ISSN 2499-4553

IJCoL

Italian Journal of Computational Linguistics

Rivista Italiana di Linguistica Computazionale

Volume 11, Number 2
december 2025
Special Issue

Bridging Theoretical Linguistics and Automated Language Processing: Emerging Synergies and Advances

aA ccademia university press

Registrazione presso il Tribunale di Trento n. 14/16 del 6 luglio 2016

Rivista Semestrale dell'Associazione Italiana di Linguistica Computazionale (AILC)

Associazione Italiana di Linguistica Computazionale

direttore responsabile
Michele Arnese

isbn 9791255001522

Accademia University Press
via Carlo Alberto 55
I-10123 Torino
info@aAccademia.it
www.aAccademia.it/IJCoL_11_2

Accademia University Press è un marchio registrato di proprietà
di LEXIS Compagnia Editoriale in Torino srl

IJCoL

Volume 11, Number 2
december 2025
Special Issue

Bridging Theoretical Linguistics and Automated Language Processing: Emerging Synergies and Advances

Guest Editors:
Alessandro Lenci, Marco Passarotti, Rachele Sprugnoli, Fabio Tamburini

CONTENTS

The Pragmatic Utility of Asking the Right Question in a Recommendation Scenario

Martina Di Bratto*
Università di Napoli Federico II,
Logogramma S.r.l.

Maria Di Maro**
Università di Napoli Federico II

The formal study of argumentation-based dialogue lacks a comprehensive reference framework, particularly from a linguistic perspective. This work addresses part of this gap by analysing whether the relationship between the semantic-syntactic structure of questions and their pragmatic features could increase the utility of the corresponding answers. A preliminary experiment was carried out to identify which forms of information-seeking request best prompt useful human responses, thereby efficiently solving decision problems. Four question types were tested alongside alternative formulations to examine the influence of pragmatic features on answer quality and cognitive effort. Results suggest that pragmatic features can affect the informativeness of answers, with polar questions tending to elicit informative or over-informative responses, while content questions displayed greater variability. Additionally, questions containing polarity items appeared to increase cognitive load, as reflected in response patterns. While these findings are necessarily tentative due to the exploratory nature of the study, they offer promising indications for the design of linguistically grounded, argumentation-based dialogue systems and point to several avenues for further research, including broader experimental designs and an expanded set of pragmatic variables.

1. Introduction

In recent years, dialogue systems have attracted considerable interest and have become an essential component of our everyday interactions with technology. These systems are designed to support natural and efficient communication between humans and machines, enabling a wide range of tasks—from information retrieval to personalised recommendations. Within this broad field, argumentation has emerged as a particularly relevant aspect (Prakken 2018). Dialogue is at the heart of argumentation, providing the dialectical structure through which the acceptability of arguments is assessed. Conversely, argumentation can serve as the foundation for dialogue, where participants exchange viewpoints in order to reach agreement on what constitutes an acceptable outcome (Black, Maudet, and Parsons 2021).

The present study focuses on a specific and still underexplored area: how argumentation principles can be operationalised in recommendation dialogues, especially in the selection and presentation of supporting arguments. While the literature offers extensive theoretical accounts, the practical integration of these insights into real dialogue system interactions remains limited. Our contribution is deliberately scoped as

* E-mail: mdibratto@logogramma.com
** Centro Interdipartimentale URBAN/ECO, Via Trasia, 31, Napoli (NA), 80135, Italy.
E-mail: maria.dimaro2@unina.it

a preliminary step towards bridging this gap, presenting a conceptual framework together with an initial experimental validation. The aim is not to claim a comprehensive solution, but to provide a foundation for more systematic, large-scale investigations.

Formal and computational argumentation are originally studied in dialogue systems with respect 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. Conversely, argumentation-based dialogue deals with phenomena that rely on the dynamic exchange of information, which can shift depending on the participants and the flow of turns. In these contexts, information is distributed among multiple agents, each of whom may choose to share — or withhold — information at different moments based on their individual strategies and objectives. This creates challenges both for communication protocols, which seek to promote fairness and efficiency, and for understanding agent behaviour.

Studies on argumentation-based dialogue are often done against the background of Walton and Krabbe's classification of dialogue types: persuasion, negotiation, inquiry, information-seeking, deliberation, and eristic (Walton and Krabbe 1995). In particular, the authors suggest that dialogues can be classified by what the participants know, what the participants seek to get from the dialogue, and what the dialogue rules are intended to bring about (Black, Maudet, and Parsons 2021). Framing the study into a certain type of dialogue is very useful for the definition of the dialogue moves needed to reach the conversational goal, especially in the field of human-machine interaction. In the specific case of this work, we will focus on specific types of information-seeking and deliberation dialogues in recommendation (Section 2).

The starting point of our research was the pursuit of conversational goals through the use of plausible arguments (Di Bratto et al. 2024; Di Bratto, Di Maro, and Origlia 2024), an illocutionary act corresponding to what happens in argumentation. As noted in (Bermejo-Luque 2019), effective argumentation involves an implicit inference that links the reasons to the conclusion, making the conclusion appear plausible or justified based on the reasons given. For this reason, the selection of the most appropriate items and features supporting the claim 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 tends to present a pattern structured in two phases, *Exploration* and *Exploitation* (E&E), intended as two types of dialogues embedded in each other (Gao et al. 2021). The main goal of a recommendation dialogue is to have the interlocutors, during the exploitation phase, agree on the selection of a specific item due to its supporting arguments. The need to select these arguments constitutes a secondary goal pursued during the exploration phase, while building the Common Ground (Clark 1996; Stalnaker 2002).

The exploitation phase can be intended as a grounded-level dialogue (Krabbe 2003, p. 83). If the recommendation had not been accepted, a shift would occur in the exploration phase. In this case, participants may then move to a meta-dialogue to have a secondary dialogue on whether the movements in the first dialogue can be judged to be correct or not by some criteria (Krabbe 2003; Macagno and Bigi 2020). Thus, a meta-dialogue's primary purpose is to help the first dialogue achieve its end successfully. The structure of the recommendation dialogue is heavily based on argumentation capabilities, whereas the information collected during the exploration phase is used to *support* the proposals used in the exploitation phase, namely activating an argumentative inference connection to conclusions.

The joint purposes of a dialogue — what Searle (2002) refer to as the interlocutors' generic "we-intentions" to engage in a cooperative activity — have been systematically categorised by Walton into the aforementioned seven distinct dialogue types (Walton and Krabbe 1995; Walton 1998; Macagno 2008; Macagno and Bigi 2020). Each type is defined by a specific initial situation and a set of normative goals pursued by the participants. These typologies provide a pragmatic framework to analyse the most recurrent forms of goal-directed dialogical interaction in both everyday and institutional contexts (Dunin-Keplicz and Verbrugge 2001; McBurney and Parsons 2009). In other words, they show how arguments become essentially intertwined with the *dialogical dimension* defined by the dialogical goal that the user proposes through their argument. Furthermore, an argument also has a *pragmatic dimension* since it is part of a dialogue and is grounded on accepted inferential rules, according to the Common Ground shared between the participants. Thus, this dimension mainly refers to how an argument is related to the individuals involved in terms of their wills, beliefs, and commitments (Kecskes 2014). This highlights how the acceptability of the arguments is strictly correlated with the shared information but also to the amount of good pieces of evidence provided in support of them, framing the evaluation of the arguments in the *epistemic dimension* as well (Macagno 2022).

