Phrasing Questions

Geert-Jan M. Kruijff
German Research Center
for Artificial Intelligence (DFKI GmbH)
Saarbrücken, Germany
gj@dfki.de

Michael Brenner
Institute for Computer Science
Albert-Ludwigs-Universität
Freiburg, Germany
brenner@informatik.uni-freiburg.de

Abstract

In a constructive learning setting, a robot builds up beliefs about the world by interacting – interacting with the world, and with other agents. Asking questions is key in such a setting. It provides a mechanism for interactively exploring possibilities, to extend and explain the robot's beliefs. The paper focuses on how to linguistically phrase questions in dialogue. How well the point of a question gets across depends on how it is put. It needs to be effective in making transparent the agent's intentions and beliefs behind raising the question, and in helping to scaffold the dialogue such that the desired answers can be obtained. The paper proposes an algorithm for deciding what to include in formulating a question. Its formulation is based on the idea of considering transparency and scaffolding as referential aspects of a question.

Introduction

Robots are slowly making their entry into "the real world." And it is slowly becoming an accepted fact of life that we cannot possibly provide such robots will all there is to know, out-of-the-box. So they need to learn. The point of socially guided (machine) learning (Thomaz 2006) is that some of that learning can be done effectively through social interaction with other agents in the environment.

This paper focuses on how a robot should phrase its questions, considering a social learning setting in which situated dialogue is the main interactive modality (Kruijff et al. 2006a; Jacobsson et al. 2007). The robot and a human use spoken dialogue to discuss different aspects of the environment. We consider learning to be driven by the robot's own, perceived learning needs. This requires dialogue to be mixed-initiative. Both the human and the robot can take the initiative in driving this "show-and-tell-then-ask" dialogue. Questions play a fundamental role in such dialogues. Assuming a robot has the ability to raise issues in need of clarification or learning for any modality, (e.g. (Kruijff, Brenner, and Hawes 2008)), the problem thus becomes how to properly *phrase* a question.

Typically, a question is represented as an abstraction over the argument of a predicate. For example, assuming $?x.P(x)$ to indicate that a question regards a parameter x of some predicate $P(x)$, a question about the color of a ball could be phrased as $?x.(ball(y) \wedge has\!-\!color(y,x))$. However, more aspects need to be taken into account, for a question to be posed in such a way that the addressee is likely to understand the question and provide a suitable answer (Ginzburg 1995b).

First of all, the phrasing needs to make *transparent* how a question arises from an agent's beliefs, what beliefs – and what gaps in an agent's beliefs – it refers to. It should make clear *what a question is about*. Furthermore, there is a reason behind raising the question. The agent has a specific goal, it intends to obtain a particular kind of answer. Not just any answer will do. Raising a question also needs to set up, *scaffold*, the right context for answering it. This is the *why* of a question, pointing to how the agent would like to see the question *resolved*.

An example in (Kruijff et al. 2006b; 2007b) provides an interesting illustration.[1] The robot is capable of figuring out when it might have mistakenly classified a particular passage in the environment as a door. At the point where it realizes this, it asks, "Is there a door here?" Unfortunately, the place where it asks this is not related to the location "here" refers to. To anyone but a developer-acting-as-user it is not transparent what the "here" means. This often leads to the user giving the wrong answer, namely "yes this room has a door" rather than, "no, there is no door between the trash bin and the table." The way the question was phrased lacked both in transparency (location reference) and in scaffolding (specific location, not the room as such).

The paper presents an approach to generating a content representation for a question. These representations reflect what is being asked after, in reference to beliefs (aboutness, transparency) and intentions (resolvedness, scaffolding). The approach explicitly regards transparency and scaffolding as *referential qualities* of a question. This way their referential nature in the larger dialogue- and situated context can be considered. Following out that idea, the approach bases its content determination algorithm on Dale & Reiter's incremental algorithm for generating referring expressions (Dale and Reiter 1995), in combination with algo-

[1] See also the video at the CoSy website's Explorer page, at http://cosy.dfki.de/www/media/explorer.y2.html.

rithms for referential context determination (Zender, Kruijff, and Kruijff-Korbayová 2009; Paraboni, van Deemter, and Masthoff 2007).

Central to the approach is establishing the information pertaining to the question. A description logic-like formalism is used to represent such information, as a conceptual structure in which propositions have ontological sorts and unique indices, and can be related through named relations. A question can then be represented as a structure in which we are querying one or more aspects of such a representation (Ginzburg 1995b; Kruijff, Brenner, and Hawes 2008). The formalism allows everything to be queried: relations, propositions, sorts. Around the formulation of a question we construct a nucleus, comprising the situation (the ”facts”) and the beliefs that have led up to the question, the question itself, and the goal content which would resolve the question. The question nucleus integrates Ginzburg's notions of aboutness, and (potential) resolvedness.

