IJCoL

**Italian Journal
of Computational Linguistics**

**Rivista Italiana
di Linguistica Computazionale**

**Computational Dialogue Modelling:
The Role of Pragmatics and Common Ground in Interaction**

Associazione Italiana di
Linguistica Computazionale
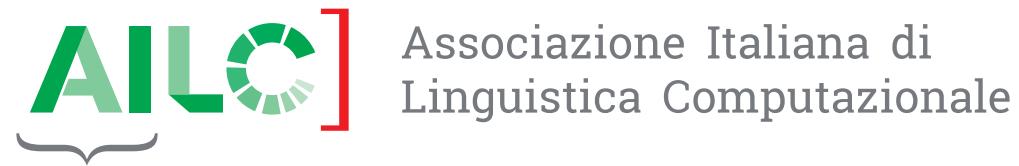
accademia
university
press

IJCoL

Volume 7, Number 1-2
june-december 2021
Special Issue

Computational Dialogue Modelling:
The Role of Pragmatics and Common Ground in Interaction

Invited editors: *Hendrik Buschmeier and Francesco Cutugno*
co-editors: *Maria Di Maro and Antonio Origlia*

CONTENTS

Computational Grounding: An Overview of Common Ground Applications in Conversational Agents

Maria Di Maro[*]
Università di Napoli 'Federico II'

This work reports on the literature on grounding in conversational agents, as one of the pragmatic aspects adopted to ensure a better communicative efficiency in dialogue systems. The paper starts with a general description of the theory of grounding. As far as its computational implications are concerned, grounding phenomena are firstly framed in the common grounding processes described in terms of grounding acts. Secondly, they are considered in the argumentation-related framework within which already grounded information are processed. Open issues and application gaps are finally highlighted.

1. Introduction

In Stalnaker's words, *"when speakers speak, they presuppose things and what they presuppose guides both what they choose to say and how they intend what they say to be interpreted. To presuppose something means to take it for granted as background information – as common ground among the participants during their conversation"* (Stalnaker 2002, p. 701). In fact, communication is a joint activity in which two speakers must share information or, in other words, they must have a common ground, i.e., mutual knowledge, mutual beliefs, and mutual assumptions, as the foundation for mutual understanding (Clark and Brennan 1991). To coordinate on this process, speakers need to update, check, or revise their common ground with a process that constantly evolves through time. The importance of focusing on this communicative process reflects the need to bridge the gap left in the study and development of dialogue systems caused by the lack of insights into the application of pragmatics to conversational agents. Although pragmatics is very important in dialogue, as it is one of the aspects governing interpretation, understanding, and efficiency, its computational application is mainly focused on the study and identification of speech acts (Leech 2003). Furthermore, in the last ten years, semantics has been a more investigated topic within the dialogue systems field with respect to pragmatics, especially as far as the understanding of the correct recognition of the received intent was concerned, as shown in the publications on dialogue systems (Figure 1).

On the other hand, as far as pragmatics is concerned, in the last ten years, the research on Common Ground has started to see a thriving impulse (Figure 2). Nevertheless, a more in-depth analysis of pragmatic phenomena, such as Clarification Requests, related to Common Ground construction and consistency checks in human-machine interaction appears to be a missing spot in the research on dialogue systems.

---

[*] Interdepartmental Center for Advances in Robotic Surgery. Department of Electrical Engineering and Information Technology. University of Naples "Federico II", Italy. E-mail: `maria.dimaro2@unina.it`
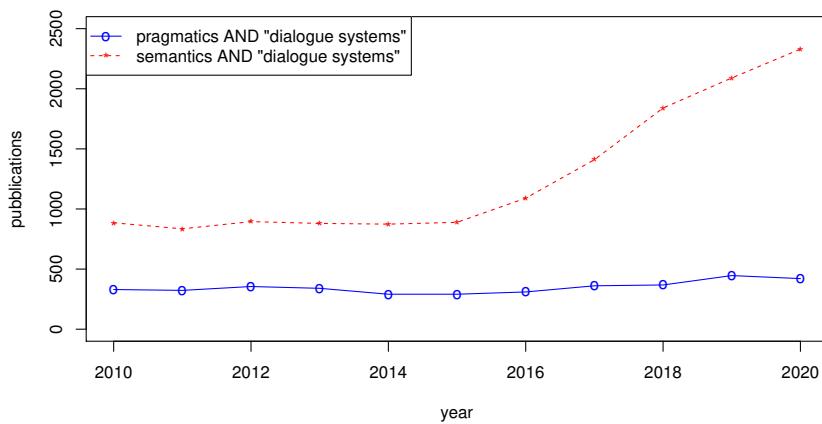
Figure 1
Number of Google Scholar's results on publications about dialogue systems applying semantics versus pragmatics from 2010 to 2020 [Retrieved on 30/04/2021].
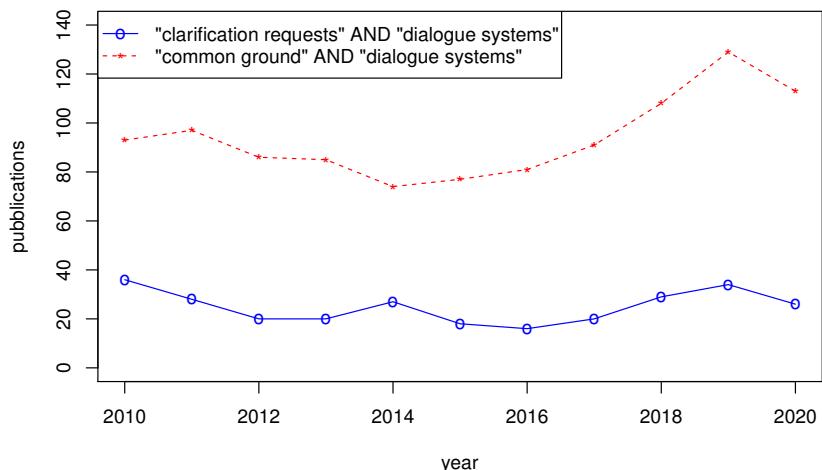
Figure 2
Number of Google Scholar's results on publications about clarification requests and common ground from 2010 to 2020 [Retrieved on 30/04/2021].

Different scholars (Bousquet-Vernhettes, Privat, and Vigouroux 2003; Beun and van Eijk 2004; Purver 2004a; Roque and Traum 2009; Hough, Zarrieß, and Schlangen 2017; Müller, Paul, and Li 2021) highlighted the urge of including pragmatic aspects in their systems to improve the communication process. This need resulted from the users' need to interact with an agent capable of cooperating on the communicative actions.

This survey aims both at presenting a literature review on grounding theories and their application in dialogue systems, and at pointing out pragmatic aspects which still need to find a computational model. The paper is organised as follows: in the next section the theory of grounding is summarised; in section 2.1, the grounding acts reported in (Traum 1994) are explained; in section 3, computational applications of the theory are reported starting from the aforementioned grounding acts to more general works; finally, open issues concerned with the processing of grounding information are presented.

2. Grounding the Grounding Process

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 was 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 in order to understand each other. Common ground, as Clark (2015) acknowledged, can be of four main types: personal, local, communal and specialised. *Personal Common Ground* 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. A part of Personal Common Ground is *Local Common Ground* that is tied to a piece of information obtained from a single exchange with an unknown or known interlocutor. According to Clark (2015), information of this type can be, for instance, the opening hours for a shop, train timetables, and so on. *Communal Common Ground* refers to the amount of information shared with people belonging to the same community, that is to say, people that share general knowledge, knowledge about social background, education (schools attended, levels of education attained), religion, nationality, and language(s). Within a larger community, a smaller one can be found: *Specialised Common Ground* pertains to those people that share particular areas of expertise about some domain of knowledge, such as colleagues, friends, or acquaintances. It is marked by specialised vocabulary of that specific domain, such as medicine, law, and so on.

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*, that is 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). According to Clark and Brennan (1991), the first main form of positive evidence for acceptance are the acts of *acknowledgement* (the complete classification of *grounding acts* is detailed in

Section 2.1), in particular: i) *back-channel responses* that include continuers such as *uh, huh* or *yeah*, used to signal that the utterance has been understood and that there is no need to initiate a repair in the next turn; ii) *assessments* (i.e., *gosh*, *really*) that are usually produced without taking the turn. A second form of positive evidence is the initiation of the *relevant next turn*: suppose A is trying to ask B a question; if B understands it, the answer will be expected in the next turn. Questions and answers constitute adjacency pairs. In other words, once the first part of the adjacency pair is uttered, the second part is considered as conditionally relevant for the next turn. The third and most basic form of positive evidence is *continued attention* provided by an attentive listener. In conversation, people monitor their partner from time to time and immediately adapt to their feedback. If A utters something and notices that B was not paying attention, A could assume that B did not understand him. B must show that he is paying attention through different social signals, like eye gaze or other communicative feedback. A can, therefore, use phatic expressions (i.e., *Are you listening?*, *You know what I mean?*) to understand if B is following, or she can elicit attentive listener feedback in B. On the other hand, B could also want to show his attention by using communicative feedback. Positive evidence of understanding, thus, is provided by communicative feedback and comes with attention that is unbroken or undisturbed (Buschmeier 2018; Buschmeier and Kopp 2018). Furthermore, according to (Clark 1996, p. 147-148), these actions are processed following the concepts of *upward completion*, i.e., *in a ladder of actions, it is only possible to complete actions from the bottom level u through any level in the ladder*, and *downward evidence*, i.e., *in a ladder of actions, evidence that one level is complete is also evidence that all levels below it are complete*.