2. Information gathering in recommendation

Building on the theoretical background and focusing on recommendation tasks, we examine dialogical goals by distinguishing between two key phases: *exploration*, characterized by information gathering, and *exploitation*, associated with deliberation. Deliberation dialogues are inherently collaborative, involving participants who work together to reach an agreement on a proposal that addresses a shared problem, while taking into account everyone's interests (Walton 2019). In this context, argumentation plays a central role, as it involves identifying proposals, supporting arguments, and critiques of alternative options (Walton 2010). In contrast, the exploration phase is typically aligned with *information-seeking* dialogues, which start from an asymmetrical situation where one party lacks information known by the other. The goal is to acquire the needed information through interaction (Walton and Krabbe 1995). However, in real-life dialogues, knowledge gaps are dynamic — new information can change the participants' epistemic states and redefine the information needed. As a result, dialogue moves in this phase can be more broadly interpreted as acts of *information sharing*, which include requesting, providing, and offering information (Macagno and Bigi 2017, 2020). From this perspective, information sharing is a more symmetrical and collaborative process than information seeking, making the exploration phase as cooperative as the deliberation phase.

To support both types of dialogues effectively, the dialogue system must select appropriate items and features. This is achieved by collecting and integrating relevant beliefs during the exploration phase. Since gathering pertinent information is critical at this stage, strategies for eliciting user feedback on candidate system beliefs are essential.

A theoretical model is needed to identify and select relevant data as system beliefs. The Data-oriented Belief Revision (DBR) model, introduced by (Paglieri and Castelfranchi 2005), evaluates the reliability and strength of data, and incorporates Toulmin's argumentation model (Toulmin 2003), as it aligns with the belief-changing process. The model identifies four key properties of data based on cognitive reasons: i) **Credibility**: a measure of the number and values of all supporting data; ii) **Importance**: a measure of the epistemic connectivity of the datum; iii) **Relevance**: a measure of the pragmatic utility of the datum; iv) **(Un-)Likeability**: a measure of the motivational appeal of the datum. The selection of a theoretical model for argument selection hinges on the measurability of certain features. As illustrated in previous studies (Di Bratto et al. 2024), our dialogue management module uses a graph database (Webber 2012) containing common knowledge from Linked Open Data. In this context, graphs offer an ideal method for measuring specific features, particularly when using the HITS algorithm (Kleinberg 1999). The algorithm assigns two key types of scores to nodes in the graph: authority and hub. A node is considered an authority if it is linked to by many other nodes, which can be seen as hubs. Authorities represent high-quality or valuable content, such as important sources in a citation network. On the other hand, a hub is a node that links to many authoritative nodes, serving as a guide or reference to valuable content. The more hubs that link to a particular node, the higher its authority score, and the more authoritative that node becomes. These scores surprisingly represent the corresponding properties of credibility and importance, respectively.

The HITS algorithm thus enables the system to analyse the graph by assigning these authority and hub scores to nodes, helping prioritise which data should be disambiguated in dialogues. The plausibility of new information is determined by how well it connects to the user's existing knowledge, and this connectivity influences whether the information is accepted as a belief. As far as relevance and (un-)likeability are concerned, on the other hand, entropy and hard evidence were used. Our system is, therefore, capable of using numerical values derived from the graph to select relevant data, which are then mapped to the cognitive properties in the DBR model, as illustrated in Table 1. Experimental studies supported this model. In Di Bratto et al. (2024), simulated dialogues were evaluated by humans, confirming that arguments selected using this model were perceived as relevant and plausible, thus enhancing recommendation acceptability. A follow-up study (Di Bratto, Di Maro, and Origlia 2024) involved real users interacting with the full dialogue system. Results showed that users found the system likable, easy to use, and controllable, though it was also perceived as unmotivating and slow. These findings, while encouraging, must be interpreted in light of the study's exploratory scope and limited sample. They serve as an initial proof-of-concept, indicating that the proposed integration of argumentation-based selection criteria can be promising, but also highlighting the need for further validation with more diverse question types, broader pragmatic variables, and a richer experimental design.

Furthermore, question formulation strategies were also crucial for effective interaction in the aforementioned experiment. The system chose between open (wh-) and yes/no (polar) questions based on domain size, computed, again, on the basis of the graph database: wh-questions were used when the domain covered was less than 10% of the total candidate domain, as they demand higher cognitive effort and are thus better suited to more focused queries. This strategy aims to optimise both the informativeness of the user's response and the user's cognitive load. At this stage, this solution should be regarded as a working hypothesis rather than a definitive principle, motivating the present work's focus on its linguistic and pragmatic underpinnings, and setting the stage for more extensive follow-up research.

3. A *good answer* with a minimal effort

Once the relevant arguments are correctly selected, the form in which feedback is requested must also be considered. Choosing the most appropriate form of questions in this context is crucial for eliciting *good* answers from users and guiding them towards

Table 1
Cognitive properties described in Paglieri and Castelfranchi (2005) mapped on computational scores

Theoretical Model	*Computational Score*
Credibility: a measure of the number and values of all supporting data, contrasted with all conflicting data, including both external and internal sources.	**Authority score**: identifies the node with a central role in the graph, as it is supported by a solid number of hub nodes.
Importance: a measure of the epistemic connectivity of the datum, i.e., the number and values of the data that would require revision if that datum were revised.	**Hub score**: identifies the node most connected to other knowledgeable nodes.
Relevance: a measure of the pragmatic utility of the datum, i.e., the number and values of the pursued goals that depend on it.	**Entropy**: identifies data that are relevant but uncertain, and thus require feedback to proceed in the dialogue.
(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 it.	**Hard evidence**: reflects how the system's beliefs guide the selection of data that align with user preferences.

achieving the communicative goal. In this work, we use the term *good answer* to refer to a response that directly contributes to resolving the decision problem at hand, by providing information that is both relevant to the questioner's goal and sufficiently precise to advance or complete the current dialogue phase. While informativeness is a necessary component of a good answer, it is not sufficient: an answer may be informative in a general sense but still fail to be *good* if it does not help the dialogue progress towards the intended decision outcome.

Since the recommendation dialogue comprises an information-sharing sub-dialogue (Macagno and Bigi 2017), corresponding to the Exploration phase, the function of the questions is crucial for extracting the relevant amount of information needed to complete the main task. Analysing the functions of questions within a specific dialogue context, such as the exploration phase of a recommendation dialogue, is a linguistically motivated approach aimed at gaining a deeper understanding of the role each question plays in dialogue management.

Table 2
Questions classification in information-sharing dialogue

Questioner's intention	
Information-seeking questions	**Discussion-seeking questions**
initial fact questions	initial opinion questions
specifying fact questions	specifying opinion questions

Savolainen (2020)'s study, for instance, focuses on the analysis of online information seeking and sharing dialogues. The author defined eleven categories of dialogue acts.