Based on the question nucleus, the algorithm starts by determining to what extend the different aspects are covered by the (dialogue) common ground between the robot and the human. For this, contextual references are resolved in a dialogue context model (Kruijff et al. 2007a), and it is established how these can be related to inferences over domain knowledge and instances (Kruijff et al. 2007b). The question nucleus is extended with these connections – or rather, with indications of the information structure or informativity of individual content – so that it includes an explicit notion of what is shared, and what is privately held information (cf. (Lochbaum, Grosz, and Sidner 1999; Grosz and Kraus 1999)).

The algorithm next decides what aspects of a question nucleus to include in the content for phrasing the question. For each aspect of the nucleus (facts, beliefs, question, goals) the algorithm uses the informativity of the aspect's content, in conjunction with similarly related but contrasting content in the dialogue context model, to determine whether to include it. Essentially, new or contrastive content will be considered, whereas salient ”old” information will not. The form in which the content will be included is determined by content-specific algorithms for generating referring expressions (e.g. (Kelleher and Kruijff 2006; Zender, Kruijff, and Kruijff-Korbayová 2009)). The decisions to include particular content can be weighted according to a comprehensibility ranking as e.g. in (Krahmer, van Erk, and Verleg 2003).

The contributions the approach aims for are, briefly, as follows. Purver and Ginzburg develop an account for generating questions in a dialogue context (Purver, Ginzburg, and Healey 2003; Purver 2004). Their focus was, however, on clarification for the purpose of dialogue grounding. A similar observation can be made for recent work in HRI (Li, Wrede, and Sagerer 2006), We are more interested in formulating questions regarding issues in building up situation awareness, including the acquisition of new ways of understanding situations (cf. also (Kruijff, Brenner, and Hawes 2008)). In issue-based (or information state-based) dialogue systems (Larsson 2002), the problem of how to phrase a question is greatly simplified because the task domain is fixed. There is little need for paying attention to transparency or scaffolding, as it can be assumed the user understands the task domain.

An overview of the paper is as follows.The paper starts with a discussion of basic issues in modeling questions and their semantics, based on (Ginzburg 1995b). Then the approach is presented. The approach starts from the assumption that a question is a dialogue, not just a single utterance. Discussed is how the content plan for such a question dialogue can be determined, providing definitions, representation, and algorithms. The paper ends with a discussion of how the approach could be integrated, evaluated, and points for further research.

Background

What is a question? Ginzburg (1995b) discusses a variety of linguistic approaches. All of them aim to provide an invariant characterization of the semantics of a question. Broadly, they have proposed the following aspects as crucial to that definition.

First, several approaches propose to see a question as an n-ary relation. The relation puts together the question with one or more contributions pertaining to answering it. The point here is to take into account the fact that a question can be discussed over several turns in a dialogue. Second, there is a sense of *aboutness* to a question. Each question can be associated with a collection of propositions, which are –intuitively– related to the question. And, finally, each question can be considered to be associated with a (possibly complex) proposition which provides an *exhaustive answer*. In other words, an exhaustive answer resolves the question.

Ginzburg suggests that all these aspects together make up a characterization of a question – not just one of them, as most approaches suggest. Furthermore, these aspects are to be understood as being *relative*. What a question is about, and how it can be resolved, should be understood relative to an agent's *goal* and *belief/knowledge state* (cf. also (Ginzburg 1995a)). The following example illustrates this.

(1) Context: a robot drives around campus, and is about to enter the DFKI building.
   a. Janitor: Do you know where you are?
      Robot: DFKI.
   b. Janitor believes the robot knows where it is.

(2) Context: a robot drives around the DFKI building, to get a cup of coffee.
   a. Janitor: Do you know where you are?
      Robot: DFKI.
   b. The janitor is not convinced the robot really knows where it is.

What counts as an answer to a question may thus vary across contexts. What a question is thus cannot be reduced to an analysis of just what counts as its answers. Instead, Ginzburg starts with setting up an ontology in which questions, propositions and facts are considered as equal citizens. This makes it possible to consider a question *in relation to* possible answers for it. The ontology is defined using situation theoretic constructs, which we will adopt throughout this paper. (All definitions as per (Ginzburg 1995a; 1995b).)