As argued by Clark and Schaefer (1989), the strength of evidence that B has understood A can depend on several factors, including the complexity of the presentation, the importance of its understanding, and the closeness among the participants. Moreover, since the acceptance phase can be recursive, as B's acceptance to A's presentation needs to be accepted as well, in Traum (1999) the *Strength of Evidence Principle*, introduced in Clark and Schaefer (1989, p. 268), is instead preferred to avoid recursion. This principle states that *"The participants expect that, if evidence e_0 is needed for accepting presentation u_0, and e_1 for accepting presentation of e_0, then e_1 will be weaker than e_0"* (Traum 1999, p. 2). In other words, the evidence is stronger when the need for acceptance is higher. The authors exemplified the principle as follows: A presents a book identification number, *f, six, two*, B accepts it by displaying it verbatim *f, six, two*; then A accepts the B's acceptance by using a weaker evidence like *yes*. Lastly, B accepts the A evidence by proceeding to the next contribution. The traditional version of this principle exhorts speakers not to expend any more effort than they need to get their addressees to understand them with as little effort. Grice (1975) used two maxims of the cooperative principle to account for the communicative effort: according to the maxim of *quantity*, the speaker must not make their contribution more informative than is required, and, according to the maxim of *manner*, they must also be brief and avoid prolixity. In detail, the general principle of least collaborative effort introduced by Clark and Wilkes-Gibbs (1986) was used by the authors to criticise the general speaker economy principle (Brown 1958) which does not always represent the right strategy for grounding. As claimed by Clark and Wilkes-Gibbs (1986), there are three main problems with this principle: i) *time pressure*, speakers tend to limit the effort for planning an utterance which could result in incorrect productions; ii) *errors* that a speakers can make during speaking that need to be repaired; iii) *ignorance* of the personal knowledge and beliefs of the interlocutor can cause improper utterances. Instead, the authors focus on the *minimisation* of the collaborative effort, as *"speakers and addressees try to minimise collaborative effort, the work*
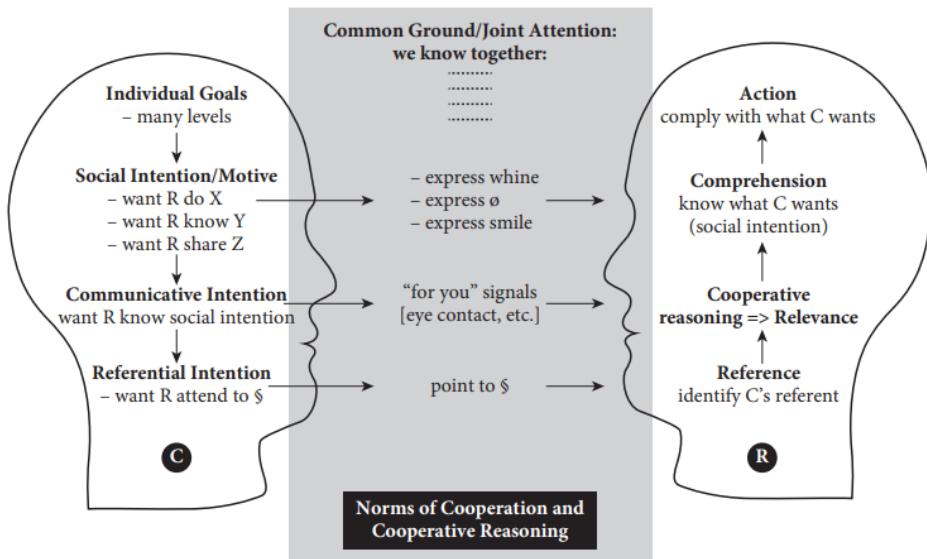
Figure 3
Summary of cooperative model of human communication (C = communicator; R = recipient);
Source: Tomasello (2010); All rights belong to their respective owners.

both speakers and addressees do from the initiation of the referential process to its completion"
(Clark and Wilkes-Gibbs 1986, p. 26).

From a more cognitive point of view, grounding, referred to as an explicit signal of cooperation in dialogue, is also represented in the cooperation model of communication reported by Tomasello (2010) (Figure 2): the communicator C has individual goals, such as goals and values pursued in their life. If for any reason, C feels that the recipient R can be of any help in the achievement of some goals, C will produce specific acts which will bring R to do something, know something, or share something. This is represented by C's social intention, which is expressed through communication. Therefore, a communication act (verbal or not verbal) is mutually manifested in the joint attentional frame. C's communicative intention is consequently shared. C can also draw R's attention to some referential situation in the external world (referential intention) designed to lead R to infer social intentions via processes of cooperative reasoning (Huang 2017, p. 282). On the other hand, R attempt to firstly identify the referent, typically within the space of the common ground, and secondly to infer the social intention, also by relating it to the common ground. Then, assuming that R understands C's social intention, R can decide whether or not to cooperate as expected (Tomasello 2010; Huang 2017).

Whereas the cognitive and linguistic aspects of grounding are naturally clear, its computational applications can be prone to diverse difficulties. Pragmatics can sometimes be subjective, contextual, ambiguous, and its phenomena can be described through one-to-many and many-to-many relationships. Their computational modelling is, therefore, challenging, although different scholars worked on some aspects as it will be summarised in this work. In the next sections, we will focus on grounding

Table 1
Conversation Act Types (Traum 1999, *adapted*); *UU* and *DU* stand respectively for 'utterance unit' and 'discourse unit'.

Discourse Level	Act Type	Sample Acts
Sub UU	**Turn-taking**	take-turn, keep-turn, release-turn, assign-turn
UU	**Grounding**	Initiate, Continue, Acknowledgement, Repair, Cancel, RequestRepair, RequestAcknowledgement
DU	**Core Speech Acts**	Inform, YesNoQuestions, Check, Evaluate, Suggest, Request, Accept, Reject
Multiple DUs	**Argumentation**	Elaborate, Summarise, Clarify, Q&A, Convince, Find-Plan

acts, as they were described in (Traum 1994), and how they could be mapped on research approaches described by different scholars. This work is, therefore, intended as a schematic literature review on some aspects of grounding that can function as a guide and lead to new studies, as research gaps are also highlighted.

2.1 Grounding acts

Traum (1994) provided a computational model of grounding. In his theory, he introduced a description of the so-called *grounding acts*, which are speech acts used to ground the traditional illocutionary speech acts (Austin 1975; Searle 1985). In other words, they correspond to *"the actions performed in producing particular utterances which contribute to this groundedness"* (Traum 1994, p. 31). In particular, he accounted for the protocol determining, for any sequence of grounding acts, whether the content of the communicated utterances is grounded or not. In table 1, its conversation acts are presented, among which the grounding acts are listed.

Each of the grounding acts considered is described as follows:

Initiate. This act is the initial utterance of a discourse unit and usually corresponds to the first utterance of the presentation phase (Clark and Schaefer 1989).

Continue. This represents the continuation of a previous act performed by the same speaker. A continue is expressed in a separate utterance unit, but is syntactically and conceptually part of the same act.

Acknowledgement. An act of acknowledgement is used to claim or demonstrate understanding of a previous utterance. It may be either a repetition or paraphrase of all or part of the utterance, an explicit signal of comprehension such as *ok* or *uh huh*, or an implicit signalling of understanding. Typical cases of implicit acknowledgement are answers to questions. Acknowledgements are also referred to by some as confirmations (Cohen and Levesque 1991) or acceptances (Clark and Schaefer 1989). Traum (1994) prefers the term 'acknowledgement' as a signal of understanding, whereas 'acceptance' is referred to a core speech act signalling agreement with a proposed domain plan.

Repair. A repair is used to change the content of the current discourse unit. This may correspond either to a correction, or it can concern the addition of material. Both solutions will change the interpretation of the speaker's intention. Repair actions should not be confused with domain clarifications. Repairs are concerned merely with the grounding of content. On the other hand, domain clarifications, which modify grounded content, are considered as argumentation acts (Traum 1994). As we will see in the next sections, this particular act that processes grounded information can have interesting computational applications.

Cancel. This act closes the current discourse unit as ungrounded. Rather than repairing the current unit, a Cancel leaves it; the speaker intention must, therefore, be possibly expressed in a new discourse unit.

RequestRepair. A request for a repair is, conversely, uttered by the interlocutor. This is equivalent to a next turn repair initiator or clarification request (Schegloff, Jefferson, and Sacks 1977). Often a RequestRepair can be distinguished from a Repair or Acknowledgement only by intonation. Implicit requests have also been studied (Schettino, Di Maro, and Cutugno 2020).

RequestAcknowledgment. The act is used as an attempt to elicit an Acknowledgement act in the other agent. This invokes a discourse obligation on the listener to respond with either the requested acknowledgement, or an explicit refusal or postponement (i.e., a followup repair or a repair request).