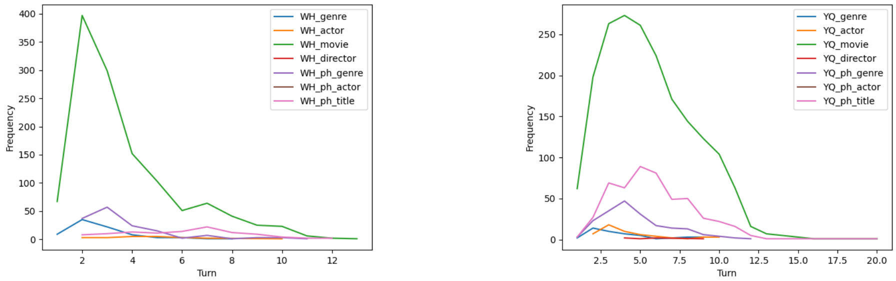
These acts include initial fact question, initial opinion answer, specifying fact question, specifying opinion question, initial fact answer, initial opinion answer, complementary fact answer, complementary opinion answer, answer agreement, answer disagreement, and quote. His conceptual framework was built on more detailed categories proposed by Jeng et al. (2017) and Wang et al. (2014). Although their studies focus on conversations occurring in Q&A communities, the generic categories identified in the above investigations are also used for this purpose. For the objective of this work, we only consider the classification referring to the questions, since the focus is on the exploration phase of an argumentative recommendation dialogue. In examining this phase, Savolainen (2020) focused on three main categories of questions identified by Jeng et al. (2017): information-seeking questions, discussion-seeking questions, and non-questions—though the latter were excluded from his analysis due to the scope of his study. Information-seeking questions are those (Fahy, Crawford, and Ally 2001, p. 4) called **vertical questions**, where a correct answer exists if the right authority or reference may be provided to support the answer; discussion-seeking questions, on the other hand, are called **horizontal questions** since there may not be a right answer for these questions, but instead, more responses are invited to obtain a plausible answer or an opinion that may shed more light on the issue at hand. These two categories have been divided into two further ones starting from the question-question and add-question categories proposed by Wang et al. (2014). The results of this classification are shown in Table 2. As the Table shows, two further functions of information-seeking questions and discussion-seeking questions have been specified: i) Initial fact/opinion questions are those used for seeking factual information or discussion by asking others to indicate their opinions about an issue; ii) Specifying fact/opinion questions are additional questions to obtain more information about an issue. In the context of this work, the focus is on the *horizontal questions*, namely, the discussion-seeking ones. Specifically, the *initial opinion question* will be explored, because of their fundamental role in the exploration phase of a recommendation dialogue. In the early stages of a dialogue, these questions are crucial as they can effectively elicit good responses that help resolve the decision problem quickly. They do this by focusing on a subset of relevant data without cognitively overloading the interlocutor by asking a few heavy questions or multiple simpler questions.

As also noted by several scholars (Hamblin 1976; Karttunen 1977; Groenendijk and Stokhof 1982, 1984), a question is best represented by the set of its appropriate or valid answers. The extent to which an answer resolves a question depends on its *usefulness* in relation to the questioner's goals (Boër and Lycan 1975; Grewendorf 1981; Ginzburg 1995). These statements, moreover, confirm that the degree of the illocutionary force of a speech act (questioner's goal) is strictly correlated to its realisation in terms of perlocutionary effect which can determine the *utility* (or *efficacy*) of the resolving answer. Grewendorf (1981) was the first to propose the use of Bayesian decision theory to evaluate the usefulness of answers but did not detail how to compute it. Van Rooy (2003b) focuses on showing how to relate questions and answers to decision problems and discuss two methods for using decision problems as contextual parameters. The first method argues that while the meaning of a question is context-independent, the concept of resolution is context-dependent, allowing an assertion to resolve a question in relation to a specific decision problem even if it does not provide a complete semantic answer (Boër and Lycan 1975; Ginzburg 1995). The second proposal, which is adopted in this work, assumes that the full meaning of an interrogative sentence is context-dependent. It holds that a question's meaning lies in the set of answers that resolve it, while also maintaining that the interpretation of a wh-interrogative remains underspecified — or intentionally ambiguous — based on its conventional meaning. According to this view, the decision problem plays a crucial role in resolving this ambiguity or underspecification. Thus, the interpretation of a question is determined by the set of answers that resolve the question in relation to the pertinent decision problem (Van Rooy 2003b).

In the experiment proposed in this work, different forms of questions will be evaluated according to the type of answer the user gives (if it is a good answer or not), taking also into consideration their cognitive effort. To the best of our knowledge, the literature lacks focused investigation on this topic; specifically, there appears to be no research directly exploring the relationship between question forms and the cognitive effort required to answer them. Nevertheless, whenever a question is posed, it effectively presents the interlocutor with a choice among a set of information they possess — that is, a range of possible resolving answers, which can vary depending on the size of the domain the question addresses (Van Rooy 2003b). When we start elaborating an answer, we are recollecting an amount of information at our disposal. Answering requires respondents to invest a great deal of cognitive effort for little or no apparent reward (Krosnick 1991). Most of the studies in this field regard mainly the construction of Question surveys, but they rely on a strong theoretical background in cognitive studies which affirm that if questions are difficult to understand respondents are likely to arrive at different interpretations (Belson 1981; Foddy 1993), to *satisfy* (i.e., to provide satisfying rather than optimal answers (Krosnick 1991)), to give incorrect answers (Schober and Conrad 1997) or to refuse answering any further question of the survey (Ganassali 2008). Consequently, an important objective in questionnaire design is to write clear questions and thus to minimise the cognitive effort required to process them. In contrast to these studies, our focus is on the cognitive effort required by the speaker to provide relevant answers that contribute to resolving a decision problem. We assume that the questioner should avoid requesting superfluous or irrelevant information when determining which action to take (Van Rooy 2003b). In the next Section, we present the data elaborated for our investigation and the methodology adopted for the experiment.

4. Data and Methods

The aim of the experiment is to define which types of information-seeking requests a dialogue system should utter to elicit from their human interlocutor a *good answer*, and solve the decision problem rapidly, satisfactorily and with the least cognitive effort. Four typologies of questions have been selected according to what has been observed in an annotated corpus of human-human conversation, i.e., INSPIRED (Hayati et al. 2020). The INSPIRED corpus comprises 1,001 English-language movie recommendation dialogues between pairs of crowd-workers. Each conversation involves a recommender, tasked with exploring the seeker's preferences, and proposing a suitable film. The dataset contains 35,811 utterances, with an average of 10.73 turns per dialogue, as recommenders were instructed to sustain at least 10 turns. Recommender utterances are manually annotated according to a scheme that captures *sociable* recommendation strategies—tactics aimed at persuading seekers to accept a suggestion. These strategies fall into two main categories: preference elicitation (gathering user information) and sociable interaction (delivering the recommendation), corresponding to the two core phases of the dialogue. Analysis by Hayati et al. (2020) shows that sociable strategies positively influence recommendation acceptance and overall dialogue quality. Further



(a) Distribution over turns of wh-questions asked by the recommender to recover information about seeker preference regarding genre, movie, actor or director.