Definition 1 (SOA, Situation, Fact)**.** A SOA (State Of Affairs) describes possible ways an actual situation might be. SOAs are either *basic*, or built up from basic ones using algebraic operations. A *basic SOA* is an atomic possibility, written as $\langle R, f : i\rangle$ with R a relation, f a mapping assigning entities to the argument roles of R, and i is a polarity i.e. $i \in \{+, -\}$. A situation s supports the factuality of a SOA σ iff $s \models \sigma$. The SOA σ is then considered a *fact* in s. To enable complex SOAs, SOAs can be structured as a Heyting algebra under a partial order '$\rightarrow$', which is closed under arbitrary meets ($\bigwedge$) and joins ($\bigvee$). Situations and SOAs together form a SOA-algebra:

1. If $s \models \sigma$ and $\sigma \rightarrow \tau$ then $\sigma \models \tau$
2. $s \not\models 0$, $s \models 1$ (FALSE,TRUE)
3. If Σ is any finite set of SOAs, then $s \models \bigwedge \Sigma$ iff $s \models \sigma$ for each $\sigma \in \Sigma$
4. If Σ is any finite set of SOAs, then $s \models \bigvee \Sigma$ iff $s \models \sigma$ for at least one $\sigma \in \Sigma$

Finally, an application operator is defined, to allow for variable assignment (and reduction):
$\lambda x.\langle R, a : b, c : x : +\rangle[x \mapsto d] = \langle R, a : b, c : d : +\rangle$ □

Using Definition 1, we can now consider a proposition to be an assertion about the truth of a possibility relative to a situation.

Definition 2 (Proposition)**.** A proposition p is a relational entity, asserting a truth regarding a SOA τ in a particular situation s: $p = (s : \tau)$. A proposition $p = (s : \tau)$ is TRUE iff τ is a *fact* of s, denoted as $s \models \tau$. □

Before defining what a question is, the notions of *resolvedness* and *aboutness* need to be defined. Resolvedness, or rather the broader concept of *potentially resolving* a question, is defined as follows. The definition distinguishes whether a (possibly complex) fact resolves a question depending on whether the question is *polar*, asking for the truth of an assertion (e.g. "Is the ball red?"), or *factive*, asking after a value (e.g. "What color is the ball?").

Definition 3 (Resolvedness conditions)**.** A SOA τ *potentially resolves* a question q if either

1. τ positively-resolves q (for 'polarity p': any information that *entails* p; for a factive question: any information that entails that the extension of the queried predicate is non-empty)
2. τ negatively-resolves q (for 'polarity p': any information that *entails* $\neg p$; for a factive question: any information that entails that the extension of the queried predicate is empty)

□

We will leave the notion of *aboutness* for the moment. Essentially, Ginzburg (1995a; 1995b) defines this as a collection of SOAs which can be associated with the content of a question q, with a SOA being about q if it subsumes the fact that q is either positively or negatively resolved. (For subsumption, recall Definition 1.)

Ginzburg's definition of what a question is then works out as follows.

Definition 4 (Question)**.** A question is an entity $(s?\mu)$ constructed from a situation s and an n-ary abstract SOA $\mu = \lambda x_1, ..., x_n \sigma(x_1, ..., x_n)$ $(n \geq 0)$:

1. μ constitutes an underspecified SOA from which the class of SOAs that are *about* q can be characterized.
2. Those SOAs which are facts of s and informationally subsume a level determined by μ constitute a class of SOAs that *potentially* resolve q.

□

The definition includes references to the relational character of a question (the abstract), and the notions of aboutness (intuitively, the space within which we are looking for an answer) and of resolvedness (the space of possible answers we are looking for, one of which will -hopefully- establish itself as fact). Finally, we already indicated above that resolvedness is an agent-relative notion. Ginzburg suggests to do so using Definition 3 as follows.

Definition 5 (Agent-relative resolvedness)**.** A fact τ *resolves* a question $(s?\mu)$ relative to a mental situation ms iff

1. Semantic condition: τ is a fact of s that potentially resolves μ
2. Agent relativisation: $\tau \implies_{ms} Goal - content(ms)$, i.e. τ entails the goal represented in the mental situation ms relative to the inferential capabilities encoded in ms.

□

Approach

The previous section presented a formal (but relatively abstract) notion of what a question is. It made clear that a question is more than a predicate with an open variable, or (alternatively) just another way of characterizing a set of propositions that would serve as exhaustive answer. Instead, a question is a relational structure, tying into a larger context. For one, this "context" provides a set of beliefs (SOAs, in Ginzburg's terms), a background within which potential answers are sought. An agent's goals help motivate to focus which beliefs are associated with the question. Another point about this "context" is that a question isn't just a single utterance, or just forming a unit with an utterance that answers it. There is a dialogue context in which this question is phrased. The question itself, and whatever utterances contribute to help clarify, refine and answer that question, may (though need not) refer to content already established in that context.