Starting from the description of grounding acts, in the next section, we will explore the studies that concentrated on their computational modelling, or of some of their aspects, in dialogue systems.

3. Computational Grounding

This section reports on pragmatics applied to dialogue modelling and automatic text processing. This branch of computational pragmatics, especially when applied to conversational agents, mostly deals with corpus data, context models, and algorithms for context-dependent utterance generation and interpretation (Huang 2017, p. 326). Nevertheless, conversational agents should be able not only to process local but also global structures of dialogues (Airenti, Bara, and Colombetti 1993). Whereas local structures are involved with linguistic rules (i.e., speech acts, turn-taking, etc.), which can be derived from corpus analysis, global structures refer to the conversation flow, that is the dialogue's action plan and how this is mutually known by dialogue participants (i.e., opening, closing, etc.). Cognitive pragmatics looks at these global structures derived from behavioural games, which in turn derive from grounding processes (Bara 1999). Different authors started including these processes in their dialogue systems architectures, especially as far as evaluating and updating common ground with their human partner. For instance, Roque and Traum (2009) have developed a dialogue system that tracks grounded information in the previous conversation. As a consequence, the dialogue system is capable of selecting its utterances using different types of evidence of the user's understanding (i.e., whether the dialogue system has just submitted material or the user has also acknowledged it, repeated it back, or even used it in a subsequent utterance) (Müller, Paul, and Li 2021).

Table 2
Computational Grounding acts state of the art

Grounding Act	References
Initiate	(Dahlbäck and Jönsson 1998)
Continue	(Schlangen and Skantze 2011) (Visser et al. 2012) (Visser et al. 2014)
Acknowledgement	(Skantze, House, and Edlund 2006) (Wang, Lee, and Marsella 2013) (Visser et al. 2012, 2014) (Eshghi et al. 2015) (Buschmeier 2018) (Buschmeier and Kopp 2018) (Schlangen 2019)
Repair	(Skantze 2008) (Swerts, Litman, and Hirschberg 2000) (Hough and Purver 2012) (Marge and Rudnicky 2015) (Purver, Hough, and Howes 2018) (Di Maro et al. 2019) (Marge and Rudnicky 2019)
Cancel	N/A
RequestRepair	(Gabsdil 2003) (Rodríguez and Schlangen 2004) (Purver 2004a) (Schlangen 2004) (Purver 2006) (Stoyanchev, Liu, and Hirschberg 2014) (Müller, Paul, and Li 2021)
RequestAcknowledgement	(Misu et al. 2011) (Buschmeier and Kopp 2014)

Using grounding strategies in conversational agents led to interesting implementations. One aspect which has not yet been investigated is concerned with the mechanisms of grounding between humans and dialogue systems. Experimental investigations have mostly studied *"how users evaluate the interaction, instead of studying interaction mechanisms"* (Müller, Paul, and Li 2021, p. 3). For instance, Roque and Traum (2009) performed a user study in which subjects interacted with their system and rated how much they felt the system understood them, put effort into understanding them, and gave appropriate responses. Conversely, what most studies do not ask is how a specific dialogue principle, such as the use of a particular type of request, is used by a system to affect user behaviours. Therefore, to learn more about human–machine dialogues mechanisms, it is important to turn to more basic experimental research methods (Müller, Paul, and Li 2021).

With the purpose of providing a structured view concerning the application of grounding in dialogue systems, we start with the classification presented in Traum (1994), and summarised in section 2.1, as a point of departure to understand which aspects of grounding has been modelled over time. As we will see, some of them are more investigated than others, while new other aspects have been considered. In table 2, the studies in which grounding acts are modelled are reported.

Initiate. In the LINLIN dialogue model (Dahlbäck and Jönsson 1998), the initiative is defined as the move whose aim is to introduce a goal. It can have different functions: update, question, answer, discourse opening, discourse continuation, discourse ending. The initiation act in dialogue systems is described in terms of presentation phase, in which form it is presented and which function it shows. This act, as reported in Clark

and Schaefer (1989), introduces something that has to be grounded, via implicit and explicit feedback, to proceed with the exchange. Since this act can also, from other aspects, overlap with other type of grounding acts, such as acknowledgement, as also reported in (Clark and Schaefer 1989), specific details are given in this section, when the corresponding grounding acts are dealt with more in detail.

Continue. For the continue act, defined as the continuation of a previous act by the same speaker, we can account for the studies on incrementality in dialogue. Dialogue processing is, indeed, *incremental*: the processing starts before the input is completed (Kilger and Finkler 1995). Systems designed for incremental processing can process the user inputs, with or without intermediate feedback, before the system output is generated. Incrementality is, for this reason, a research aspect comparable to what has been studied for continue grounding acts. Here, the aspect of grounding is referred to the fact that the previous act is considered as understood and grounded, in that no repair is needed, and therefore the current speaker can go on with the contribution which it refers to. In (Schlangen and Skantze 2011), a model for incremental processing architecture is presented. In their model, this act corresponds to the *incremental unit*, which is the "minimal amount of characteristic input". The incremental processing is composed of a left buffer, a processor, and a right buffer, as represented in Figure 4. The authors also point out for future application the necessity to connect such model for incremental processing and grounding of interpretations in previous processing with models of dialogue-level grounding in the information-state update tradition (Larsson and Traum 2000). For example, the study of self-correction could be a starting point in the connection of sub-utterance processing and discourse-level processing (Ginzburg, Fernández, and Schlangen 2007). Visser et al. (2012, 2014) define incremental understanding in terms of pairs of frames generated every 200 milliseconds, where the first frame is a prediction of the meaning of the complete user utterance, although not yet fully uttered, whereas the second frame is the sub-frame of what the user said so far. Here, feedback of different kinds are analysed before the completion of the utterance.
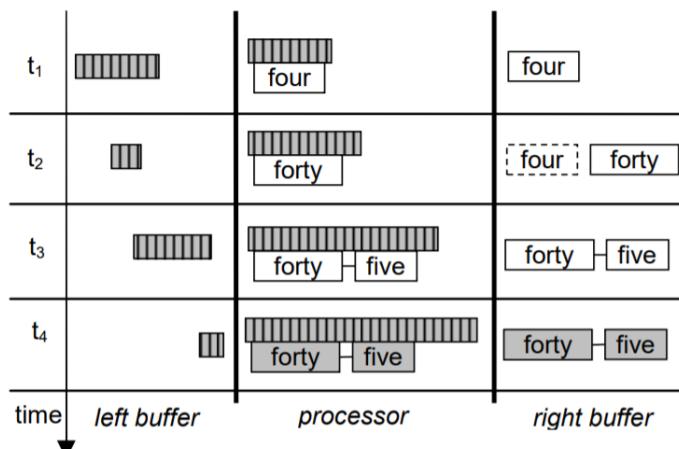

Figure 4
Speech recognition as an example of incremental processing; Source: Schlangen and Skantze (2011); All rights belong to their respective owners.

Acknowledgment. The use of acknowledgement feedback in human-machine interaction has been deeply investigated. In Skantze et al. (2006), for instance, investigate feedback produced both by the user and the system. These were used along with other types of feedback categorising the subjects' responses based on pragmatic meaning. In Wang et al. (2013), a Listener Feedback Model for virtual agents in multi-party conversations is presented, for which the importance of using such systems is underlined. The use of understanding feedback were also studied in incremental models as signals used to update the grounding state (Visser et al. 2012, 2014; Eshghi et al. 2015). In (Buschmeier and Kopp 2018; Buschmeier 2018), acknowledgement acts are studied as attentiveness markers: *"artificial conversational agents should have the capability to use such a mechanism, too, because it would allow them to approach potential or upcoming problems in understanding (and other listening related communicative functions) before they become more serious and require costly repair actions"* (Buschmeier and Kopp 2018, p. 1220). Acknowledgement acts are important for collaborative goals, as also pointed out in Schlangen (2019), and more generally also in Benotti and Blackburn (2021).

Repair. The repair act is aimed at grounding information which may not be clear to either the user or the system. Purver et al. (2018, p. 426), indeed, describe repair as one of *"the primary strategies by which interaction participants achieve and maintain shared understanding"*. This set of strategies is specifically used to highlight and/or resolve miscommunications or potential miscommunications. In Di Maro et al. (2021), different miscommunication scenarios are listed. Starting from Allwood et al. (1992) four basic communicative functions, the communication levels *contact*, *perception*, *understanding*, *intention* were defined. At each level, one or many problems can occur, which are triggered by specific linguistic or informational issues. According to the type of problem, a specific repair strategy can be used to ensure the grounding process to be successful. For instance, in Marge and Rudnick (2015, 2019), recovery strategies were studied in three different scenarios: referential ambiguity (more than one possible action), impossible-to-execute (zero possible actions), and executable (one possible action). In Hough and Purver (2012, p. 143), repair acts are incrementally generated *"in line with psycholinguistic evidence of preference for locality and the availability of access to the semantics of repaired material"*. Prosodic features of repairs are also investigated both from the perspective of users (Di Maro et al. 2019) and of machines (Swerts, Litman, and Hirschberg 2000).

