15 (6): 1009–40.
- Prakken, Henry. 2008. "A formal model of adjudication dialogues." *Artificial Intelligence and Law* 16 (3): 305–28.
- Prakken, Henry. 2018a. "Historical overview of formal argumentation." in: *Handbook of Formal Argumentation*, 73–141. London: College Publications.
- Prakken, Henry. 2018b. *Historical Overview of Formal Argumentation* (vol. 1). London: College Publications.
- Purver, Matthew. 2006. "Clarie: Handling clarification requests in a dialogue system." *Research on Language and Computation* 4: 259–88.
- Purver, Matthew Richard John. 2004. *The Theory and Use of Clarification Requests in Dialogue*. PhD diss., University of London.
- Rienstra, Tjitze, Matthias Thimm, and Nir Oren. 2013. "Opponent models with uncertainty for strategic argumentation." in *Twenty-Third International Joint Conference on Artificial Intelligence*. Beijing: AAAI Press.
- Rodríguez, Kepa Joseba, and David Schlangen. 2004. "Form, intonation and function of clarification requests in German task-oriented spoken dialogues." in: *Proceedings of Catalog (the 8th Workshop on the Semantics and Pragmatics of Dialogue; SemDial04)*.
- Rosenfeld, Sophia. 2011. *Common Sense: A Political History*. Cambridge: Harvard University Press.
- Ruiz-Dolz, Ramon, Joaquin Taverner, Stella M Heras Barberá, and Ana García-Fornes. 2024. "Persuasion-enhanced computational argumentative reasoning through argumentation-based persuasive frameworks." *User Modeling and User-Adapted Interaction* 34 (1): 229–58.
- Russo, Valentina, Azzurra Mancini, Marco Grazioso, and Martina Di Bratto. 2022. "Graph-based representations of clarification strategies supporting automatic dialogue management." *Italian Journal of Computational Linguistics* 8 (8-1).
- Saba, Walid S. 2023. "Stochastic LLMS do not understand language: towards symbolic, explainable and ontologically based LLMS." in: *International Conference on Conceptual Modeling*, 3–19. Berlin: Springer.
- Sap, Maarten, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, et al., "Atomic: An atlas of machine commonsense for if-then reasoning." in: *Proceedings of the AAAI Conference on Artificial Intelligence* (vol. 33), 3027–35.
- Sbisa, Marina. 1987. "Acts of explanation: A speech act analysis." *Argumentation: Perspectives and Approaches* 2:7.
- Schopenhauer, Arthur. 2004. *The art of always being right: Thirty eight ways to win when you are defeated (tb saunders, trans.)*. London: Gibson Square, (Original work published 1831).
- Searle, John R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Stalnaker, Robert. 2002. "Common ground." *Linguistics and philosophy* 25 (5/6): 701–21,
- Stange, Sonja, Hendrik Buschmeier, Teena Hassan, Christopher Ritter, and Stefan Kopp. 2019. "Towards self-explaining social robots. verbal explanation strategies for a needs-based architecture." in: *AAMAS 2019 Workshop on Cognitive Architectures for HRI: Embodied Models of Situated Natural Language Interactions (MM-Cog)*, May 2019. Montreal, Canada.
- Traum, David, Johan Bos, Robin Cooper, Staffan Larsson, Ian Lewin, Colin Matheson, and Massimo Poesio. 1999. "A model of dialogue moves and information state revision." *Tech. Rep., Tech. rept. Deliverable*.
- Van Eemeren, Frans H, and Rob Grootendorst. 2003. "A pragma-dialectical procedure for a critical discussion." *Argumentation*, 17 (4): 365–86.
- Vassiliades, Alexandros, Nick Bassiliades, and Theodore Patkos. 2021. "Argumentation and explainable artificial intelligence: a survey." *The Knowledge Engineering Review* 36: e5.
- Visser, Jacky, Barbara Konat, Rory Duthie, Marcin Koszowy, Katarzyna Budzynska, and Chris Reed. 2020. "Argumentation in the 2016 US presidential elections: annotated corpora of television debates and social media reaction." *Language Resources and Evaluation* 54 (1): 123–54.
- Von Wright, Georg Henrik. 1997. "Explanation and understanding of actions." *Contemporary Action Theory Volume 1: Individual Action*, 1–20. Berlin: Springer.
- Walton, Douglas N. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, State University of New-York Press.
- Walton, Douglas N. 1984. *Logical Dialogue-Games*. Lanham, Maryland: University Press of America.
- Walton, Douglas. 2005. "Justification of argumentation schemes." *The Australasian Journal of Logic* 3: 1–13.
- Walton, Douglas, and Erik CW Krabbe. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. Dordrecht: SUNY Press.
- Webber, Jim. 2012. "A programmatic introduction to Neo4j." in: *Proceedings of the 3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity*, 217–8. Cambridge: ACM.

- Yuan, Tangming, David Moore, and Alec Grierson. 2004. “*Human-computer Debate, a Computational Dialectics Approach*,” Unpublished PhD diss., Leeds Metropolitan University.
- Zang, Liang-Jun, Cong Cao, Ya-Nan Cao, Yu-Ming Wu, and Cun-Gen Cao. 2013. “A survey of commonsense knowledge acquisition.” *Journal of Computer Science and Technology* 28 (4): 689–719.
- Zhou, Pei, Karthik Gopalakrishnan, Behnam Hedayatnia, Seokhwan Kim, Jay Pujara, Xiang Ren, et al. 2021. “Commonsense-focused dialogues for response generation: An empirical study.” in: *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 121–32.