(b) Distribution over turns of positive polar questions asked by the recommender to recover information about seeker preference regarding genre, movie, actor or director.

Figure 1
Distribution over turns of the two main forms of questions annotated in the INSPIRED corpus (Hayati et al., 2020).

work (Di Bratto et al. 2021) links these strategies to domain-specific elements (e.g., movies, genres, plots), offering deeper insights into dialogue dynamics.

With respect to the question types emerging from additional examination of this corpus, negative polar question (NPQ), positive polar question (YQ or PQ), wh-question (WHQ) and alternative question (AQ) have been annotated. Specifically, the results show that PQ and WHQ are the most frequent forms used by the speakers. As Figure 1 shows, the two forms are mostly distributed at the beginning of the conversations given the nature of the dialogues. In fact, recommendation dialogues typically begin with an information-sharing phase, aimed at gathering the most relevant details, and then transition into a deliberation phase, during which an item is proposed and the choice is supported through coherent argumentation.

Given the corpus analysis results, PQ and WHQ have been selected, and three different domains of application have been chosen: movies, cooking, and music. The domain changing was fundamental since the set of questions presented to the participants was semantically the same, i.e., they asked feedback regarding the same arguments but using different linguistic formulas, as shown below:

Qual è il tuo genere preferito di film?[a]
Qual è il tuo genere di cucina preferito?[b]
Qual è il tuo genere di musica preferito?[c]

a What is your favourite movie genre?
b What is your favourite food?
c What is a type of music you would like to listen to?

If these questions were asked on the same domain of application, the participants could have developed a bias risking to invalidate the answers. The arguments were selected based on observations from the INSPIRED corpus in the movie domain. As shown in Figures 1, the distribution of questions was analyzed with respect to the types of arguments addressed by the interrogatives. In particular, we considered questions containing keywords such as *genre*, *actor*, *movie*, and *director*. Figure 1 (a) and (b) illustrate the distribution of wh-questions and PQ, respectively, used by the recommender

Table 3
Forms of questions selected for the experiment

Labels	*Questions*	*Labels + Example*
Wh-pre	Il tuo genere preferito qual è?	Il tuo genere preferito qual è? Mi faresti un esempio?
Wh-post	Qual è il tuo genere preferito?	Qual è il tuo genere preferito? Mi faresti un esempio?
Pq	Hai un genere preferito?	Hai un genere preferito? Mi faresti un esempio?
Pq-pi	Hai qualche genere preferito?	Hai qualche genere preferito? Mi faresti un esempio?

to elicit user preferences. In both cases, questions predominantly target movie-related preferences, with clear peaks in the early turns (around turn 2 for wh-questions and turns 4–5 for PQs). By contrast, questions about actors, directors, and genres appear less frequently and are more evenly distributed across the dialogue. The plots also highlight placeholder (wh) questions, which represent more specific instantiations of general queries (e.g., "*What about action films?*" instead of "*What is your favourite genre?*"). These findings suggest that the dialogue strategy is designed to recover movie-specific information early on, while using placeholders to progressively refine arguments and narrow user preferences in later turns. Consistently, recommenders often begin by asking about a user's favourite genre, preferred "types of movies," or a favourite movie. These arguments were used and reported into the other two domains, i.e., asking about favourite music and food.

Starting from the two main groups of requests, wh-questions and PQs, an alternative form has been proposed for each typology in order to verify whether two questions with the same semantic value could trigger different answers only by adopting different pragmatic features:

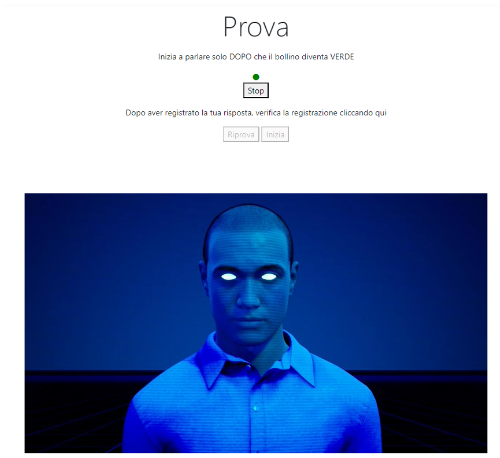
- WQs present a change of topic position. Wh-pre has the focus (i.e., the argument) at the beginning of the interrogatives and Wh-post presents the focus at the end.
- PQs present the canonical reading and the one with the so-called mention-*some* reading (pq-pi), i.e., with the addition of a polarity item, such as *any* or *some*, that makes the question more general and unbiased thanks to the domain extension they provide (Van Rooy 2003a).

As illustrated in Table 3, which presents the set of questions, each group was repeated with the addition of an explicit request for an example, in order to assess whether this would lead to more informative responses from the speaker.

The research hypotheses are the following:

H1 Pragmatic features of questions can help to solve the underspecified semantic meaning of interrogatives thus solving the decision problem through a pragmatically informative answer;

(a) Instructions for the participants presented in the initial page of the software developed for the data collection.



(b) The second page of the site showing the virtual agent Jason asking the question of interest and the mechanism to start the recording of the answer.

Figure 2
First pages of the software developed for the data collection.

H2 Cognitive effort in giving a pragmatically informative answer can vary according to the pragmatic feature of the question and not just to the semantic conventional meaning.

In the next Sections, the experimental setup will be described along with data preparation for the analysis.

5. Experimental Setup

Data collection was carried out online using a software specifically designed for administering the test. Screenshots of the software are shown in Figures 2: in figure (a), the initial page with the necessary information for participants is presented, as illustrated in Box 1; figure (b) shows the interface used to collect participants' answers through questions posed by a virtual human: the virtual human's video asking the questions was pre-recorded and then uploaded to the platform.

> **Box 1:** Good morning! Thank you for choosing to participate. In the following screens, you will meet Jason, a virtual human who will ask you some simple questions on various topics. Record your answers by speaking right after Jason asks his question. The microphone will turn on automatically, and the start of the recording will be indicated by a green dot. Stop the recording by clicking the "Stop" button after you've answered, in order to move on to the next question. Listen carefully, as it won't be possible to replay the question. Make sure you are in a quiet and calm environment. Before starting the test, you will have a chance to do a trial run to ensure everything is working correctly.