Cancel. Among the investigated breakdown recovery strategies, the cancel act appears to be not so explored. In fact, although the speaker could leave the interaction without giving any further explanation or without trying to repair, thus without modifying their common ground, when the system does not understand them, it is unlikely to find studies which focus on how a system does not try to recover the interaction. This act, in fact, could be more interestingly investigated when its adoption is caused by the analysis of multiple discourse units. In this scenario, when the modification of the common ground interests a previous discourse unit, the repair could imply a higher cost or effort. As a consequence, the dialogue can go on without accepting the last utterance and a re-planning is therefore needed. A parallelism can be drawn, for example, with a car's satellite navigation system that prefers to recompute the route, when repairing the misunderstood action could be too difficult or impossible for the driver. Nevertheless, actions on multiple discourse units are studied in Traum (1994) as argumentation acts where negotiation is important. On the other hand, as far as the last action is concerned, the problem is usually not followed by a cancellation, but by a repair.

RequestRepair. This act is investigated in dialogue systems especially in terms of clarification requests. Clarification requests (CRs) are one of the pragmatic tools used in conversation to prove, ensure, and maintain the mutual understanding of the communicated message between the interlocutors. Purver (2004a, 2004b, 2006) stated that interlocutors initiate a CR when a problem in processing the previous utterance occurs. For this reason, they are also called *anaphoric feedback*, as they refer to what has previously been uttered. Furthermore, CRs are considered as meta-communicative tools as well, since they function as an acknowledgement of the level of understanding of an utterance (Ginzburg and Macura 2005). The use of CRs is also described in terms of cognitive-pragmatic instruments adopted for grounding purposes. As pointed out by scholars, such as Clark (1996), to pursue the aim of succeeding in their joint activity, interlocutors need to ground what has been communicated. Among the scholars who pursued the intent of categorising different types of CRs, Purver (2004b) classified CRs according to form and reading, where *form* refers to the surface form, such as when i) an element from the previous utterance is used in the request (*reprise*), ii) an element from the previous utterance is used in combination with a wh-interrogative pronoun (*wh-reprise*), or iii) when a reformulation or a generic question is adopted (*non-reprise*). *Reading*, conversely, refers to the compromised item that the request questions about, such as a constituent or a clause. This classification established a precise way to describe how CRs can be automatically recognised or selected by a system and opened the way to further investigations, also concerning the causes and problems triggering the initiation of such requests. For instance, Rodriguez and Schlangen (2004) introduce the notion of *problem*, causing the instantiation of a CR. In fact, different kinds of problems, such as acoustic or lexical ones, can determine the adoption of a different informative CR.

Clarification is then a fundamental part of the grounding process. Through the pragmatic tool of CRs the interlocutors can maintain the mutual understanding of the communicated message during a conversation. Clarifications are usually uttered in a context of miscommunication. Following Hirst et al. (1994), miscommunication can be partitioned into three different types: Misunderstanding, non-understanding, and misconception.

Misunderstandings are not immediately detected, since the hearer thinks that what has been understood is the right message, but it is not the one the speaker intended to convey.

The second type of miscommunication is non-understanding that occurs when the hearer finds the message uttered by the speaker ambiguous, or, as Gabsdil (2003) noticed, when the hearer is uncertain about the interpretation given to the message. In this case, even the form in which the requests are formulated can vary. Uncertain interpretations can coarsely be associated with single polar questions, whereas ambiguous understanding is more likely to result in alternative questions or wh-questions. Furthermore, non-understanding in general can occur on several different communicative levels, ranging from establishing contact among the dialogue partners to the intended meaning or function of the utterance in context, as previously also pointed out. Clark (1996) listed four basic levels of communication in a framework that represents the interaction as a joint activity of the dialogue participants: i) execution/attendance, ii) presentation/identification, iii) signal/recognition, iv) proposal/consideration. As Gabsdil (2003) pointed out, on the lowest level, dialogue participants establish a communication channel, which is then used to present and identify signals on level two. On level three, these signals are interpreted before their communicative function is evaluated on the proposal/consideration level. The framework of joint actions requires that dialogue participants coordinate their individual actions on all of those levels.

Gabsdil (2003) combined the cause of non-understanding with Clark's four levels of communication, giving some examples and organising a coarse-grained classification of clarifications. Connected to these levels, two main readings for clarifications were proposed by Ginzburg and Cooper (2001). Their "clausal reading" can be related to the presentation/identification level and their "constituent reading" to the signal/recognition level. Clausal readings are used *"to confirm the content of a particular sub-utterance"* (Ginzburg and Cooper 2001, p. 1), and it can roughly be paraphrased as "Are you asking/asserting that X?" or "For which X are you asking/asserting that Y?". Constituent readings, on the other hand, *"elicit an alternative description or ostension to the content (referent or predicate etc.) intended by the original speaker of the reprised sub-utterance."* (Ginzburg and Cooper 2001, p. 1).

Misconceptions, finally, occur when the *"hearer's most likely interpretation suggests that beliefs about the world are unexpectedly out of alignment with the speaker's"* (Hirst et al. 1994). Clarifications in response to misconceptions usually convey extra-linguistic information like surprise or astonishment.

As already anticipated, CRs can occur in different forms and readings. The correlation between form and function of CRs has also been investigated by Rodriguez and Schlangen (2004), who presented a multidimensional classification of CRs forms and a fine-grained correlation between them and their functions. The study has been carried out in a corpus of German task-oriented dialogues, the "Bielefeld Corpus" [1], which contains 22 dialogues consisting of about 3962 turns, and 36,000 words. In the experimental setup, a dialogue participant was supposed to give instructions to the interlocutor to build a model plane. The authors pointed out some features used to describe the surface form of CRs. Concerning the attribute *Mood*, the possible values are declarative, polar questions, alternative questions, wh-questions, imperative and others; for *Completeness* are particles (*Pardon?*), partial fragments or complete sentences; for *Relation* are literal repetition of the unclear part, the addition of a part to the repetition, reformulation of the problematic utterance, or independent (i.e., no part of the utterance are repeated or reformulated); finally, for *Boundary tone* are rising or falling intonation.

Rodriguez and Schlangen (2004) posed the foundation for the identification of problems that could cause misunderstanding, taking into account the CRs readings proposed earlier, but trying to define them in a more fine-grained way. The authors devised a multidimensional classification scheme where form and function are meta-features taking sub-features as attributes. They start from the models of Clark (1996) and Allwood (1995) concerning the four levels of communication mentioned before, adding other types of sub-levels. Each of those levels is a possible locus for communication problems. This dimension specifies the extent and severity of the problem. The extent, as the authors argued, describes whether a specific CR points to a problematic element in the problematic utterance or not. The severity, on the other hand, describes which action the CR initiator requests from the interlocutor: the CR initiator can ask for a reformulation of the problematic utterance, probably triggered by a complete understanding failure, or they can ask for a confirmation of the previous hypothesis of which they are not certain. The scholars also classified the answers to CRs that can be i) yes/no answers, ii) repetitions or reformulations of the unclear element, iii) elaborations of the problematic utterance with the addition of new elements, iv) word definitions, or, lastly, v) no reaction at all. As a consequence, the satisfaction of CR initiators to the reaction

---

1 http://www.sfb360.uni-bielefeld.de/

of the CR addressee can be classified as happy or unhappy, according to the right or wrong interpretation of the CR.

Stoyanchev et al. (2014) point out how important it is for the communicative efficiency in human-machine interaction to have clarification requests which are not generic but *targeted*, in that they are based on contextual information. For instance, in Müller et al. (2021), rephrasing strategies are used to ask for correctness before grounding the received information.

As it will pointed out at the end of this section, other types of misunderstanding repair strategies to be considered are more classifiable as related to argumentation acts of grounded information, a field of research that is becoming worth exploring.

RequestAcknowledgment. In the "Media Equation" (Reeves and Nass 1996), it is hypothesised that *"people will give more spontaneous back-channels to a spoken dialogue system that makes more spontaneous back-channel-inviting cues than a spoken dialogue system that makes less spontaneous ones"*. Based on this hypothesis, Misu et al. (2011) presented the basis for a dialogue system capable of eliciting back-channels from users. For this purpose, they constructed a dialogue-style TTS which makes use of back-channel-inviting cues, whose application resulted in the more user's spontaneous back-channels, informative for the system. Similarly, Buschmeier and Kopp (2014) defined when the system should elicit feedback in the user in order to avoid undesirable dialogue states. In fact, the system needs feedback when i) its belief about the user's mental state is not informative enough; ii) its belief about the user's mental state has not changed in a long time; iii) its belief about the user's mental state is different from a desired one deriving from a previous communicative action by the agent. In Buschmeier and Kopp (2018), the same result as in Misu (2011) was reported: participants provided more feedback with an attentive listener agent, that is with agent capable of a) interpreting communicative listener feedback from users, b) adapting their production to the users' needs, whose interpretation is based on the processed feedback, and c) eliciting feedback through feedback elicitation cues when needed. The use of such feedback is moreover important to other grounding acts, such as *initiate*.

