Available online at www.sciencedirect.com
ScienceDirect
Computer Speech & Language 00 (2017) 1–16
www.elsevier.com/locate/csl

Modeling the clarification potential of instructions: Predicting clarification requests and other reactions

Luciana Benotti[a,*], Patrick Blackburn[b]

[a] *Logics, Interaction and Intelligent Systems Group, FAMAF, Universidad Nacional de Córdoba (UNC), Córdoba, Argentina*
[b] *Philosophy and Science Studies Section, IKH, Roskilde University (RUC), Roskilde, Denmark*

Received 17 April 2016; Accepted 20 January 2017

Abstract

We hypothesize that conversational implicatures are a rich source of clarification requests, and in this paper we do two things. First, we motivate the hypothesis in theoretical, practical and empirical terms and formulate it as a concrete clarification potential principle: *implicatures may become explicit as fourth-level clarification requests*. Second, we present a framework for generating the clarification potential of an instruction by inferring its conversational implicatures with respect to a particular context. We evaluate the framework and illustrate its performance using a human−human corpus of situated conversations. Much of the inference required can be handled using classical planning, though as we shall note, other forms of means-ends analysis are also required. Our framework leads us to view discourse structure as emerging via opportunistic responses to task structure.

Keywords: Clarification requests; Level-sensitive Gabsdil test; Conversational implicatures; Dialogue systems; Classical planning; Micro-planning; Negotiability; Tacit acts; Explicatures; Emergent structure; Opportunistic theories of communication

1. Introduction

We cannot fully understand the meaning of language in conversation without understanding the mechanisms that make conversation such a robust process. Nor should these mechanisms be seen as peripheral; arguably they are central to an adequate theory of meaning:

> *The adequacy of a semantic theory involves the ability to characterize for any utterance type the* ***contextual update*** *that emerges in the aftermath of successful exchange and the range of* ***possible clarification requests*** *otherwise – this is, arguably, the early 21st century analogue of truth conditions (Ginzburg, 2012, p. 8).*

That is, clarification requests are not a necessary evil but an intrinsic mechanism of language. Interpreting an utterance centrally involves characterizing its range of possible clarification requests, its *clarification potential* as we shall call it. Now, dialogue system designers have already realized the practical interest of clarification requests: see Gabsdil (2003), Purver (2004), Rodríguez and Schlangen (2004), Skantze (2007), Rieser and Lemon (2010),

[*] Corresponding author.
E-mail addresses: luciana.benotti@gmail.com, benotti@famaf.unc.edu.ar (L. Benotti), patrick.rowan.blackburn@gmail.com (P. Blackburn).

http://dx.doi.org/10.1016/j.csl.2017.01.008
0885-2308/ 2017 Elsevier Ltd. All rights reserved.

Stoyanchev et al. (2013). Moreover, in sociolinguistics and discourse analysis, where clarifications are known as *repairs*, they have been a central theme for more than three decades now; see Schegloff (1987) as a representative example. However, the theoretical scope of the phenomena and its wider implications for a theory of meaning are still being delineated; our main goal in this paper is to contribute to this discussion.

In this paper, we shall model the clarification potential of a single utterance type: instructions in task-oriented interactive settings. The following exchange illustrates the interactions we target:

Ann(1): Turn it on.
Bill(2): By pushing the red button?
Ann(3): Yeah.
Adapted from (Rodríguez and Schlangen, 2004, p. 102).

To spell this out a little, in order to carry out Ann’s request (turning something on) it is necessary to push the red button. By uttering (1), Ann has conveyed (by exploiting contextual knowledge of the task domain) that Bill should carry out a “push the red button” action although she did not say this explicitly. Bill might have known what was required and pushed the red button without further ado – but, for some reason, he chose instead to check with Ann that this was the required action.

Roughly speaking, our inference framework takes as input sentences like (1) and explains how (2) can be generated: it indicates what kinds of knowledge need to be represented and what kinds of inferences are involved in the process of generating utterances like (2). That is, it explains why this example constitutes a coherent dialogue by saying how the clarification is *relevant* to the instruction. Our framework makes explicit the relations between the instruction, its clarification and the context of the conversation. We do so by linking clarification with a central notion from pragmatics, namely the Gricean notion of conversational implicature.

As we discuss in Section 2, conversational implicatures are *negotiable*. And dialogue provides an intrinsic mechanism for carrying out negotiations of meaning: clarification requests. We hypothesize that conversational implicatures are a rich source of clarification requests: clarification requests make explicit what is *tacitly* conveyed by implicatures. In Sections 3 and 4 we present a framework for calculating the clarification potential of an instruction by inferring its contextualized conversational implicatures: each instruction is rooted in its context through utterance level *micro-planning*. The core inference method we use for micro-planning is classical AI planning, though other forms of means-ends analysis (such as inferring the next relevant action by computing affordabilities) are also required. In Section 5 we empirically evaluate the predictions of our inference framework, and in Section 6 we discuss the picture our framework gives rise to: discourse structure emerges as an opportunistic response to task structure. Section 7 concludes.

2. Definitions and motivations

In this section, we motivate our framework from the *practical perspective* of dialogue system designers, from the *theoretical perspective* of pragmatics, and from the *empirical perspective* of a human–human corpus. We first review a method of identifying clarification requests proposed in the dialogue system literature. Our review makes clear the necessity of further refinement, and we sketch what is required. We then discuss motivations from the perspective of pragmatics; in particular, we introduce the central notions of conversational implicatures and their negotiability. We view conversational implicatures as key to defining the *clarification potential* of an utterance. Lastly, we empirically motivate our work by presenting a corpus of task-oriented conversations.

2.1. Practical: defining clarification requests

Giving a precise definition of a clarification request (CR) is a difficult task. For a start, one might think that CRs are realized as questions; however corpus studies indicate that the most frequent realization of CRs is the declarative form (see Purver, 2004 for discussion). Indeed, although the form (including intonation pattern) of a CR exhibits some correlations with the CR function (Rodríguez and Schlangen, 2004), form is not generally a reliable indicator of the role the CR is playing.

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008

Gabsdil (2003) proposes a simple and elegant test for identifying CRs. Gabsdil's test says that CRs (as opposed to other kinds of contributions in dialogue) cannot be preceded by explicit acknowledgments. Consider the following example:

Lara: There's only two people in the class.
(a) Matthew: Two people?
(b) (??) Matthew: Ok. Two people?
Adapted from (Purver et al., 2003, p. 241).