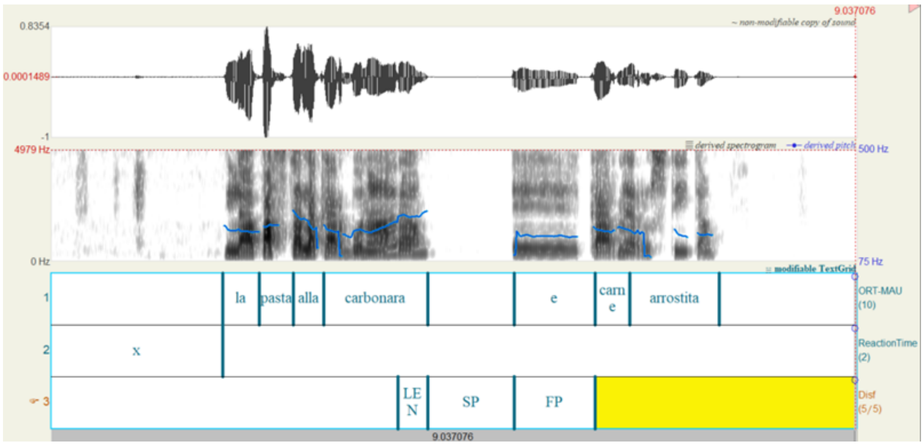


Figure 3
A screenshot of Praat program showing the data levels of annotation

The questions posed by Jason during the experiment are reported in Table 4. The link to the website was shared on Prolific[1], an online crowd-sourcing platform popular among researchers, which allowed us to recruit participants quickly and efficiently. The site showed Jason asking six different questions to each participant. Three of the six questions were randomly chosen, one for each domain, from the set of 72 questions proposed in Table 4 to ensure that at least six responses were collected for each question, for a total of 432 stimuli collected from 144 participants. The other three were filler questions asking participants to rate on a scale from one to ten how much they like movies, cooking, and music[2]. The filler questions served two main purposes: first, to gather information about the participants' level of preparedness in the relevant domain, which could affect their cognitive effort during the responses (this data was not included in the final analysis of the results, which will be discussed in future studies), and second, to act as a distraction from the target questions. As explained in the instructions, participants could start speaking after Jason finished the question. They could then respond freely without any time constraints, which is crucial for controlling the reaction time and evaluating the cognitive effort involved in providing the information. Finally, their responses were recorded and saved for further annotation and analysis, as illustrated in the next Section.

5.1 Data Annotation

Participants' answers were annotated using the Praat software (Boersma and Van Heuven 2001). As shown in Figure 3, three levels of annotation were defined:

1 https://www.prolific.com/
2 **Filler-Movie**: Su una scala da uno a dieci, quanto ti piace guardare film? (En. On a scale from one to ten, how much do you like watching movies?) **Filler-Cooking**: Su una scala da uno a dieci, quanto ti piace cucinare? (En. On a scale from one to ten, how much do you like cooking?) **Filler-Music**: Su una scala da uno a dieci, quanto ti piace ascoltare musica? (En. On a scale from one to ten, how much do you like listening to music?)

i) the Transcription level, where the automatic transcription obtained using the WebMAUS tool [3] has been manually corrected in case errors occurred; this level was considered to calculate answers' Informativeness based on their syntactic cohesion.

ii) The Reaction Time level incorporated a set boundary placed at the beginning of the verbalisation of the analysed answers to measure the time participants were required to respond to questions of a specific type; this level was necessary for the analysis of the cognitive effort of the user to give the answer.

iii) the Hesitation level, where hesitations have been recognised and annotated following the scheme in (Origlia et al. 2019; Schettino, Origlia, and Cutugno 2024). In fact, hesitations are observed to be connected to moments requiring higher cognitive load in speech planning, such as conceptualisation, search for lexical items, and selection of new information in discourse (Levelt 1993; Bortfeld et al. 2001; Barr 2001). Therefore, hesitations are expected to specifically occur when the type of information-seeking request requires a higher cognitive load. This analysis is left to future work.

In this work, we only considered the first two levels of annotation to calculate preliminary results concerning the cognitive effort implied by the use of a specific question type during information-seeking scenarios in a recommendation dialogue. The next Section will illustrate the results collected from this analysis.

6. Results and discussion

Given that the aim of this work was to evaluate the *usefulness* of answers to specific types of question forms, their usefulness can be assessed by the extent to which they help questioners achieve their goals.

Regarding **H1**, from a quantitative point of view, the reduction in the relevant sub-graph for the recommendation task would be the ideal measure. However, since this study involves three different domains, it would require three compatible graphs covering each domain to evaluate this aspect using this metric. While the movies graph is available, the construction of a cooking domain graph is currently ongoing, and the music domain graph would need to be built from scratch. For these reasons, in this work a first *surface* measure is used: the cohesion of the syntactic tree. This is described as the average distance, in tokens, between a dependent and its head, normalised by the number of tokens in the sentence and it has been found to be an indicator of complexity for language comprehension (Liu 2008), also related to cognitive load in linguistic processing (Fedorenko, Woodbury, and Gibson 2013). A measure of *informativeness*, based on syntactic cohesion was presented in Tewari et al. (2020) with the goal of modeling Grice's Maxim of Quantity for short texts, which is adequate for the case of short answers. In this work, the informativeness measure is used as an indicator of *usefulness*. Following the reference work, a text is considered *informative* if, on a normalised scale of -1 to 1, the score is close to 0. The text is considered *under-informative* if the score is close to -1 and *over-informative* if it is close to 1. For the purposes of this work, an utterance is

3 https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface

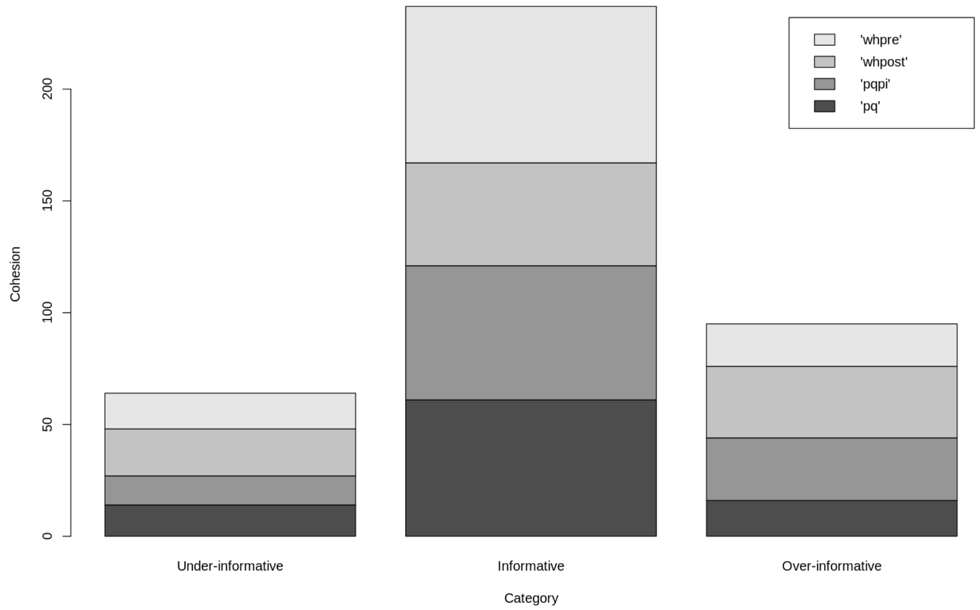


Figure 4
Distributions of answer types over the three informativeness categories.

considered *informative* if it is less than one standard deviation away from the average of the scores in the corpus. The distribution of answer types to different question forms is described in Figure 4. This distribution was investigated using the Chi-square test. The analysis is meant to highlight if the distributions of answers to the different questions forms is different from a uniform distribution, testifying for an influence of the question form over the type of answer. As shown in Table 5, answers to pq and pqpi tend not to be under-informative, and are either informative and over-informative. Moreover, pqpi has a higher chance to elicit over-informative content from the interlocutor, with respect to pq. Wh-questions, on the contrary, may receive answers belonging to any category, with a tendency to informative for wh-pre. An explanation for these results can be found in the literature: based on some observations by Krifka (1989, 1991, 1994), it seems that a Negative polarity item (but also the positive one) is used in an information-seeking question to turn a biased question into an unbiased one. He motivates this analysis by stating that general questions are normally preferred to specific ones. Van Rooy (2003a), on the other hand, tries to give a deeper explanation to this assumption by saying that polar information-seeking general questions are preferred because their *average Informativeness* is higher. Assumption that seems to be confirmed by these preliminary observations. However, a better utility measure to evaluate these hypotheses could be the reduction of the research space in the graph database. In other words, if the information provided by the answer to a pqpi diminishes the size of the objects domain over which the interrogative ranges, this means that the answer is indeed more pragmatically informative than others. To do so, there is a need for further annotated linguistic data and other dataset assembling a graph database for each domain of application.