In this section, we focused on defining a parallel between theory and application, by describing some works on dialogue systems which explicitly applied grounding acts in the dialogue. As a take-home message, it can be pointed out that theory was diversely adapted to the available technology and different new methodologies were implemented. A perfect mapping between theory and application has not yet been reached. Some aspects of grounding were therefore more investigated than others, and some others became crucial. In general, the importance of the grounding process has been variously highlighted, starting from uncertainty signalisation (Fernández et al. 2007; Hough and Schlangen 2017), to different degrees of grounding (Roque and Traum 2008; Roque 2009; Petukhova et al. 2015), to the use of grounding in dialogue systems evaluation (Curry, Hastie, and Rieser 2017; Zou 2020). The research on dialogue systems, in fact, has always underlined the need to test and evaluate their functionality and performances. Nevertheless, the evaluation of dialogue systems has always been a problematic task to carry out. When Turing (1950) suggested the imitation game as a possible evaluation of the intelligence a machine can show, he was thinking of replacing the question whether a machine is able to think with its imitation capabilities. The concept of thinking has always been difficult to define. Instead, the imitation game could actually be a valid and answerable question to pose. To answer this question positively, the evaluator should not be able to tell the difference between the machine

and the human interlocutor, in that the machine succeeded in imitating intelligent human-like behaviour. Here, the concept of intelligence needs some in-depth consideration. Gottfredson (Gottfredson 1997, p. 13) defined *intelligence* as the *"ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience"*. As we may easily comprehend, this definition is far from the possibility a machine can have to imitate some behaviours. If the aim is not only to reproduce, but also to evaluate some intelligent aspects a machine could have, we may need to adopt different tests. Therefore, the Turing test, although sometimes still used, can conversely represent the desirable goal of an intelligent agent, which shows behaviours that are human-like, rather than an evaluation tool for system performances. At the same time, the question that could here be raised is whether we really want a system to be completely indistinguishable from human beings and why we want that. Conversely, we might want systems capable of showing their specific intelligent features which might be suitable for *artificial beings* only. Similarly to Turing, Schatzmann et al. (2005) evaluate two aspects: i) human-like system's responses; ii) how well the user models cover the variety of the user population in the training data. Even here, what is missing is a shareable framework to carry out this evaluation and an in-depth description of how the system is actually working.

Whichever is the way we imagine our dialogue system to be, the evaluation should rather consider some specific traits of what we call intelligence, or better, in this case, of interactional intelligence. With *interactional intelligence*, we mean the ability to recognise intentions, beliefs, and aptitudes towards the dialogic exchange and the ability to respond appropriately (Levinson 1995; Buschmeier 2018). As we will see in section 4, this capability makes the system argumentation-skilled. For goal-oriented dialogue systems, the completeness of the task, dialogue length, and user satisfaction are usually taken into account. On the other hand, for general purpose dialogue systems, approaches like next utterance classification and word perplexity are preferred (Serban et al. 2018). To the present day, fully satisfactory automatic classification metrics for dialogue systems do not exist. Nevertheless, the combination of different methodologies could lead to better results. Grounding acts can, in this sense, also be used as a methodology to evaluate dialogue system's performances. More specifically, Curry et al. (2017) report a comparison between systems using explicit feedback and systems using implicit feedback. In Zou (2020), on the other hand, evaluation techniques are compared and faults are highlighted in that not *"all aspects of dialogue from naturalness and coherence to long-term engagement and flow"* are captured. One possible evaluation metrics could consider usability principles (Dix et al. 2003), namely learnability, flexibility, robustness. Specifically, as far as the robustness principle is concerned, that is the level of support that the system provides to the user in completing and assessing a task successfully, dialogue systems can make their internal states observable through verbal or non-verbal interaction, thus via grounding acts. More in detail, when problems occur in information processing, the observable character of such states can be utilised to recover the problems. In section 4, the need for this type of analysis will be better detailed.

3.1 Latest datasets for grounding acts

In order to make the process of grounding modelling possible, dialogue datasets are needed. In the past years, many dialogue datasets have been collected to study grounding and grounding-related problems (Serban et al. 2018). Nevertheless, the latest corpora collections are particularly important to mention as they are mostly concerned with collecting large amount of data in order to be used to train dialogue systems

with machine learning, which indeed need more data when compared to past collections. Different techniques can be used to model and train dialogue systems: whereas some use online learning (Liu and Mazumder 2021), reinforcement learning (Pietquin 2007; Young et al. 2010), probabilistic reasoning (Skantze 2007; Stoyanchev, Lison, and Bangalore 2016; Rossignol, Pietquin, and Ianotto 2010), or graph representations (Liu and Mei 2020; Mi et al. 2020; Chaudhuri, Rony, and Lehmann 2021), many grounding phenomena are learned and modelled in conversational agents via machine learning algorithms. It is important to point out that grounding can be better observed in spontaneous conversation, as eliciting it can be easier for some aspects (i.e., feedback) rather than for others. For this purpose, in the past, there have been works on agents interacting with humans applying improvisation (Bruce et al. 2000; Martin, Harrison, and Riedl 2016; Winston and Magerko 2017). Nevertheless, there are not so many corpora collecting such spontaneous dialogues, and the ones available are also far too small for machine learning purposes (Busso and Narayanan 2008). In the SPOLIN corpus (Cho and May 2020), 6,760 English Improv dialogues, comprising 90,000 turn pairs, have been collected. The *improvisational theatre* dialogues considered here are important for grounding purposes, as in this form of theatre everything is performed without a script, a scenery, or other established environment; for this reason, everything must be grounded via interactions. The specific aim of this dataset was to study *yes-ands* turns, where an acknowledgement act was combined with a new next relevant contribution. Similarly, different common grounding phenomena, like the ones described in Traum (1994), are observable in the collection presented in Udagawa and Aizawa (2019, 2020), comprising 6,760 dialogues, and whose aim is to be adopted in the training of end-to-end dialogue systems. End-to-end dialogue systems, in fact, are usually based on neural networks (Shang, Lu, and Li 2015; Vinyals and Le 2015; Sordoni et al. 2015; Dodge et al. 2016; Serban et al. 2016) and need large amount of data. For the same purposes, Chen et al. (2021) collected 10K human-to-human dialogues containing 55 distinct user intents. The few amount of appropriate dialogue corpora for grounding applications in dialogue systems in various languages can be still considered as the Achilles' heel of the data-driven research, like the machine learning-based one.

4. "What the heck are you saying?" Corrective dialogues and grounded information

As reported in the previous sections, different scholars highlighted the urge of including grounding processing in their systems, for which argumentation of grounded information needs more investigation. In this section, the attention will be focused on grounding-related corrective dialogues. In this context, the argumentative nature of some of such dialogues, in the form of Common Ground Inconsistencies, will also be taken into account.

Among the most investigated grounding aspects, corrective dialogues have drawn much attention as their adoption improves the communication process. This resulted from the users' need to interact with an agent capable of cooperating with the communicative actions. Human interlocutors always contribute with questions, answers, and feedback (Beun and van Eijk 2004). For instance, a corrective dialogue is a particular type of dialogue occurring when: i) the user notices an error in the system and corrects it; ii) the user changes their mind; iii) the user's beliefs are in contradiction with the system's beliefs and expectations. In the first two cases, the corrective dialogue is initiated by the user (it corresponds to the grounding act of Repair), whereas, in the last case, it is initiated by the system (it corresponds to the grounding act of RequestRepair) (Bousquet-Vernhettes, Privat, and Vigouroux 2003). One example of corrective dialogue

in human-machine interaction is the one presented in Beun and van Eijk (2004). The authors focused on a particular communicative problem related to conceptual discrepancies between a computer system and its user. Starting from the assumption that both the system and its user have a mental representation of a domain, the *mental* representation of the system, e.g., the ontology, contains conceptualisations that are made explicit in a formal language. Despite their possible incompleteness and inaccuracy, this information can be used to trace the system's reasoning about concepts, items, and their properties. Most importantly, this representation also allows the detection of conceptual discrepancies, arising when the system observes that the user applies an incorrect action to a particular object. The authors also stated that, although feedback of different kinds are now generally used in such systems, there is still no accurate *mathematical theory* for natural communicative behaviours and their computational model to human-machine interaction, especially as far as conceptual discrepancies are concerned. What is still missing is, therefore, a reference model guiding the adoption of a specific type, content, and form of the feedback that has to be generated in a particular situation (Beun and van Eijk 2004).

While conceptual discrepancies can be concerned with the last dialogue state whose inaccuracy can lead to a RepairRequest + Repair or directly to a Repair act, some inconsistency can also refer to a previous stage of the interaction, as in Khouzaimi et al. (2015). In this case, Traum (1994) considers the consequent acts as argumentation ones, as already grounded information are now being negotiated. The linguistic activity of argumentation is pragmatically regulated by a sequence of purposive speech acts in conflict (Walton and Godden 2006), as it represents the discussion of opposing ideas to find the truth, namely dialectics. Dialectics in dialogue systems can be framed in the field of formal and computational argumentation, where two main research topics are listed: argumentation-based inference and argumentation-based dialogue. Argumentation-based inference concentrates on establishing what conclusions can be drawn starting from incomplete or inconsistent information. Argumentation-based inference models work similarly to Hegel's dialectic, since they investigate statements from a logical point of view without considering multiple participants. Historically, the first one who described an Abstract Argumentation Framework was Dung (1995). On the other hand, Pollock (1987) first established the basis for formal argumentation-based inference.