Gabsdil argues that (a) in the example above *is* a CR because (b) is odd (odd turns are marked with (??) in the examples). In (b), Matthew first acknowledges Lara's turn and then indicates that her turn contains information that he finds controversial.[1]

On the other hand, (b) in the example below is fine and hence (a) *is not* a CR: the lieutenant acknowledges the sergeant's turn and then moves on to address what has become the most pressing topic in the conversation:

Sergeant: There was an accident sir
(a) Lieutenant: Who is hurt?
(b) Lieutenant: Ok. Who is hurt?
Adapted from (Traum, 2003, p. 391).

Gabsdil's original test incorrectly discards cases that clearly *are* CRs, such as the example (presented by Gabsdil himself as a CR):

G: I want you to go up the left hand side of it towards the green bay and make it a slightly diagonal line, towards, sloping to the right.
F: Ok. So you want me to go above the carpenter?
Adapted from (Gabsdil, 2003, p. 30).

The problem is that the level of evidence contributed by F's acknowledgment is ambiguous. For instance, the Ok could (conceivably) mean:

- *Ok, so you want to talk to me* (level 0, the level of attention).
- *Ok, I heard you* (level 1, the level of perception).
- *Ok, I understand all the words and I identified all the referents* (level 3, the level of recognition).
- *Ok, I did it* (level 4, the highest level, the level of uptake).

Thus we propose modifying Gabsdil's test to make it level-sensitive. In particular, in order to signal that all the levels have been successful and that no CR related to any of them is expected, the simple acknowledgment needs to be replaced by clear evidence of uptake (level 4). This works for Gabsdil's example:

G: I want you to go up the left hand side of it towards the green bay and make it a slightly diagonal line, towards, sloping to the right.
(??) F: Ok, I did it. So you want me to go above the carpenter?

In this case, *So you want me to go above the carpenter?* is either weird or far more likely to be interpreted as a question about an action that comes *after* having successfully followed G's instruction (that is, as a contribution that is not a CR). Which of these two alternatives is actually the case would be determined by further specification of the dialogue task.

More generally, if the addressee wants to uptake the speaker proposal then he or she has two options: either to signal uptake (and then, by downward closure, the speaker knows that all lower levels succeeded) or to explicitly

[1] This could be a felicitous move, but this would require a very marked intonation or a long pause which would induce some kind of "backtracking" effect.

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008

indicate the problem using a CR (at any level). Table 1 illustrates, for each level, possible CRs. The terminology used is adapted from Rodríguez and Schlangen (2004).

This approach to CR identification and classification is useful not only for instructions but also for other types of utterances. The following is an extension of Grice's classic implicature example (physical actions are between square brackets):

A: I am out of petrol.
B: There is a garage around the corner.
A: [A goes to the garage and then meets B again]
(??) A: Ok, I got petrol at the garage. And you think it's open?
Adapted from (Grice, 1975, p. 311).

After acknowledging a contribution at level 4 (which A's *Ok, I got petrol at the garage* clearly does) it is really hard to go on and ask a CR about that contribution (A's *And you think it's open?* is a bizarre follow-up). This is useful – the catch is that defining what is evidence in level 4 is not trivial; it depends on the *clarification potential* of the source contribution.

2.2. Theoretical: defining clarification potential

Modeling how listeners draw inferences from what they hear is a basic problem for theories of natural language understanding. An important part of the information an utterance conveys is inferred in context by taking into account the goal-oriented nature of conversation. We again illustrate with Grice's example:

A: I am out of petrol.
B: There is a garage around the corner.
⇝ B thinks that the garage is open.
(Grice, 1975, p. 311).

B's answer *conversationally implicates* (⇝) information that is relevant to A. In Grice's terms, B made a relevance implicature: he would be flouting the conversational maxim of relevance unless he believes that it is possible that the garage is open. A conversational implicature (CI) is different from an entailment in that it is *cancelable* without contradiction. B can append material that is inconsistent with the CI – *but I don't know whether it's open*. Since the CI can be canceled, B knows that it does not necessarily hold and thus both B or A are able to *reinforce* or *clarify* it without any sense of anomaly. That is, conversational implicatures are *negotiable*. For more on the negotiability of implicature see Benotti (2010) and Benotti and Blackburn (2010, 2014).

It is often controversial whether something is actually a CI or not (people have different intuitions, which is not surprising given that people have different background assumptions). The problem with CIs is that, by definition, they are not explicit: they provide *tacit* meanings. And this leads us to a central theme of the paper: *in dialogue, CRs provide good evidence of the implicatures that have been made, because level 4 CRs make implicatures explicit*. Take for example the clarification request which can naturally follow Grice's original example:

A: and you think it's open?

Table 1
Level-sensitive CR classification schema by Rodríguez and Schlangen (2004).

Num	Level	Kind of problem	Example
4	Uptake	Obstacle for carrying out the proposal	By pressing the red button? And you think it's open?
3	Recognition	Lexical problem	What's a "rebreather"?
		Reference problem	Which button?
2	Perception	Acoustic problem	What did you say?
1	Attention	Establish contact	Are you talking to me?

B will have to answer and support the implicature – for instance with *yes, it's open till midnight* – if he wants to get it added to the common ground. Otherwise, if he did not mean it, he can well reject it without contradiction with *well, you have a point there, they might have closed.*

To sum up: our hypothesis is that CIs are a rich source of CRs. And our method for generating the *clarification potential* of an utterance will be to infer (some of) the CIs of that utterance with respect to a particular context. We will verify the clarification potential thus generated by comparing it with the actual CRs in corpora.

In order to make this hypothesis more concrete we reformulate it as the following principle:

Clarification Potential Principle (CPP): implicatures become explicit as fourth level clarification requests when they cannot be grounded in the context and task in which the conversation is situated.

2.3. Empirical: instructions situated in conversation

The SCARE corpus described in Stoia et al. (2008) consists of fifteen spontaneous English dialogues associated with an instruction giving task.[2] We selected this corpus because the associated task is quite complex, taking on average 11 min to complete. Moreover the task is multimodal and is situated in a game; indeed, the conversations cannot be completely understood without watching the accompanying videos. The language used is deeply grounded in the game world the dialogue participants share.