Table 4
Question forms used in the experiment for the three domains: movie, cooking, and music

Form	*Underspecif.*	*U_example*	*Specifying*	*S_example*	*Iperspecif.*	*I_example*
			Movie			
wh-post	Quale è il tuo genere preferito di film?	Quale è il tuo genere preferito? Per esempio?	Quale è un tipo di film che ti piacerebbe vedere?	Quale è un tipo di film che ti piacerebbe vedere? Per esempio?	Quale è un film che ti piacerebbe vedere?	Quale è un film che ti piacerebbe vedere? Per esempio?
wh-pre	Il tuo genere preferito quale è?	Il tuo genere di preferito qual è? Per esempio?	Ti piacerebbe vedere un film, di che tipo?	Ti piacerebbe vedere un film, di che tipo? Per esempio?	Ti piacerebbe vedere quale film?	Ti piacerebbe vedere un film? Per esempio?
pq	Hai un genere preferito?	Mi fai un esempio del tuo genere preferito?	Hai un tipo di film preferito?	Mi fai un esempio di un tipo di film preferito?	Hai un film preferito?	Mi fai un esempio del tuo film preferito?
pq-pi	Hai qualche genere preferito?	Mi fai un esempio di qualche genere preferito?	Hai qualche tipo di film preferito?	Mi fai un esempio di qualche tipo di film preferito?	Hai qualche film preferito?	Mi fai un esempio di qualche film preferito?
			Cooking			
wh-post	Quale è il tuo genere di cucina preferito?	Quale è il tuo genere di cucina preferito? Per esempio?	Quale è un tipo di cucina che ti piacerebbe assaggiare?	Quale è un tipo di cucina che ti piacerebbe assaggiare? Per esempio?	Quale è un piatto che ti piacerebbe assaggiare?	Quale è un piatto che ti piacerebbe assaggiare? Per esempio?
wh-pre	Il tuo genere di cucina preferito quale è?	Il tuo genere di cucina preferito qual è? Per esempio?	Ti piacerebbe assaggiare una cucina, di che tipo?	Ti piacerebbe assaggiare una cucina, di che tipo? Per esempio?	Ti piacerebbe assaggiare quale piatto?	Ti piacerebbe assaggiare un piatto? Per esempio?
pq	Hai un genere di cucina preferito?	Mi fai un esempio del tuo genere di cucina preferito?	Hai un tipo di cucina preferita?	Mi fai un esempio di un tipo di cucina che preferisci?	Hai un piatto preferito?	Mi fai un esempio del tuo piatto preferito?
pq-pi	Hai qualche genere di cucina preferito?	Mi fai un esempio di qualche genere di cucina preferito?	Hai qualche tipo di cucina preferita?	Mi fai un esempio di qualche tipo di cucina preferita?	Hai qualche piatto preferito?	Mi fai un esempio di qualche piatto preferito?
			Music			
wh-post	Quale è il tuo genere preferito di musica?	Quale è il tuo genere preferito di musica? Per esempio?	Quale è un tipo di musica che ti piacerebbe ascoltare?	Quale è un tipo di musica che ti piacerebbe ascoltare? Per esempio?	Quale è la canzone che ti piacerebbe ascoltare?	Quale è una canzone che ti piacerebbe ascoltare? Per esempio?
wh-pre	Il tuo genere preferito di musica quale è?	Il tuo genere di musica preferito qual è? Per esempio?	Ti piacerebbe ascoltare musica, di che tipo?	Ti piacerebbe ascoltare una musica, di che tipo? Per esempio?	Ti piacerebbe ascoltare quale canzone?	Ti piacerebbe ascoltare una canzone? Per esempio?
pq	Hai un genere preferito di musica?	Mi fai un esempio del tuo genere preferito di musica?	Hai un tipo di musica preferito?	Mi fai un esempio di un tipo di musica preferito?	Hai una canzone preferito?	Mi fai un esempio della tua canzone preferita?
pq-pi	Hai qualche genere preferito di musica?	Mi fai un esempio di qualche genere preferito di musica?	Hai qualche tipo di musica preferito?	Mi fai un esempio di qualche tipo di musica preferito?	Hai qualche canzone preferita?	Mi fai un esempio di qualche canzone preferita?

Table 5
The Chi-square table comparing the distributions of answers over informativeness categories with respect to the uniform distributions

Form	*Category*	*Observed frequency*	*Expected frequency*	*Chi*	*Pr(>Chi)*	
pq	Under-informative	14	33	11.93388	6.615e-03	**
pqpi	Under-informative	13	33	13.22314	3.318e-03	**
whpost	Under-informative	21	33	4.76033	3.495e-01	
whpre	Under-informative	16	33	9.55372	2.395e-02	*
pq	Informative	61	33	25.91736	4.276e-06	***
pqpi	Informative	60	33	24.09917	1.098e-05	***
whpost	Informative	46	33	5.58678	2.172e-01	
whpre	Informative	70	33	45.25620	2.074e-10	***
pq	Over-informative	16	33	9.55372	2.395e-02	*
pqpi	Over-informative	28	33	0.82645	1.000e+00	
whpost	Over-informative	32	33	0.03306	1.000e+00	
whpre	Over-informative	19	33	6.47934	1.310e-01	

Concerning **H2**, the cognitive effort in terms of reaction time was considered. For the statistical analysis, a linear model was fit to the data using speakers and stimulus IDs as grouping factors. The differences in average response times for the considered syntactic forms were then evaluated pair-wise to check for statistical significance. From a deeper inspection, the statistically significant differences observed for pq were mainly due to the pqpi form, as shown in Figure 6 and Table 6. This outcome is very interesting since it seems that the polarity item can be considered as a salient pragmatic feature affecting the cognitive effort of the user. An interpretation may lie on the domain extension that a polarity item triggers (Van Rooy 2003a). Specifically, the more the objects to reason about are, the easier it is to select the relevant one. This means that give a little choice to the user like asking "Hai un film preferito?" [4], can give a major cognitive load than answering to a more general question like "Hai qualche film preferito?" [5], where the domain results wider.

7. Conclusions

The formal study of argumentation-based dialogue comprises a limited number of unifying perspectives and still lacks a comprehensive reference framework (Prakken

4 "Do you have a favourite movie?"
5 "Do you have any favourite movie?"