Argumentation-based inference is different from argumentation-based dialogue, in that the former is a formal method which is applied to a single entity to decide about the truth of an argument. On the other hand, argumentation-based dialogue considers problems arising from dialogues among different agents. In such cases, information is, in fact, distributed among the agents, who may or may not be willing to share it at different points in time due to individual strategies and goals. A solid argumentation-based dialogue theoretical framework is, in fact, still missing because of the complexity of the phenomenon in question: *"the study of argumentation-based dialogue consists of a variety of different approaches and individual systems, all exciting work but with few unifying accounts or general frameworks"* (Prakken 2017, p. 53). Among the types of dialogue that are studied in argumentation-based studies, we mention persuasion, negotiation, information seeking, deliberation, inquiry, and quarrel (Walton 1984; Walton and Krabbe 1995). These classes, however, are not meant to be absolute, as multiple goals may be present during a single dialogue. Among the ones listed, persuasion dialogues appear to have been studied the most (Yuan, Moore, and Grierson 2004; Prakken 2008). As far as deliberation dialogues are concerned, the collaboration, here, takes place to find an optimal solution to a problem for which the involved agents have not yet a solution.
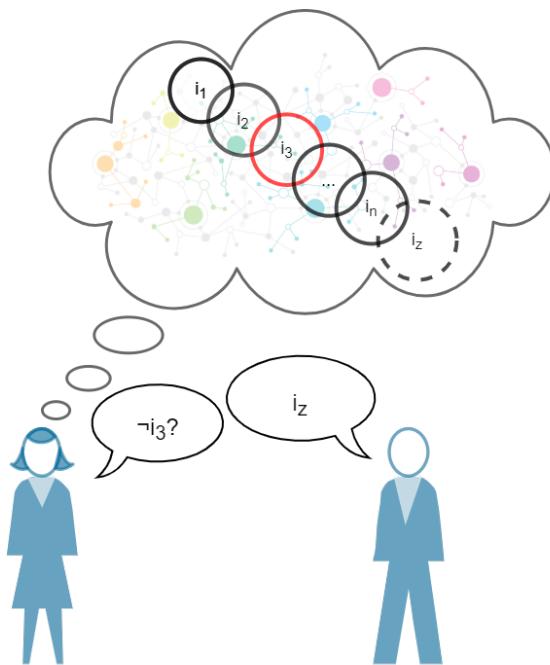
Figure 5
Representation of the Common Ground CRs elicitation scenario

For this type of dialogue, an interesting result was found. In case of a two-agents system adhering strictly to the communication protocol, forming their claims on the basis of their knowledge and adopting a collaborative attitude, it was demonstrated that the agreed solution is always acceptable to both parties (Black and Atkinson 2010). This results from employing argumentation, whose usefulness in dialogue systems, designed for deliberation, was demonstrated in Kok et al. (2010).

The problem that characterises argumentation-based dialogue with respect to argumentation-based inference is, therefore, the presence of different agents in the setting. This introduces multiple, not necessarily aligned, knowledge and, possibly, conflicting goals in the pursuit of a solution to a problem. Pragmatic strategies adopted in such situations are to be investigated, as they are generally concerned with grounded information. Based on the analysis of map-tasks, whose structure can be compared to deliberation dialogue for their goals, an argumentation-based act trigger was identified (Di Maro et al. 2021), namely Common Ground Inconsistencies, which can lead the interlocutor to the adoption of clarification requests, as its corresponding argumentation-based act. Similarly to the aforementioned conceptual discrepancies, Common Ground Inconsistencies refer to problems with grounded information.

In Figure 5, a Common Ground Inconsistency scenario eliciting a Common Ground clarification request (Common Ground CR) is displayed. With Common Ground CR we refer to clarification requests with an argumentative function. In fact, they do not help the speaker ground a piece of information, but they refer to previous discourse units, where that piece of information was already grounded. In the current state of the dialogue, a new evidence clashes with the grounded one, and, therefore the Common

Ground CR is uttered[2] (Di Maro 2021; Di Maro, Origlia, and Cutugno 2021a). As in Figure 5, in the mind of the female agent A, the Communal Common Ground is stored to guide the process of accumulating information in the Personal Common Ground. The information (i_1, i_2, i_3, ..., i_n) are communicated by the male agent B to A, and sequentially stored in her Personal Common Ground. When B utters a new information i_z, this is represented as a new item candidate to be part of the Personal Common Ground. This representation has generates a bias/evidence conflict (Domaneschi, Romero, and Braun 2017), in that the presence of the new item i_z in the Personal Common Ground clashes with the presence of another item i_3, whose validity is now questioned. This conflict represents a Common Ground Inconsistency and is translated in the Common Ground CR $\neg i_3$?, whose form, function and illocutive effect are reported in Di Maro et al. (2021b). As also highlighted in Di Maro et al. (2021), polar questions are especially important to express Common Ground Inconsistencies, in that their epistemic stance is clearly expressed compared to other types of questions (Domaneschi, Romero, and Braun 2017). Finally, differently from other CRs, Common Ground CRs do not necessary refer to the immediately previous utterance, but to previously - correctly or wrongly - grounded information.

5. Conclusion

In Human-Machine interaction, the study and application of pragmatic aspects has covered few phenomena, although their importance was recognised in various linguistic studies. On the one hand, error handling and requests for clarification have always had a central role, since the correct understanding and the consequent task completion of the system are the desired goals. On the other hand, back-channels and acknowledgement feedback have also been investigated to ensure grounding. If commercial systems try to identify possible mistakes which can be caused by users or by technology limits, their ability to understand the real cause of problems to adequately signal them and let the human user correct them is still a frontier not exhaustively explored. The complexity of possible misunderstanding and conflicting situations makes it necessary to study the communicative strategies used to efficiently handle the related interaction problems.

As mentioned at the beginning of this work, the aim pursued here was also to stimulate further investigations and applications of pragmatics, and especially grounding, in conversational agents, by underlying application gaps. In fact, whereas semantics has been a more investigated topic within the dialogue systems field with respect to pragmatics, where speech acts modelling drew more attention. Furthermore, although CRs and corrective dialogues are widely studied in linguistics, their application in dialogue systems is still limited, especially when referred to already grounded information. Further investigations on grounding-related problems concerning dialogue states which do not necessary correspond to the current dialogue state but to previous steps of the dialogue history are therefore needed. This could, moreover, expand the study on argumentation-based dialogues leading to the foundation of a shared theoretical framework.

---

2 We are aware that Clarification Requests are generally used to correctly update the common ground. Nevertheless, the term *Common Ground CR* refers here, as in the mentioned studies, to requests used to check what is already stored in the common ground.

References

Airenti, Gabriella, Bruno Giuseppe Bara, and Marco Colombetti. 1993. Conversation and behavior games in the pragmatics of dialogue. *Cognitive Science*, 17(2):197–256.

Allwood, Jens. 1995. An activity based approach to pragmatics. In *Abduction, belief and context in dialogue: Studies in computational pragmatics*. John Benjamins.

Allwood, Jens, Joakim Nivre, and Elisabeth Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–26.

Austin, John Langshaw. 1975. *How to do things with words*, volume 88. Oxford University Press.

Bara, Bruno Giuseppe. 1999. *Pragmatica cognitiva: i processi mentali della comunicazione*. Bollati Boringhieri.

Benotti, Luciana and Patrick Blackburn. 2021. Grounding as a collaborative process. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 515–531, Online, April.

Beun, Robbert-Jan and Rogier M. van Eijk. 2004. Conceptual discrepancies and feedback in human-computer interaction. In *Proceedings of the conference on Dutch directions in HCI*, page 13. Association for Computing Machinery, New York, NY, United States, June.

Black, Elizabeth and Katie Atkinson. 2010. Agreeing what to do. In *International Workshop on Argumentation in Multi-Agent Systems*, pages 12–30, Toronto, Canada, May. Springer.

Bousquet-Vernhettes, Caroline, Régis Privat, and Nadine Vigouroux. 2003. Error handling in spoken dialogue systems: toward corrective dialogue. In *ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems*, Château d'Oex, Vaud, Switzerland, August.

Brown, Roger. 1958. How shall a thing be called? *Psychological review*, 65(1):14.

Bruce, Allison, Jonathan Knight, Samuel Listopad, Brian Magerko, and Illah R. Nourbakhsh. 2000. Robot improv: Using drama to create believable agents. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, volume 4, pages 4002–4008, San Francisco, CA, USA, August. IEEE.

Buschmeier, Hendrik. 2018. *Attentive Speaking. From Listener Feedback to Interactive Adaptation*. Ph.D. thesis, Faculty of Technology, Bielefeld University, Bielefeld, Germany.

Buschmeier, Hendrik and Stefan Kopp. 2014. When to elicit feedback in dialogue: Towards a model based on the information needs of speakers. In *International Conference on Intelligent Virtual Agents*, pages 71–80, Boston, August. Springer.