The corpus was collected using the Quake environment, a first-person virtual reality game. The task consists of a direction giver (DG) instructing a direction follower (DF) on how to complete several tasks in a simulated game world. The corpus contains the collected audio and video, as well as word-aligned transcriptions.

The DF had no prior knowledge of the world map or tasks and relied on his partner, the DG, to guide him on completing the tasks. The DG had a map of the world and a list of tasks to complete. The partners spoke to each other through headset microphones; they could not see each other. As the participants collaborated on the tasks, the DG had instant feedback of the DF's location in the simulated world, because the game engine displayed the DF's first person view of the world on both the DG's and DF's computer monitors.

We analyzed the 15 transcripts that constitute the SCARE corpus while watching the associated videos to get familiar with the experiment and evaluate its suitability for our purposes. We then randomly selected one dialogue; its transcript contained 449 turns and its video lasted 9 min and 12 s. In this transcript, 291 turns (65%) were uttered by the DG and 158 (35%) by the DF. 73% of the turns uttered by the DF included positive acknowledgments – *yeah, mhm, alright, done* – and descriptions of the visible area – *there are two cabinets here* – and 27% constituted clarifications according to the level-sensitive version of the Gabsdil test we discussed in Section 2.1.

We classified the clarifications according to the levels of communication (see Table 1): 65% belong to the level 4, and 31% belonged to level 3 (most of them related to reference resolution). Only 4% of the CRs were acoustic (level 2) since the channel used was very reliable. No CRs at level 1 were found.

Below we include an extended example, extracted from the SCARE corpus, of the phenomena that we shall model. Between square brackets we indicate forms of non-linguistic communication:

DG(1): we have to put it in cabinet nine [pause]
DF(2): yeah [pause] they're not numbered [laughs]
DG(3): [laughs] where is cabinet nine [pause]
DG(4): it's [pause] kinda like back where you started [pause] so
DF(5): ok [pause] so I have to go back through here?
DG(6): yeah
DF(7): and around the corner?
DG(8): right
DF(9): and then do I have to go back up the steps?
DG(10): yeah
DF(11): alright this is where we started

[2] The corpus is freely available for research at http://slate.cse.ohio-state.edu/quake-corpora/scare/.

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008

DG(12): ok [pause] so your left ca-[pause] the left one
DF(13): so how do I open it?
DF(14): one of the buttons?
DG(15): yeah, it's the left one
DF(16): makes sense
DF(17): alright so we put it in cabinet nine

Of the 17 turns, 9 were uttered by the DF and 8 by the DG. Of the 9 turns by the DF, 6 of them are CRs in level 4. Turn (2) is a CR of instruction (1). Turns (5), (7) and (9) are CRs of instruction (4). Utterance (11) shows evidence of uptake of instruction (4) so this instruction cannot be further clarified following the level sensitive version of Gabsdil's test. Turns (13) and (14) are CRs of utterance (12). The evidence of uptake in level 4 of instruction (12) is completed by a physical action of the DF in the game world: opening the cabinet by pressing the left button while uttering (16). Finally, turn (17) together with the corresponding physical action are evidence of uptake in level 4 of instruction (1). We will take a closer look at this dialogue fragment in Sections 4 and 6.

This concludes our discussion of the linguistic background; we are now ready to present the inference framework we have devised for working with such data. In the following section we discuss the formal resources we use to represent the static and dynamic elements of the SCARE environment; once this is done, we turn to the inference framework itself.

3. Representing the context of the instructions

Our inference framework uses four information resources whose content depends on the information available to the participants of the situated interaction being modeled. The information resources described here are for modeling asymmetric interactions. That is, as in the SCARE corpus, one of the participants (the direction giver: DG) has complete information about how the world works and the task that has to be accomplished but cannot modify the world, while the other (the direction follower: DF) can modify the world but has only partial information about the world and no information about the task. In this section we describe each of these resources in turn and illustrate their content using the SCARE experimental setup.

3.1. The world model

The first information resource required is a world model, a knowledge base representing the game world's physical state: it contains complete and accurate information about the properties of individuals, for example, that something is a *button* or a *cabinet*. Relationships between individuals are also represented here, including the relationship between an object and its location. This knowledge base can be viewed as a *first-order model* (a *relational structure* Chang and Keisler, 2012) or − which pretty much amounts to the same thing − as a *relational database* (Abiteboul et al., 1995).

The content of the world model for the SCARE setup, such as the functions associated with the buttons in the game world and what the various cabinets contain, is automatically extracted from the Quake game environment. The content is modified during the interaction as the DF performs actions in the game world.

3.2. The interaction model

This knowledge base represents what the DF knows about the world in which the interaction is situated. The information the DF learns by observing the world while navigating through it is incrementally added to this knowledge base. Like the world model, this knowledge can be viewed as a first-order model; indeed, it is a submodel of the world model.

In the SCARE setup, the DF's initial instructions include no factual information. The only information that the DF receives are pictures of various objects in the world (so she can recognize them). Thus we can assume that the interaction model for the SCARE experiment starts out empty.

During the interaction, the interaction model is automatically updated with the information about the world that becomes visible to the dialogue participants as the DF moves inside the game world. It is also updated with the effects of the actions that the DF executes. For example, the DF learns that a particular button opens a given cabinet when she performs this action. A simplifying assumption of this model is that we assume that the DF's memory is perfect; we return to this point later as this is an assumption that has concrete consequences.

Since this interaction model starts empty, everything that is added here can be observed by both the DF and the DG, so we will assume that the information included here is mutually believed by them both.

3.3. The world actions

The framework also includes the definitions of the actions that can be executed in the world (physical actions such as *take* or *open*). Each action is formalized as a STRIPS-like operator (Fikes et al., 1972) detailing its arguments, preconditions and effects. The preconditions indicate the conditions that the world must satisfy if the action is to be executed; the effects determine how the action changes the world when it is executed.

In SCARE, these actions specify complete and accurate information about how the world behaves, and, together with the world model, they are assumed to represent what the DG knows about the world. The SCARE world action database contains a representation of the specification of the Quake controls received by both participants (for example, that pressing buttons in the game world can cause things to move). As with the world model, the world actions are automatically extracted from the Quake game environment.

3.4. The potential actions

The potential actions include a definition of how the DF conceptualizes the actions that she can perform on the world. This knowledge base may (but need not) coincide with the world actions. If it does not coincide it means that the DF has misconceptions about some actions.