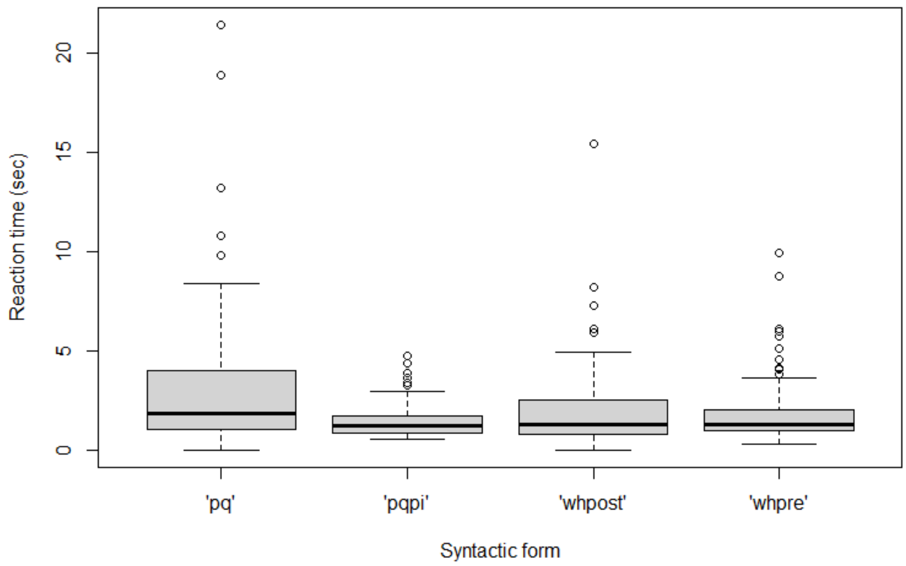


Figure 5
Boxplot of reaction times distribution over the different question forms.

Table 6
Significant results yielded by the pairwise comparison among fixed effects of the Linear Mixed Model having Reaction Times as dependent variable

Comparison	*p-value*
pq/pqpi	0.0033**
pq/whpost	0.0798
pq/whpre	0.0266*
pqpi/whpost	0.6756
pqpi/whpre	0.8853
whpost/whpre	0.9771

2005). This work, following previous work on the matter, aimed at filling part of this gap from a linguistic perspective. For this reason, in this work, a preliminary experiment was carried out to explore the correlation between the semantics of questions and the pragmatics of answers. The central idea is that whether an answer resolves a question depends on how useful it is in relation to the questioner's goals. According to Van Rooy (2003b), this can be achieved by relating questions and answers to decision problems.

The experiment was designed to investigate which types of information-seeking requests might help a dialogue system prompt effective human answers, thereby supporting faster and more satisfactory resolution of decision problems with manageable cognitive effort. Four question forms were tested: two wh-questions differing in topic position, and two PQs, one canonical and one containing a polarity item. For each form, an alternative version was created to examine whether pragmatic variation alone—without semantic change—could influence responses. Participants were recruited via Prolific to answer the questions using software specifically designed for the experiment, which recorded their responses. The audio files were then annotated and analysed to test two hypotheses: first, that the pragmatic features of a question can influence the answer, and second, that the cognitive effort required to answer may vary depending on the type of question.

The results revealed that, in relation to the first hypothesis, answers varied depending on the presence of pragmatic features. Specifically, responses to PQs were generally either informative or over-informative, but rarely under-informative. In contrast, wh-questions could elicit answers across all categories, with a tendency towards being informative for Wh-pre questions. Regarding the second hypothesis, statistically significant differences observed for PQs were primarily due to the pqpi form. This might be attributed to the domain expansion introduced by the polarity item, which in turn increased the cognitive load.

Given the exploratory nature of the study, these findings should be interpreted as indicative rather than definitive. They nonetheless point to promising directions for further investigation, such as analysing hesitation patterns as potential markers of cognitive load and expanding the set of pragmatic variables and question types. Future research will aim to integrate these insights into the FANTASIA framework (Origlia et al. 2019, 2022), contributing to the gradual development of a linguistically motivated approach to argumentation-based dialogue.

Acknowledgments

We would like to express our sincere gratitude to Prof. Fabrizio Macagno from Universidade Nova de Lisboa for his invaluable theoretical support on argumentation and his guidance in validating the question forms employed in this study. We also wish to thank Prof. Antonio Origlia from the University of Naples Federico II for his assistance with the experiment setup and with the data analysis.

References

Barr, Dale J. 2001. Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cavé, I. Guaïtella, and S. Santi, editors, *Oralité et gestualité: Interactions et comportements multimodaux dans la communication*. Paris: L'Harmattan, Paris, pages 597–600.

Belson, William A. 1981. *The design and understanding of survey questions*. Gower, Aldershot, England.

Bermejo-Luque, Lilian. 2019. Giving reasons does not always amount to arguing. *Topoi*, 38(4):659–668.

Black, Elizabeth, Nicolas Maudet, and Simon Parsons. 2021. Argumentation-based dialogue. *Handbook of Formal Argumentation, Volume 2*.

Boër, Steven E. and William G. Lycan. 1975. Knowing who. *Philosophical Studies*, 28(5):299–344.

Boersma, Paul and Vincent Van Heuven. 2001. Speak and unspeak with praat. *Glot International*, 5(9/10):341–347.

Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober, and Susan E. Brennan. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and speech*, 44(2):123–147.

Clark, Herbert H. 1996. *Using Language*. Cambridge University Press, Cambridge, UK.

Di Bratto, Martina, Maria Di Maro, and Antonio Origlia. 2024. On the use of plausible arguments in explainable conversational AI. In *Proceedings of Interspeech 2024: Speech and Beyond*, Kos, Greece, September 2024.