Phrasing a question, in other words, means we need to provide the possibility for such contextual factors to influence how the content of a question is determined. Once the agent has determined that it needs to raise a question, and about what (e.g. cf. (Kruijff, Brenner, and Hawes 2008) for questions in situated forms of learning), it needs to establish how best to communicate the question. In this paper, we suggest to do this as follows. We will begin by further explication of the notion of question, using a structure we term the *question nucleus*. The question nucleus captures more explicitly the relation between beliefs and intentions that are active in a current context, and how they determine the space of possible answers (or complexes of those). Then, we sketch several algorithms. The first group of algorithms concern *context determination*. Intuitively, these algorithms determine what beliefs and potential answers form the relevant background for the question. The background specifies what can be assumed to be known, (and can thus be referred to or even silently assumed), both in terms of content and intentions in the the dialogue- and situated context. How a question is to be phrased relies on what it needs to explicate relative to that background, to effectively communicate it. This is then finally done by the *content determination* algorithm. The result of this algorithm is a logical form, expressed in a (decidable) description logic. The logical form specifies the core content for the question, which a content planner subsequently can turn into one or more fully-fledged utterances.

The following definition defines more precisely what we mean by a logical form, based on (Blackburn 2000; Baldridge and Kruijff 2002). We will use the same formalism to describe SOAs (cf. Definition 1).

Definition 6 (Logical forms)**.** A logical form is a formula ϕ built up using a sorted description logic. For a set of propositions $PROP = \{p, ...\}$, an inventory of ontological sorts $SORT = \{s, ...\}$, and a set of modal relations $MOD = \{R, ...\}$, $\phi = p \mid i : s \mid \psi \wedge \psi' \mid \langle R \rangle \psi \mid @_{i:s}\psi$. The construction $i : s$ identifies a nominal (or index) with ontological sort s. The at-operator construction $@_{i:s}\psi$ specifies that a formula ψ holds at a possible world uniquely referred to by i, and which has ontological sort s. □

A standard Kripke-style model-based semantics can be defined for this language (Blackburn 2000). Intuitively, this language makes it possible to build up relational structures, in which propositions can be assigned ontological sorts, and referred to by using i as indices. For example, $@_{b1:entity}(\mathbf{ball} \wedge \langle Property \rangle (c1 : color \wedge \mathbf{red})$ means we have a "ball" entity, which we can uniquely refer to as $b1$, and which has a (referable) color property. (An alternative, equal way of viewing this formula is as a conjunction of elementary predications: $@_{b1:entity}\mathbf{ball} \wedge @_{b1:entity}\langle Property \rangle c1 : color \wedge @_{c1:color}\mathbf{red}$.)

Question nucleus

We start by defining the notion of *question nucleus*. The function of a question nucleus is twofold. First, it should capture the question's background in terms of associated beliefs and intentions, and what space of expected answers these give rise to. An expected answer is naturally only as specific (or unspecific) as is inferable on the basis of what the agent knows.

Definition 7 (Expected answer)**.** An expected answer a for a question q is a proposition $a = (s : \tau)$, with τ potentially resolving q as per Definition 3. τ is a logical formula (Definition 6) which can be underspecified, both regarding the employed ontological sorts, and arguments. □

Effectively, assuming that the agent has a collection of ontologies which provide a subsumption structure ($a \sqsupset b$ meaning a subsumes b, i.e. b is more specific), an expected answer can be said to define a "level" of specifity (Definition 4) according to subsumption. Following up on the ball example, assume the agent has an ontology which defines $material - property \sqsupset \{color, shape\}$. An expected answer to a question, what particular shape the ball has, would take the form $@_{b1:entity}(\mathbf{ball} \wedge \langle Property \rangle (s1 : shape))$. All the proposition specifies is that there is an identifiable shape. If the question would be about any, or some unknown, property of the ball, an expected answer could be phrased as $@_{b1:entity}(\mathbf{ball} \wedge \langle Property \rangle (m1 : material - property))$. Using the available ontological structure, and relational structure between formulas, we can formulate expected answers at any level of specifity without requiring the agent to already know the answer (cf. also (Kruijff, Brenner, and Hawes 2008)).

Definition 8 (Question nucleus)**.** A *question nucleus* is a structure $qNucleus = \{r, BL, XP, AS\}$ with:

1. A referent r relative to which the question q (part of XP) is phrased.
2. BL (*Beliefs*) is a set of private and shared beliefs, about agent intentions and facts in the current context (cf. (Lochbaum, Grosz, and Sidner 1999; Grosz and Kraus 1999)).
3. XP (*Execution Plan*) is a continual plan with an execution record (Brenner and Nebel 2008) for resolving a question $q = (s?\mu)$.
4. AS (*Answer Structure*) is a finite $\sqsupset$-structure over propositions $p_1, ...$ which potentially resolve q, and which are implied by BL.