In SCARE, the potential actions include the representation of actions that the DF tried during her pre-corpus collection training. In this training phase, the DF could move and act freely in a training room of the SCARE world. She could learn, for instance, that the effect of pressing a button can be to open a cabinet (if it was closed) or to close it (if it was open). As with the world actions, such knowledge can be represented as STRIPS-like operators.

We have now specified the different information elements that constitute the context of the SCARE experiments. These elements play a central role in the inference of the clarification potential: in order to infer the clarification potential of an instruction it is crucial to understand the facts and dynamics of these elements, and the knowledge that the dialogue participants have about them. Note that while the interaction model and the world model both change during an interaction, the potential actions and the world actions do not. These actions represent how the context can evolve from one state to the next; that is, they represent the causal links of the SCARE game world.

4. Predicting clarification requests and other reactions

We are now ready to present a framework that spells out how the Clarification Potential (CP) of an instruction is inferred and used in conversation; we introduce the framework incrementally in the subsections that follow. Following the Wilson and Sperber (2004) terminology for classifying conversational implicatures, we have classified potential clarifications into *implicated premises, implicated conclusions* and *explicatures*. We will treat each of these types in turn, illustrating our discussion with fragments of human−human dialogue drawn from the SCARE corpus. We begin by discussing the generic inference framework that underlies our work.

4.1. A generic inference framework for CP

Our inference framework links the CP of an instruction with its CRs via the following four steps:

Step 1: Pick an instruction from the corpus.
Step 2: Calculate the CP of the instruction using the interaction model and the potential actions.
Step 3: Predict the CRs using the CP just calculated together with the world model and the world actions.
Step 4: Compare the predictions with the corpus.

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008

In the remainder of this section we show how to flesh out this generic specification to handle the three kinds of CP we are targeting (implicated premises, implicated conclusions, and explicatures) and much of this discussion will revolve around Step 2 of the process, which is where we make use of AI planning. But first a more general remark about the role played by AI planning in our approach.

AI planning is central to our inference framework: we are interested in finding out which tasks can be handled using current off-the-shelf AI planners, and which tasks require something more (and what it is such tasks require). But we use planning differently from earlier work on utterance interpretation. Papers such as Lochbaum (1998), Carberry and Lambert (1999), Blaylock and Allen (2005) use shared-plan recognition. Now, in shared-plan recognition approaches, a whole dialogue is mapped to one shared plan, with each utterance adding a constraint to the partially filled-out plan. Our approach, on the other hand, uses planning at the utterance level rather than the dialogue level: *each instruction is interpreted as the template for a plan*. That is, we use AI-planning to perform what is called *micro-planning*. Micro-planning has previously been used for natural language generation by Koller and Stone (2007), Garoufi (2014). Here, building on ideas from Benotti (2010) and Benotti and Blackburn (2011), we use micro-planning for the interpretation of instructions instead. In our view, much important dialogue structure emerges from this interactive use of local planning, and we discuss this in Section 6.

4.2. An inference framework for implicated premises

Let us see how AI classical planners can be used to further specify Steps 2 and 3 of our generic inference framework for the case of implicated premises. We first spell out the formal framework and then illustrate it with fragments from the SCARE corpus.

Step 1: Pick an instruction from the corpus.

Step 2: When the DG gives an instruction, the DF has to interpret it in order to know what actions he has to perform. The interpretation consists in trying to construct a plan that links the current state of the interaction with the preconditions of the instruction. An action language used by AI planners such as STRIPS is used to specify the world action and potential action databases introduced in the previous section. Furthermore, the world model and the interaction model are relational structures that can be directly expressed as a set of literals, which is the format used to specify the initial state of a planning problem. Thus these information resources constitute almost everything that is needed in order to specify a complete *planning problem* as expected by current AI classical planners; indeed, the only element that is missing is the goal. With a set of action schemes (i.e. action operators), an initial state and a goal as input, a planner is able to return a sequence of actions (i.e. a plan) that, when executed in the initial state, achieves the goal (if an appropriate sequence exists).
In short, the specification of such planning problem is as follows:

- The preconditions of the instruction are the *goal*.
- The dialogue model is the *initial state*.
- The potential actions are the *action operators*.

Given this information, an off-the-shelf planner, such as FastForward designed by Hoffmann and Nebel (2001) will find a sequence of actions (that is, a plan), or say that no plan exists. *This sequence of actions is the set of implicated premises of the instruction.* It represents the modifications that the interaction model needs to undergo in order to be able to execute the instruction uttered. To put it another way: *the plan makes tacit knowledge explicit*.

Step 3 (if there is a plan): After inferring the plan, an attempt is made to execute the plan on the the world model using the world actions. Whenever the plan fails, the tacit act that failed is a predicted clarifications.

Step 3 (if there is no plan): All the preconditions that cannot be linked to the initial state by a plan are added to the set of predicted clarifications.

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008

Step 3 (if there is more than one plan): The plans are ranked in some way (e.g., by length) and the tacit acts of the higher ranked plan are part of the predicted clarifications of the instruction.

Step 4: Compare the predictions with the corpus.

To summarize: the *implicated premises* of an instruction is the sequence of actions that links the interaction model when the instruction was uttered with the preconditions of the instruction. This framework gives rise to three possible scenarios: there is a sequence which fails (failed plan), there is no sequence (no plan), there is more than one possible sequence (multiple plans). We will illustrate each of them in turn.

4.2.1. The plan fails

Let us illustrate the framework just introduced to analyze an implicated premise example from the SCARE corpus. In this example the participants are trying to move a picture from one wall to another. Let us go step by step:

Step 1: The instruction that is being interpreted is DG(1).

DG(1): well, put it on the opposite wall

Step 2: The preconditions of this instruction in the potential actions are to have the picture (in the DG's hands) and to be near the target wall. The inferred plan involves *picking up the picture* in order to achieve the precondition of *having the picture*, and *going to the wall* in order to achieve the precondition of *being near the wall*. That is, the actions *picking up the picture* and *going to the wall* are part of the CP of what the DG said.

Step 3: The plan inferred by the DF fails in the game world because the picture is *not takeable* and thus it cannot be picked up, resulting in a predicted clarification: *picking up the picture*. The correct plan to achieve (1) involves pressing a button instead of taking the picture.