Buschmeier, Hendrik and Stefan Kopp. 2018. Communicative listener feedback in human-agent interaction: Artificial speakers need to be attentive and adaptive. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, pages 1213–1221, Stockholm, Sweden. International Foundation for Autonomous Agents and Multiagent Systems.

Busso, Carlos and Shrikanth S. Narayanan. 2008. Scripted dialogs versus improvisation: Lessons learned about emotional elicitation techniques from the iemocap database. In *Ninth annual conference of the international speech communication association (ISCA)*, Brisbane, Australia, September.

Chaudhuri, Debanjan, Md Rashad Al Hasan Rony, and Jens Lehmann. 2021. Grounding dialogue systems via knowledge graph aware decoding with pre-trained transformers. In *European Semantic Web Conference*, pages 323–339, Online, August. Springer.

Chen, Derek, Howard Chen, Yi Yang, Alex Lin, and Zhou Yu. 2021. Action-based conversations dataset: A corpus for building more in-depth task-oriented dialogue systems. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online, June 6–11.

Cho, Hyundong and Jonathan May. 2020. Grounding conversations with improvised dialogues. page 2398–2413, Online, July 5 - 10.

Clark, Eve V. 2015. Common ground. In *The Handbook of Language Emergence*. Wiley, Chichester, UK, page 328–353.

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

Clark, Herbert H. and Susan E. Brennan. 1991. Grounding in communication. In Lauren B.
    Resnick, John M. Levine, and Stephanie D. Teasley, editors, *Perspectives on Socially Shared
    Cognition*, pages 222–233, Washington, DC, USA. American Psychological Association.
Clark, Herbert H. and Edward F. Schaefer. 1989. Contributing to discourse. *Cognitive science*,
    13(2):259–294.
Clark, Herbert H. and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process.
    Cognition, 22(1):1–39.
Cohen, Philip R. and Hector J. Levesque. 1991. Confirmations and joint action. In *Proceeding of the
    International Joint Conferences on Artificial Intelligence Organization (IJCAI)*, pages 951–959,
    Sydney, Australia, August.
Curry, Amanda Cercas, Helen Hastie, and Verena Rieser. 2017. A review of evaluation techniques
    for social dialogue systems. In *Proceedings of the 1st ACM SIGCHI International Workshop on
    Investigating Social Interactions with Artificial Agents*, pages 25–26, Glasgow, UK, 13 November.
Dahlbäck, Nils and Arne Jönsson. 1998. A coding manual for the linköping dialogue model.
    unpublish manuscript.
Di Maro, Maria. 2021. "Shouldn't I use a polar question?" Proper question forms disentangling
    inconsistencies in dialogue systems. *Upublished Dissertation, Università degli Studi di Napoli
    Federico II*.
Di Maro, Maria, Hendrik Buschmeier, Stefan Kopp, and Francesco Cutugno. 2021. Clarification
    requests negotiating personal common ground. In *Proceedings of the XPRAG.it (Poster)*, Online,
    July. URL = https://osf.io/tmjrf/.
Di Maro, Maria, Antonio Origlia, and Francesco Cutugno. 2021a. Common ground
    inconsistencies in dialogue systems: conflict patterns implied by polar question forms
    (submitted). *Speech Communication*.
Di Maro, Maria, Antonio Origlia, and Francesco Cutugno. 2021b. Polarexpress: Polar question
    forms expressing bias-evidence conflicts in Italian. *International Journal of Linguistics*,
    13(4):14–35.
Di Maro, Maria, Jana Voße, Francesco Cutugno, and Petra Wagner. 2019. Perception breakdown
    recovery in computer-directed dialogues. In *Proceedings of the First International Seminar on the
    Foundations of Speech (SEFOS 2019)*, Sønderborg, Denmark, December.
Dix, Alan, Alan John Dix, Janet Finlay, Gregory D. Abowd, and Russell Beale. 2003.
    Human-computer interaction. Pearson Education.
Dodge, Jesse, Andreea Gane, Xiang Zhang, Antoine Bordes, Sumit Chopra, Alexander Miller,
    Arthur Szlam, and Jason Weston. 2016. Evaluating prerequisite qualities for learning
    end-to-end dialog systems. In *Proceedings of the International Conference on Learning
    Representations*, San Juan, Puerto Rico, May 2 - 4.
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).
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):321–357.
Eshghi, Arash, Christine Howes, Eleni Gregoromichelaki, Julian Hough, and Matthew Purver.
    2015. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th
    International Conference on Computational Semantics*, pages 261–271, London, UK, April.
Fernández, Raquel, Andrea Corradini, David Schlangen, and Manfred Stede. 2007. Towards
    reducing and managing uncertainty in spoken dialogue systems. In *Proceedings of the 7th
    International Workshop on Computational Semantics (IWCS07)*, Tilburg, Netherlands, January.
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*,
    pages 28–35, Palo Alto, California, March.
Ginzburg, Jonathan and Robin Cooper. 2001. Resolving ellipsis in clarification. In *Proceedings of
    the 39th Annual Meeting of the Association for Computational Linguistics*, pages 236–243, Toulouse,
    France, July.
Ginzburg, Jonathan, Raquel Fernández, and David Schlangen. 2007. Unifying self-and
    other-repair. In *Decalog 2007: proceedings of the eleventh Workshop on the Semantics and Pragmatics
    of Dialogue*, Rovereto, Italy, May 30 - June 1. Rotooffset Paganella.
Ginzburg, Jonathan and Zoran Macura. 2005. The emergence of metacommunicative interaction:
    Some theory, some practice. In *Proceedings of the 2nd International Symposium on the Emergence*

and Evolution of Linguistic Communication, pages 35–40, Hatfield, UK, April.

Gottfredson, Linda S. 1997. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24:13–23.

Grice, Herbert Paul. 1975. Logic and conversation. In *Speech acts*. Brill, pages 41–58.

Grice, Paul. 1989. *Studies in the way of words*. Harvard University.

Hirst, Graeme, Susan McRoy, Peter Heeman, Philip Edmonds, and Diane Horton. 1994. Repairing conversational misunderstandings and non-understandings. *Speech Communication*, 15(3-4):213–229.

Hough, Julian and Matthew Purver. 2012. Processing self-repairs in an incremental type-theoretic dialogue system. In *Proceedings of the 16th SemDial workshop on the semantics and pragmatics of dialogue*, pages 19–21, Paris, France, September.

Hough, Julian and David Schlangen. 2017. It's not what you do, it's how you do it: Grounding uncertainty for a simple robot. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 274–282, Vienna, Austria, March. IEEE.

Hough, Julian, Sina Zarrieß, and David Schlangen. 2017. Grounding imperatives to actions is not enough: A challenge for grounded nlu for robots from human-human data. In *GLU 2017 International Workshop on Grounding Language Understanding*, pages 88–91, Stockholm, Sweden, 25 August.

Huang, Yan. 2017. *The Oxford Handbook of Pragmatics*. Oxford University Press.

Khouzaimi, Hatim, Romain Laroche, and Fabrice Lefevre. 2015. Turn-taking phenomena in incremental dialogue systems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1890–1895, Lisbon, Portugal, September.

Kilger, Anne and Wolfgang Finkler. 1995. Incremental generation for real-time applications. *Tech. rep. RR-95-11. Saarbrücken, Germany: Deutsches Forschungszentrum für Künstliche Intelligenz*.

Kok, Eric M., John-Jules Ch. Meyer, Henry Prakken, and Gerard A.W. Vreeswijk. 2010. A formal argumentation framework for deliberation dialogues. In *International Workshop on Argumentation in Multi-Agent Systems*, pages 31–48, Toronto, Canada, May. Springer.

Larsson, Staffan and David R. Traum. 2000. Information state and dialogue management in the trindi dialogue move engine toolkit. *Natural language engineering*, 6(3-4):323–340.

Leech, Geoffrey. 2003. Pragmatics and dialogue. In *The Oxford Handbook of Computational Linguistics*. Oxford University Press.

Levinson, Stephen C. 1995. Interactional biases in human thinking. In *Social intelligence and interaction*. Cambridge University Press, pages 221–260.

Liu, Bing and Sahisnu Mazumder. 2021. Lifelong and continual learning dialogue systems: learning during conversation. *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-2021)*, February.

Liu, Bing and Chuhe Mei. 2020. Lifelong knowledge learning in rule-based dialogue systems. *arXiv preprint arXiv:2011.09811*.

Marge, Matthew and Alexander Rudnicky. 2015. Miscommunication recovery in physically situated dialogue. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 22–31, Prague, Czech Republic, September.

Marge, Matthew and Alexander I. Rudnicky. 2019. Miscommunication detection and recovery in situated human–robot dialogue. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 9(1):1–40.

Martin, Lara J., Brent Harrison, and Mark O Riedl. 2016. Improvisational computational storytelling in open worlds. In *International Conference on Interactive Digital Storytelling*, pages 73–84, Los Angeles, CA, USA, November. Springer.

Mi, Jinpeng, Jianzhi Lyu, Song Tang, Qingdu Li, and Jianwei Zhang. 2020. Interactive natural language grounding via referring expression comprehension and scene graph parsing. *Frontiers in Neurorobotics*, 14.