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 *Proceedings of the Eighth Italian Conference on Computational Linguistics (CLiC-it 2021)*, Milan, Italy, June-July 2022.
Di Bratto, Martina, Antonio Origlia, Maria Di Maro, and Sabrina Mennella. 2024. Linguistics-based dialogue simulations to evaluate argumentative conversational recommender systems. *User Modeling and User-Adapted Interaction*, 34.
Dunin-Keplicz, Barbara and Rineke Verbrugge. 2001. The role of dialogue in cooperative problem solving. In *Fifth Symposium on Logical Formalization of Commonsense Reasoning (Common Sense 2001)*, New York, USA, May.
Fahy, Patrick J., Gail Crawford, and Mohamed Ally. 2001. Patterns of interaction in a computer conference transcript. *International Review of Research in Open and Distributed Learning*, 2(1):1–24.
Fedorenko, Evelina, Rebecca Woodbury, and Edward Gibson. 2013. Direct evidence of memory retrieval as a source of difficulty in non-local dependencies in language. *Cognitive science*, 37(2):378–394.
Foddy, William H. 1993. *Constructing questions for interviews and questionnaires: Theory and practice in social research*. Cambridge university press.
Ganassali, Stéphane. 2008. The influence of the design of web survey questionnaires on the quality of responses. *Survey research methods*, 2(1):21–32.
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*, 2:100–126.
Ginzburg, Jonathan. 1995. Resolving questions, I. *Linguistics and philosophy*, 18(5):459–527.
Grewendorf, Günther. 1981. Answering as decision making: A new way of doing pragmatics. In *Possibilities and limitations of pragmatics*. John Benjamins, page 263.
Groenendijk, Jeroen and Martin Stokhof. 1982. Semantic analysis of "wh"-complements. *Linguistics and philosophy*, pages 175–233.
Groenendijk, Jeroen Antonius Gerardus and Martin Johan Bastiaan Stokhof. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam.
Hamblin, Charles L. 1976. Questions in montague english. In *Montague grammar*. Elsevier, pages 247–259.
Hayati, Shirley Anugrah, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, and Zhou Yu. 2020. Inspired: Toward sociable recommendation dialog systems. *arXiv preprint arXiv:2009.14306*.
Jeng, Wei, Spencer DesAutels, Daqing He, and Lei Li. 2017. Information exchange on an academic social networking site: A multidiscipline comparison on researchgate q&a. *Journal of the Association for Information Science and Technology*, 68(3):638–652.
Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and philosophy*, 1:3–44.
Kecskes, Istvan. 2014. *Intercultural pragmatics*. Oxford University Press.
Kleinberg, Jon M. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632.
Krabbe, Erik C. W. 2003. Metadialogues. In Frans H. Van Eemeren, J. Anthony Blair, Charles A. Willard, and A. Francisca Snoeck Henkemans, editors, *Anyone who has a view: Theoretical contributions to the study of argumentation*. Springer, pages 83–90.
Krifka, Manfred. 1989. Polarity phenomena and alternative semantics. In Martin Stokhof and Leen Torenvliet, editors, *Proceedings of the seventh Amsterdam Colloquium*, pages 277–301, Amsterdam, The Netherlands, December 19—22, 1989. Institute for Language, Logic and Information Universiteit van Amsterdam.
Krifka, Manfred. 1991. Some remarks on polarity items. In D. Zaefferer, editor, *Semantic universals and universal semantics*. De Gruyter Berlin, pages 150–189.
Krifka, Manfred. 1994. The semantics and pragmatics of weak and strong polarity items in assertions. In Mandy Harvey and Lynn Santelmann, editors, *Proceedings of the 4th Semantics and Linguistic Theory Conference (SALT)*, pages 195–219, Rochester, , New York, USA, May 6-8, 1994.
Krosnick, Jon A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3):213–236.
Levelt, Willem J. M. 1993. *Speaking: From intention to articulation*, volume 1. Cambridge/London: MIT press.
Liu, Haitao. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2):159–191.
Macagno, Fabrizio. 2008. Dialectical relevance and dialogical context in walton's pragmatic theory. *Informal logic*, 28(2):102–128.

Macagno, Fabrizio. 2022. Argumentation profiles: A tool for analyzing argumentative strategies. *Informal Logic*, 42(1):83–138.
Macagno, Fabrizio and Sarah Bigi. 2017. Analyzing the pragmatic structure of dialogues. *Discourse Studies*, 19(2):148–168.
Macagno, Fabrizio and Sarah Bigi. 2020. Analyzing dialogue moves in chronic care communication: Dialogical intentions and customization of recommendations for the assessment of medical deliberation. *Journal of Argumentation in Context*, 9(2):167–198.
McBurney, Peter and Simon Parsons. 2009. Dialogue games for agent argumentation. *Argumentation in artificial intelligence*, pages 261–280.
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*, 78:13613–13648.
Origlia, Antonio, Martina Di Bratto, Maria Di Maro, and Sabrina Mennella. 2022. Developing embodied conversational agents in the unreal engine: the fantasia plugin. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)*, pages 6950–6951, Lisboa, Portugal, October 10 - 14, 2022.
Origlia, Antonio, Renata Savy, Violetta Cataldo, Loredana Schettino, Alessandro Ansani, Isora Sessa, Alessandra Chiera, and Isabella Poggi. 2019. Human, all too human: Towards a disfluent virtual tourist guide. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization (UMAP 2019)*, pages 393–399, Larnaca Cyprus, June 9 - 12, 2019.
Paglieri, Fabio and Cristiano Castelfranchi. 2005. Revising beliefs through arguments: Bridging the gap between argumentation and belief revision in mas. In *Argumentation in Multi-Agent Systems: First International Workshop, ArgMAS 2004, New York, NY, USA, July 19, 2004, Revised Selected and Invited Papers 1*, pages 78–94. Springer.
Prakken, Henry. 2005. Coherence and flexibility in dialogue games for argumentation. *Journal of logic and computation*, 15(6):1009–1040.
Prakken, Henry. 2018. *Historical overview of formal argumentation*, volume 1. College Publications.
Savolainen, Reijo. 2020. Dialogue processes in online information seeking and sharing: a study of an asynchronous discussion group. *Information Research*, 25(3).
Schettino, Loredana, Antonio Origlia, and Francesco Cutugno. 2024. Though this be hesitant, yet there is method in't: Effects of disfluency patterns in neural speech synthesis for cultural heritage presentations. *Computer Speech & Language*, 85:101585.
Schober, Michael F. and Frederick G. Conrad. 1997. Does conversational interviewing reduce survey measurement error? *Public opinion quarterly*, 61(4):576–602.
Searle, John R. 2002. *Consciousness and language*. Cambridge University Press.
Stalnaker, Robert. 2002. Common ground. *Linguistics and philosophy*, 25(5/6):701–721.
Tewari, Maitreyee, Suna Bensch, Thomas Hellström, and Kai-Florian Richter. 2020. Modelling grice's maxim of quantity as informativeness for short text. In *ICLLL 2020: The 10th International Conference in Languages, Literature, and Linguistics, Japan, November 6-8, 2020*.
Toulmin, Stephen E. 2003. *The uses of argument*. Cambridge university press.
Van Rooy, Robert. 2003a. Negative polarity items in questions: Strength as relevance. *Journal of semantics*, 20(3):239–273.
Van Rooy, Robert. 2003b. Questioning to resolve decision problems. *Linguistics and Philosophy*, 26:727–763.
Walton, Douglas. 2010. Burden of proof in deliberation dialogs. In *Argumentation in Multi-Agent Systems: 6th International Workshop, ArgMAS 2009, Budapest, Hungary, May 12, 2009. Revised Selected and Invited Papers 6*, pages 1–22. Springer.
Walton, Douglas. 2019. How the context of dialogue of an argument influences its evaluation. *Informal Logic a Canadian approach to Argument*, pages 196–233.
Walton, Douglas and Erik CW Krabbe. 1995. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press.
Walton, Douglas N. 1998. *The new dialectic: Conversational contexts of argument*. University of Toronto Press.
Wang, G. Alan, Harry Jiannan Wang, Jiexun Li, Alan S. Abrahams, and Weiguo Fan. 2014. An analytical framework for understanding knowledge-sharing processes in online Q&A communities. *ACM Transactions on Management Information Systems (TMIS)*, 5(4):1–31.
Webber, Jim. 2012. A programmatic introduction to Neo4j. In *Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity (SPLASH '12)*, pages 217–218, Tucson, Arizona, USA, October 19 - 26, 2012.