Step 4: In the corpus, the predicted clarification *picking up the picture*, foreshadowed by (2) and (3), is finally made explicit by the CR in (4), as predicted by the model.

DF(2): ok control picks the [pause]
DF(3): control is supposed to pick things up and [pause]
DF(4): am I supposed to pick this thing?

4.2.2. There is no plan

In the case that no plan can be inferred, our framework predicts that the instruction preconditions for which no plan can be found will be part of the clarification potential of the instruction. Consider the example first introduced in Section 2.

Step 1: In the dialogue below, the DG utters the instruction (1) knowing that the DF will not be able to follow it; the DG is just thinking aloud.

DG(1): we have to put it in cabinet nine [pause]

Step 2: If taken seriously, this instruction would have the precondition *the reference to* cabinet nine*is resolved*. However this precondition cannot be achieved by any action, because the DF does not know the numbers of the cabinets. Hence a planner can find no plan for this planning problem.

Step 3: The framework then predicts a CR related to resolving the referent *cabinet nine*.

Step 4: Both participants know that the DF does not know the numbers, as only the DG can see the map. That is why the CR in (2) is received with laughs and the DG continues his loud thinking in (3) while looking at the map.

DF(2): yeah [pause] they're not numbered [laughs]
DG(3): [laughs] where is cabinet nine [pause]

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008

Our framework would not be able to produce a clarification move as precise as the DG did in (3) asking for the location of cabinet nine, because the planner will just say there is no plan for resolving the reference *cabinet nine*. However, using the information that the framework can output, namely *the referent cabinet nine cannot be resolved*, a more general clarification such as *which one is cabinet nine?* can be produced, asking about the identity and not the location of the referent.

4.2.3. *The plan is uncertain*

When more than one plan can be inferred for the given instruction, the alternative plans will be part of the clarification potential of the instruction. Why? Because the DF cannot be certain which plan the DG had in mind. We can see the following dialogue (which continues the fragment just given) as an instance of this case.

Step 1: Now, the DG refines the instruction given in (1) with the location of the target.

DG(4): it's [pause] kinda like back where you started [pause] so

Step 2: And the DF comes up with a plan that achieves the precondition of the instruction *put* uttered in (1) of being near the destination of the action (cabinet nine) namely: going back to *where you started*.

Step 3: Uttering the steps of the plan that were not made explicit by the instruction is a frequently used method for confirming the clarification potential of an instruction. The DF clarifies when he is not certain that the plan he found is exactly the one that the DG had in mind.

Step 4: The DF incrementally grounds the shortest plan he found by making it explicit in (5), (7), and (9) and waits for confirmation before executing each action. Finally the DF gives evidence of uptake in of instruction (4) in turn (11).

DF(5): ok [pause] so I have to go back through here?
DG(6): yeah
DF(7): and around the corner?
DG(8): right
DF(9): and then do I have to go back up the steps?
DG(10): yeah
DF(11): alright this is where we started

Thus the DF clarifies hypotheses when he is not certain that the plan found is exactly what the DG wants; for example, when there is more than one possible plan. However there may be other sources of uncertainty; for example, because the DF's memory is imperfect. We discuss such cases in Section 5. The rest of the running example involves interactions between different kinds of conversational implicatures and is analyzed in Section 6.

4.3. *An inference framework for implicated conclusions*

Not all clarifications of an instruction correspond to implicated premises, nor can they all be inferred using classical AI planning. Consider the following example:

DG(1): now, on the wall on the right turn and face that
DG(2): press the button on the left
DF(3): [presses the button and a cabinet opens]
DF(4): put it in this cabinet?
DG(5): put it in that cabinet, yeah

The question in (4) is not making explicit an implicated premise. Implicated premises are *necessary* in order to execute the instruction, implicated conclusions are not. Rather, implicated conclusions are *normally* drawn from the instruction and the context. Below, we propose a practical inference task (distinct from classical AI planning) to handle these cases.

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008

Step 1: In turns (1) and (2), the DG told the DF to press a button with no further explanation. As a result of pressing the button, a cabinet opened in (3).

Step 2: The inference of implicated conclusions can be defined intuitively as a practical inference task which involves finding the set of *next relevant actions*. The input of this means-ends task is different from that of a planning problem. It too has an initial state, and a set of possible actions, but it will also contain one observed action (in the example, action (3)) instead of the goal. Inferring the *next relevant action* consists in comparing the affordabilities (i.e. the set of executable actions) of the initial state and the affordabilities of the state after the observed action was executed. The output of this inference task, the set of next relevant actions, are those actions that were enabled by the observed action (that is, they were not possible in the initial state). In the example, the next relevant action that is inferred using this method is *put the thing you are carrying in the cabinet that just opened*.

Step 3: As far as we have observed, in the SCARE corpus the DF never executes a next relevant act without clarifying it beforehand. Next relevant actions are possible follow ups − but they are not certain. The urge to clarify such cases in the SCARE corpus is probably a result of the experimental setup, which lowered dialogue participants (DPs) scores if they perform wrong actions.

Step 4: In the example above, the next relevant action that will be inferred is “put the thing you are carrying in the cabinet that just opened”, just what the DF verbalizes in (4). In (5) the DG confirms this hypothesis.

4.4. An inference framework for explicatures

In the SCARE corpus, we encountered clarification requests in level four that are neither implicated premises nor implicated conclusions. They correspond to what has been called *explicatures* by Wilson and Sperber (2004).

For instance, in the following exchange the DG says *take the stairs* but the DF does not know whether to go downstairs or upstairs:

DG(1): there should be some stairs [pause] take the stairs [pause]
DF(2): up? [pause]
DG(3): yeah

Thus there are the two possible values that the missing parameter of this action can take (up or down) and the DF clarifies in (2) which was intended. In our framework we model explicatures as missing parameters of task actions. They are clarified when they cannot be inferred from context.