Misu, Teruhisa, Etsuo Mizukami, Yoshinori Shiga, Shinichi Kawamoto, Hisashi Kawai, and Satoshi Nakamura. 2011. Toward construction of spoken dialogue system that evokes users' spontaneous backchannels. In *Proceedings of the SIGDIAL 2011 Conference*, pages 259–265, Portland, Oregon, June.

Müller, Romy, Dennis Paul, and Yijun Li. 2021. Reformulation of symptom descriptions in dialogue systems for fault diagnosis: How to ask for clarification? *International Journal of Human-Computer Studies*, 145:102516.

Petukhova, Volha, Harry Bunt, Andrei Malchanau, and Ramkumar Aruchamy. 2015. Experimenting with grounding strategies in dialogue. In *Proceedings of the 19th Workshop on the*

Semantics and Pragmatics of Dialogue (SEMDIAL 2015 goDIAL), Gothenburg, Sweden, August.

Pietquin, Olivier. 2007. Learning to ground in spoken dialogue systems. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 4, pages IV–165, Honolulu, HI, USA, June. IEEE.

Pollock, John L. 1987. Defeasible reasoning. *Cognitive science*, 11(4):481–518.

Prakken, Henry. 2008. A formal model of adjudication dialogues. *Artificial Intelligence and Law*, 16(3):305–328.

Prakken, Henry. 2017. Historical overview of formal argumentation. *IfCoLog Journal of Logics and their Applications*, 4(8):2183–2262.

Purver, Matthew. 2004a. Clarie: The clarification engine. In *Proceedings of the 8th Workshop on the Semantics and Pragmatics of Dialogue (Catalog)*, pages 77–84, Barcelona, Spain, July.

Purver, Matthew. 2004b. *The Theory and Use of Clarification Requests in Dialogue*. Ph.D. thesis, King's College, University of London, London, UK.

Purver, Matthew. 2006. Clarie: Handling clarification requests in a dialogue system. *Research on Language and Computation*, 4(2-3):259–288.

Purver, Matthew, Julian Hough, and Christine Howes. 2018. Computational models of miscommunication phenomena. *Topics in cognitive science*, 10(2):425–451.

Reeves, Byron and Clifford Nass. 1996. *The media equation: How people treat computers, television, and new media like real people*. Cambridge university press Cambridge, UK.

Rodríguez, Kepa Joseba and David Schlangen. 2004. Form, intonation and function of clarification requests in german task-oriented spoken dialogues. In *Proceedings of the 8th Workshop on the Semantics and Pragmatics of Dialogue*, Barcelona, Catalonia, Spain, July.

Roque, Antonio. 2009. *Dialogue management in spoken dialogue systems with degrees of grounding*. University of Southern California.

Roque, Antonio and David Traum. 2008. Degrees of grounding based on evidence of understanding. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 54–63, Columbus, Ohio, USA, June.

Roque, Antonio and David Traum. 2009. Improving a virtual human using a model of degrees of grounding. In *Twenty-First International Joint Conference on Artificial Intelligence*, Pasadena, California, USA, July.

Rossignol, Stéphane, Olivier Pietquin, and Michel Ianotto. 2010. Simulation of the grounding process in spoken dialog systems with bayesian networks. In *International Workshop on Spoken Dialogue Systems Technology*, pages 110–121, Shizuoka, Japan, October. Springer.

Schatzmann, Jost, Kallirroi Georgila, and Steve Young. 2005. Quantitative evaluation of user simulation techniques for spoken dialogue systems. In *6th SIGdial Workshop on DISCOURSE and DIALOGUE*, Lisbon, Portugal, September.

Schegloff, Emanuel A, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.

Schettino, Loredana, Maria Di Maro, and Francesco Cutugno. 2020. Silent pauses as clarification trigger. In *Laughter and Other Non-Verbal Vocalisations Workshop: Proceedings (2020)*, Bielefeld, Germany, October.

Schlangen, David. 2004. Causes and strategies for requesting clarification in dialogue. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL 2004*, pages 136–143, Cambridge, Massachusetts, USA, April 30 - May 1.

Schlangen, David. 2019. Grounded agreement games: Emphasizing conversational grounding in visual dialogue settings. *Computing Research Repository Journal*, August.

Schlangen, David and Gabriel Skantze. 2011. A general, abstract model of incremental dialogue processing. *Dialogue & Discourse*, 2(1):83–111.

Searle, John R. 1985. *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

Serban, Iulian Vlad, Ryan Lowe, Peter Henderson, Laurent Charlin, and Joelle Pineau. 2018. A survey of available corpora for building data-driven dialogue systems: The journal version. *Dialogue & Discourse*, 9(1):1–49.

Serban, Iulian Vlad, Alessandro Sordoni, Yoshua Bengio, Aaron C. Courville, and Joelle Pineau. 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Conference of the Association for the Advancement of Artificial Intelligence*, volume 16, pages 3776–3784, Phoenix, Arizona, USA, February.

Shang, Lifeng, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational*

Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, July.

Skantze, Gabriel. 2007. Making grounding decisions: Data-driven estimation of dialogue costs and confidence thresholds. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, pages 206–210, Antwerp, Belgium, September.

Skantze, Gabriel. 2008. Galatea: A discourse modeller supporting concept-level error handling in spoken dialogue systems. In *Recent Trends in Discourse and Dialogue*. Springer, pages 155–189.

Skantze, Gabriel, David House, and Jens Edlund. 2006. User responses to prosodic variation in fragmentary grounding utterances in dialog. In *Ninth International Conference on Spoken Language Processing*, Pittsburgh, Pennsylvania, USA, September.

Sordoni, Alessandro, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Denver, Colorado, May–June.

Stalnaker, Robert. 2002. Common ground. *Linguistics and philosophy*, 25(5/6):701–721.

Stoyanchev, Svetlana, Pierre Lison, and Srinivas Bangalore. 2016. Rapid prototyping of form-driven dialogue systems using an open-source framework. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 216–219, Los Angeles, September.

Stoyanchev, Svetlana, Alex Liu, and Julia Hirschberg. 2014. Towards natural clarification questions in dialogue systems. In *AISB symposium on questions, discourse and dialogue*, volume 20, April.

Swerts, Marc, Diane Litman, and Julia Hirschberg. 2000. Corrections in spoken dialogue systems. In *Sixth International Conference on Spoken Language Processing*, Beijing, China, October.

Tomasello, Michael. 2010. *Origins of human communication*. MIT press.

Traum, David R. 1994. A computational theory of grounding in natural language conversation. Technical report, Rochester Univ NY Dept of Computer Science.

Traum, David R. 1999. Computational models of grounding in collaborative systems. In *Psychological Models of Communication in Collaborative Systems: Papers from the AAAI Fall Symposium*, pages 124–131, North Falmouth, MA, USA, November.

Turing, A. M. 1950. Computing machinery and intelligence. *Mind*, LIX(236):433–460, 10.

Udagawa, Takuma and Akiko Aizawa. 2019. A natural language corpus of common grounding under continuous and partially-observable context. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 01, pages 7120–7127, Hilton Hawaiian Village, Honolulu, Hawaii, USA, January 27 – February 1.

Udagawa, Takuma and Akiko Aizawa. 2020. An annotated corpus of reference resolution for interpreting common grounding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 05, pages 9081–9089, Hilton New York Midtown, New York, USA, February.

Vinyals, Oriol and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Visser, Thomas, David Traum, David DeVault, and Rieks op den Akker. 2012. Toward a model for incremental grounding in spoken dialogue systems. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents*, Santa Cruz, CA, USA, September.

Visser, Thomas, David Traum, David DeVault, and Rieks op den Akker. 2014. A model for incremental grounding in spoken dialogue systems. *Journal on Multimodal User Interfaces*, 8(1):61–73.

Walton, Douglas and David M. Godden. 2006. The impact of argumentation on artificial intelligence. *Considering Pragma-Dialectics: A Festschrift for Frans H. Van Eemeren on the Occasion of his 60th Birthday*, pages 287–299.

Walton, Douglas and Erik C.W. Krabbe. 1995. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press.

Walton, Douglas N. 1984. *Logical Dialogue-Games*. University Press of America, Lanham, Maryland.

Wang, Zhiyang, Jina Lee, and Stacy Marsella. 2013. Multi-party, multi-role comprehensive listening behavior. *Autonomous Agents and Multi-Agent Systems*, 27(2):218–234.

Winston, Lauren and Brian Magerko. 2017. Turn-taking with improvisational co-creative agents. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 13, 1, Snowbird, Utah, USA, October.

Young, Steve, Milica Gašić, Simon Keizer, François Mairesse, Jost Schatzmann, Blaise Thomson, and Kai Yu. 2010. The hidden information state model: A practical framework for pomdp-based spoken dialogue management. *Computer Speech & Language*, 24(2):150–174.

Yuan, Tangming, David Moore, and Alec Grierson. 2004. Human-computer debate, a computational dialectics approach. *unpublished doctoral dissertation, Leeds Metropolitan University*.

Zou, Yiqian. 2020. An experimental evaluation of grounding strategies for conversational agents. *Master Thesis. Department of Philosphy, Linguistics and Theory of Science. University of Gothenburg*.