There is evidence in the corpus that the DF expects the DG not to provide parameters of actions that can be inferred from context. For instance, in (1) in the following dialogue, the DG specifies which way to leave the room. However, this is the only exit of the room in which the DF is currently located and the DF makes this explicit in (2):

DG(1): Let’s leave the way [pause] we came
DF(2): that’s the only way

It was suggested to the first author[3] that we could have used a bigger size for the unit of context update (that is, actions) and then the example about *putting the picture on the opposite wall* from Section 4.2 could also be treated as a missing parameter. On this view, the semantic content of the utterance *putting the picture on the opposite wall* has a missing *manner* parameter that can be filled by *by picking it up*, and we would avoid all the “heavy reasoning needed by planning”. Now, this certainly could be done in the picture example, but we don’t think that the heavy reasoning can be avoided for long. If we increased the size of our unit of context update every time we find a CR that can be resolved by supplying an extra argument, we could, in principle, end up with an infinite number of additional arguments. For example, the SCARE corpus contains a dialogue (reproduced in Section 2) where it takes 17 turns to finally ground the instruction *put it in cabinet nine*. It seems plausible that could give ourselves a pragmatic analog

[3] P.c. with Jonathan Ginzburg.

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008

of the semantic problem that (Davidson, 1969) solved several decades ago when he proposed his event semantics. So we do not think that all CRs can be treated as explicatures.[4]

5. Empirical evaluation

We used the level-sensitive version of the Gabsdil test discussed in Section 2 to detect CRs in the SCARE corpus. We then applied the inference framework described in the previous section to classify those CRs into:

Implicated premises (58%).
Implicated conclusions (16%).
Explicatures (11%).

That is, items were classified by the inference task used to analyze them, and we were able to account for 85% of the CRs that appeared in the SCARE corpus in this way. In other words: the performance of Step 2 of our inference framework is 85% in the SCARE corpus. The percentage of each kind of CR is shown in Fig. 1. Most of them are implicated premises, thus most of them were analyzed using classical AI planning.

The CRs not covered by the classification (15% in the SCARE corpus) seem to be accounted for by memory lapses. That is, people do not completely remember (or trust) the instructions given for the experiments or what they (or their partner) said a few turns before, and they sometimes seek clarification because of this. We will see an example of this type shortly.

We also classified implicated premises according to *why* they had to be made explicit in a CR. We list the reasons here and display the percentages found on the corpus in the pie chart:

Wrong plan: the plan inferred is not executable (16%).
Not explainable: no plan can be inferred (12%).
Ambiguous: more than one plan can be inferred (32%).

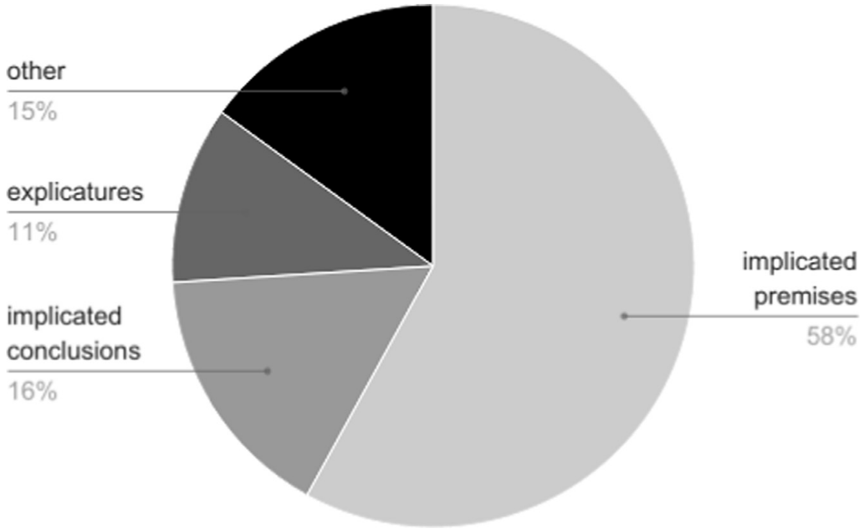

Fig. 1. Results of the evaluation of the coverage of the framework.

[4] Actually, we think that the other direction is more interesting: probably all explicatures can be treated as implicatures. This would require a smaller size for the context update than task-meaningful physical actions. It would imply that the interactions between the explicatures of an utterance would not need to be revised in the light of the implicatures; rather, all the not-explicit content would be developed in parallel embedded within the overall process of constructing a hypothesis about the speaker meaning. Whether such a framework can be properly formalized is a task for further work.

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008

Our framework is able to correctly predict when an implicated premise (inferred in the Step 2) will be realized as a CR in 60% of the cases. Most of these are cases when more than one plan was inferred. The remaining 40% of CRs are implicated premisses correctly inferred in Step 2, but Step 3 of our framework did *not* predict that they would be made explicit as a CR. In other words: conversational partners in the SCARE corpus clarify more often than our model predicts. Again, this is related to the DF memory not being perfect – the idealisation assumed in the the interaction model. Here is an example of a memory lapse induced CR:

DG(1): you've to [pause] like jump on it or something [pause]
DF(2): I don't know if I can jump

Here the DF does not remember that he can jump using the space bar (as stated in the instructions he received). The fact that people's memory is not reliable is intrinsic to communication, and modeling such limitations is one of the many challenges that a complete theory of communication will have to cope with.

6. The emergence of dialogue structure

Our inference framework is based on micro-planning: that is, it uses AI planning (and other inference techniques) *locally* – at the utterance level rather than the dialogue level. But micro-planning give rise to interesting interactional possibilities, and from these interactions dialogue structure may emerge in ways that are only visible in retrospect. We begin by finishing the analysis of the long example introduced at the end of Section 2 and partially analyzed in Section 4. Here is the initial instruction and the last part of the dialogue again:

DG(1): we have to put it in cabinet nine [pause]
...
DG(12): ok [pause] so your left ca-[pause] the left one
DF(13): so how do I open it?
DF(14): one of the buttons?
DG(15): yeah, it's the left one
DF(16): makes sense
DF(17): alright so we put it in cabinet nine

In (12), the target cabinet is identified but the DF is not able to find a plan that achieves another precondition of the instruction *put* uttered in (1), namely that the destination container is opened, so he directly produces a CR about the precondition in (13). However, the DF does not stop here and wait for an answer – instead he continues with (14). That is, the plan failure prompts the DF to continue the conversation using the partial information at his disposal.[5] This is an example of how an ambiguous plan case (*going back to the starting spot*) can interact with a no plan case (*opening the cabinet*).

Uncertain plans and failed plans can also interact. In the following example, the DF comes up with two plans that are supposed to achieve the instruction given in (1) and (2). One plan involves *pressing control* and the other sequence involves *jumping on the object*. Now, the DF learned (in the pre-corpus collection training phase) to pick up objects by pressing Ctrl, so he silently tries this plan first and then verbalizes in (3) the second plan, his dispreferred plan:

DG(1): we wanna move those three boxes
DG(2): so they're all three on the left side of the table
DF(3): ok may be I try jumping in and up there
DG(4): I'm not sure [pause] uh [pause]
DG(5): may be you can just press control [pause]
DF(6): I tried that and no luck

[5] In this situation a classical planner will just say "there is no plan" – that is, off-the-shelf planning technology can currently generate (13) but not (14). But although (14) cannot be obtained by a *classical* AI planner, new-generation *non-classical* planners that find plans when information is incomplete (Gaschler et al., 2015) may soon be able to model the DF's behavior here.

The DG does not know that the DF tried the shortest plan and failed, so he suggests it explicitly in (5).

These examples illustrate how the dialogue structure starts to emerge from the task structure. And we believe that dialogue structure really is an emergent property. It is impossible to specify in advance what actions each participant is to take in a long conversation. Conversations are created piece by piece as the participants negotiate purposes and then fulfill them; following Clark (1996), we call this the *opportunistic* view of conversation. However, conversations often look planed and goal-oriented *in retrospect*. Viewed as a whole, a conversation consists of a hierarchy of parts: conversation, sections, adjacency pairs, and turns. Where does the structure come from? We believe that the structure of the task in which the conversation is embedded has a strong impact on the structure of the conversation that emerges. What is the status of this structure? According to traditional plan recognition based models of dialogue (such as those described by Lochbaum, 1998; Carberry and Lambert, 1999; Blaylock and Allen, 2005) it reflects a plan that the DF and DG agree upon in order to reach their goals. In the opportunistic view, on the other hand, much of the structure emerges as the DF and DG do what they need to do in order to deal with the more local projects that arise in course of the conversation. That is: dialogue structure is a trace of the opportunities *taken*, not of the opportunities *considered*.

As the conversation above unfolded, it could have taken very different directions depending on what the DPs did. It is easy to see this in the SCARE corpus: all the participants had to perform the same task, yet the resulting interactions can be quite different. For instance, two other DPs performed the "box moving" task far more efficiently as follows:

DG(1): what we need to do is to move these boxes by pressing [pause]
DG(2): turn around [pause]
DF(3): [turns around]
DG(4): it's gonna be one of these buttons again [pause]
DG(5): and it's the one on the left
DF(6): ok [presses the button]

Roles can switch in conversations, and the DG can take advantage of the clarification potential of his own utterances. This is precisely what the dialogue just given illustrates. The DG gives the instruction to *move these boxes* in (1) and knows that the plan to achieve it is to *turn around*, and *look at the buttons*, and *press the left one*. So he uses these CIs to further specify this instruction in (2)–(4).

This type of dialogue structure has been characterized by saying that the DG is instructing in a top-down (or pre-order) fashion, first verbalizing a higher action in the hierarchy and then verbalizing the sub-actions; see Bard et al. (2008), Foster et al. (2009). However, under such a view, it is not so easy to explain how roles can switch and, more importantly, why some steps are omitted; that is, left *tacit*. For instance, in the DG instructions just given, the sensing action of looking at the buttons is not made explicit. Also, if the DG had *not* taken all the initiative in this sub-dialogue, the turns could also have been taken by the DF. This option is actually illustrated by yet another exchange from the SCARE corpus involving a third pair of DPs:

DG(1): you're gonna wanna move the boxes so you see now there's like two on the left and one on the right
DF(2): so let me guess [pause] like the picture [pause] the buttons move them
DG(3): aha that's true so you wanna turn around so you're facing the buttons
DF(4): [turns around]
DG(5): and you wanna push the button that's on the left
DF(6): ok [presses the button]

In this exchange it is the DF and not the DG the one who first makes explicit the need for button pressing. And, in our view, there is not a big difference between the two dialogues just given. Indeed, we selected the examples so that the parallelism is clear: the utterances can be mapped one by one (though they are not exactly in the same order).[6] Certainly this is not necessary: the utterances can be broken down in many different ways in real

[6] In passing: you may have noticed that the utterances in the last dialogue are longer and more articulate than in the previous two: these last two DPs were girls, the previous two pairs were guys. The gender of the DPs plays an interesting role in the SCARE corpus, one we are currently investigating.

conversation. However the two examples vividly make the point that there is a guiding structure behind these conversations, namely the task structure. The task structure opens up opportunities that either of the DPs can choose make use of or not. When and how they adopt these opportunities shapes the emergent structure of the dialogue.

7. Conclusions

Conversational implicatures are negotiable; this is the characteristic that distinguishes them from other kinds of inferences (such as entailments). Human−human dialogues use a sophisticated mechanism for carrying out negotiations of meaning, namely clarification requests. Implicature and clarification requests seem to fit well together, and to investigate their interaction, we reviewed theoretical work from pragmatics, practical work from the dialogue system community and empirical evidence from spontaneous dialogues situated in an instruction giving task. This led us to hypothesize that *implicatures become explicit as fourth level clarification requests*. We then presented a framework in which (part of) the clarification potential of an instruction was generated by inferring its conversational implicatures. We believe that this is a step towards defining a clear functional criteria for identifying and classifying clarification requests. We also believe that it supports a view of dialogue structure emerging as an opportunist response to task structure.

But more remains to be done. The empirical results we presented here are suggestive but preliminary; we are currently in the process of evaluating their reliability by measuring inter-annotator agreement. We are also interested in the role of forms of means-ends reasoning other than classical planning, a topic only touched on in this paper, and in which much fundamental work (such as analysis of computational complexity) remains to be done. Finally, we are considering automatically generating the world actions in domains with more uncertainty than that of the SCARE game setup. Currently, there are two main techniques for doing this: learning planning operators by reading text and improving them with reinforcement learning as proposed by Branavan et al. (2012), or learning planning operators from noisy sensory data like that obtained by robots acting in the real world (Zhuo and Kambhampati, 2013).

But while much remains to be done, we believe that the interplay between conversational implicatures and clarification mechanisms will eventually play a role in the development of *opportunistic* theories of communication.

Acknowledgment

This work was partially supported by grants ANPCyT-PICT-2014-1833, CIN-CONICET-PDTS-2015, PID-MINCYT-CBA-2014 and grants by the Argentinean Ministry of Science, Technology and Productive Innovation.

References

Abiteboul, S., Hull, R., Vianu, V., 1995. Foundations of Databases: The Logical Level. Addison-Wesley Longman Publishing Co., Inc.

Bard, E.G., Hill, R., Foster, M.E., 2008. What tunes accessibility of referring expressions in task-related dialogue? In: Proceedings of the Thirtieth Annual Meeting of the Cognitive Science Society (CogSci 2008).

Benotti, L., 2010. Implicature as an Interactive Process. Université Henri Poincaré, INRIA Nancy Grand Est, France Ph.D. thesis. Supervised by P. Blackburn.

Benotti, L., Blackburn, P., 2010. Negotiating causal implicatures. In: Proceedings of the Eleventh Annual Meeting of the Special Interest Group on Discourse and Dialogue. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 67–70.

Benotti, L., Blackburn, P., 2011. Classical planning and causal implicatures. In: Beigl, M., Christiansen, H., Roth-Berghofer, T., Kofod-Petersen, A., Coventry, K., Schmidtke, H. (Eds.), Modeling and Using Context. Lecture Notes in Computer Science. vol. 6967, Springer Berlin Heidelberg, pp. 26–39.

Benotti, L., Blackburn, P., 2014. Context and implicature. In: Brézillon, P., Gonzalez, J. (Eds.), Context in Computing: A Cross-Disciplinary Approach for Modeling the Real World. Springer New York, New York, NY, pp. 419–436.

Blaylock, N., Allen, J., 2005. A collaborative problem-solving model of dialogue. In: Proceedings of the Sixth SIGdial Workshop on Discourse and Dialogue. Lisbon, Portugal, pp. 200–211.

Branavan, S.R.K., Kushman, N., Lei, T., Barzilay, R., 2012. Learning high-level planning from text. In: Proceedings of the Fiftieth Annual Meeting of the Association for Computational Linguistics: Long Papers, vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 126–135.

Carberry, S., Lambert, L., 1999. A process model for recognizing communicative acts and modeling negotiation subdialogues. Comput. Linguist. 25 (1), 1–53.

Chang, C.C., Keisler, H.J., 2012. Model Theory. third ed. Dover Publications Inc.

Clark, H., 1996. Using Language. Cambridge University Press, New York.

Davidson, D., 1969. The Logic of Decision and Action. University of Pittsburgh Press, Chapter The Logical Form of Action Sentences, pp. 3–17.

Fikes, R., Hart, P., Nilsson, N., 1972. Learning and executing generalized robot plans. Artif. Intell. 3, 251–288.

Foster, M.E., Giuliani, M., Isard, A., Matheson, C., Oberlander, J., Knoll, A., 2009. Evaluating description and reference strategies in a cooperative human-robot dialogue system. In: Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09).

Gabsdil, M., 2003. Clarification in spoken dialogue systems. In: Proceedings of the AAAI Spring Symposium. Workshop on Natural Language Generation in Spoken and Written Dialogue, pp. 28–35.

Garoufi, K., 2014. Planning-based models of natural language generation. Lang. Linguist. Compass 8 (1), 1–10.

Gaschler, A., Kessler, I., Petrick, R.P.A., Knoll, A., 2015. Extending the knowledge of volumes approach to robot task planning with efficient geometric predicates. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2015). Seattle, Washington, USA.

Ginzburg, J., 2012. The Interactive Stance: Meaning for Conversation. Oxford University Press.

Grice, P., 1975. Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), Syntax and Semantics: Speech Acts. 3, Academic Press, pp. 41–58.

Hoffmann, J., Nebel, B., 2001. The FF planning system: fast plan generation through heuristic search. JAIR 14, 253–302.

Koller, A., Stone, M., 2007. Sentence generation as a planning problem. In: Proceedings of the Forty-fifth Annual Meeting of the Association of Computational Linguistics. Association for Computational Linguistics, Prague, Czech Republic, pp. 336–343.

Lochbaum, K.E., 1998. A collaborative planning model of intentional structure. Comput. Linguist. 24 (4), 525–572.

Purver, M., 2004. The Theory and Use of Clarification Requests in Dialogue. Ph.D. thesis. King’s College, University of London.

Purver, M., Ginzburg, J., Healey, P., 2003. On the means for clarification in dialogue. Current and New Directions in Discourse and Dialogue. Kluwer Academic Publishers, pp. 235–255.

Rieser, V., Lemon, O., 2010. Learning human multimodal dialogue strategies. Nat. Lang. Eng. 16, 3–23.

Rodríguez, K., Schlangen, D., 2004. Form, intonation and function of clarification requests in german task oriented spoken dialogues. In: Proceedings of the 2004 Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL), pp. 101–108.

Schegloff, E., 1987. Some sources of misunderstanding in talk-in-interaction. Linguistics 8, 201–218.

Skantze, G., 2007. Ph.D. thesis Error Handling in Spoken Dialogue Systems. Ph.D. thesis. KTH − Royal Institute of Technology, Sweden.

Stoia, L., Shockley, D., Byron, D., Fosler-Lussier, E., 2008. SCARE: a situated corpus with annotated referring expressions. In: Proceedings of the 2008 International Conference on Language Resources and Evaluation (LREC).

Stoyanchev, S., Liu, A., Hirschberg, J., 2013. Modelling human clarification strategies. In: Proceedings of the 2013 ACL Special Interest Group Conference on Discourse and Dialogue (SIGDIAL). Association for Computational Linguistics, Metz, France, pp. 137–141.

Traum, D., 2003. Semantics and pragmatics of questions and answers for dialogue agents. In: Proceedings of the 2003 International Workshop on Computational Semantics (IWCS), pp. 380–394.

Wilson, D., Sperber, D., 2004. Relevance theory. In: Horn, L., Ward, G. (Eds.), The Handbook of Pragmatics. Blackwell, Oxford, pp. 607–632.

Zhuo, H.H., Kambhampati, S., 2013. Action-model acquisition from noisy plan traces. In: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence. AAAI Press, pp. 2444–2450.

Please cite this article as: L. Benotti, P. Blackburn, Modeling the clarification potential of instructions: Predicting clarification requests and other reactions, Computer Speech & Language (2017), http://dx.doi.org/10.1016/j.csl.2017.01.008