The beliefs BL specify what the agent knows about r, what the agent presumes to be shared knowledge about r, and what the agent presumes other agents could know about r. BL is based on the dialogue leading up to the question, any previous actions involving r, and a domain model of agent competences (Brenner and Kruijff-Korbayová 2008). XP makes explicit that phrasing a question constitutes a dialogue, with an associated plan for communicating the question and a record for how far the question has been fully answered. This record maintains which aspects (elementary predications) of the question are still open ("under discussion," similar to the Question-Under-Discussion construct of (Ginzburg 1995b)). The AS is a set of propositions, relating those propositions to the aspect(s) of the question they would potentially resolve (and thus to the execution record in XP). AS is based on propositions implied by BL (relative to r, q) and is $\sqsupset$-structured according to ontological structure. □

Contextually determining aboutness

Asking a question starts with the agent having determined what it is it needs to know about some referent r, e.g. an area in the environment, an object – or, more specifically, relations or properties. (To allow for group referents, we will consider r to be a *set*.) Next the question nucleus is built up, starting with the beliefs about the question, BL.

We adopt the approach to belief modeling described in (Brenner and Kruijff-Korbayová 2008). Beliefs are formulated as relational structures with *multi-valued state variables* (MVSVs). These state variables are used for several purposes. First, they can indicate domain values, as illustrated by the sorted indices in the examples above. The color $c1$ would be a Property-type state variable of the entity $b1$, and could take domain values in the range of that ontological sort. Important is that the absence of a value for an MVSV is interpreted as *ignorance*, not as falsehood: $@_{b1:entity}(\mathbf{ball} \wedge \langle Property \rangle (s1 : shape))$ means the agent does not know what shape the ball has, not that it has no shape (as per a closed-world assumption). In a similar way, state variables are used for expressing *private beliefs*, and mutual or *shared beliefs* (Lochbaum, Grosz, and Sidner 1999; Grosz and Kraus 1999). A private belief of agent a_1 about content ϕ is expressed as $(K\{a_1\}\phi)$ whereas a mutual belief, held by several agents, is expressed as $(K\{a_1, a_2, ...\}\phi)$. Secondly, MSVSs can be quantified over, for example using the ? to express a question: $?s1.@_{b1:entity}(\mathbf{ball} \wedge \langle Property \rangle (s1 : shape))$ represents a question regarding the shape of the referent $b1$.

As an agent perceives the environment, we assume it builds up beliefs about the instances it perceives, and what relations can be observed or inferred to hold between them. For example, see (Brenner et al. 2007) for a robot manipulating objects in a local visual scene, or (Kruijff et al. 2007b) for a robot exploring an indoor environment. Furthermore, we assume that the agent's planning domains include models of agent capabilities – what another agent is capable of doing, including talking (and answering questions!) about particular aspects of the environment (Brenner and Kruijff-Korbayová 2008). Finally, if the agent has been engaged in a dialogue with another agent, and discussed the referent-in-question r before, we assume that the (agreed-upon) content discussed so far constitutes shared beliefs, held by all agents involved.

Algorithm 1 : Determine(BL) (*sketch*)

Require: `BELs` is a set of private and mutual beliefs the agent holds, (including beliefs about capabilities); r is the referent (set) in question

```
BL = ∅
for b ∈ BELs do
   if b includes a MVSV m ∈ r then
      BL = BL ∪ b
   end if
end for

return BL
```

Algorithm 1 sketches the basis of the algorithm for establishing BL. Those beliefs are gathered which refer explicitly to the referent the question is about. Note that BL may end up being empty. This means that r has not been talked about, nor does the agent know whether another agent could actually offer it an answer to what it would like to know more about.

Contextually determining resolvedness

The beliefs BL about the referent in question r state what the agent already believes about r (privately, or shared), and what it believes about another agent's capabilities. Next, these beliefs need to be structured such that potentially resolving answers can be derived. We assume that we can make use of the ontological sorts, and the structuring over these sorts provided by domain ontologies, to organize beliefs. The organization we are after first of all relates a belief to a potentially resolving answer, by combining it (inferentially) with the ?-quantified, ontologically sorted MVSVs in the question to yields a partially or completely reduced logical form (Definition 1). Secondly, the organization relates beliefs by (sortal) subsumption over the potentially resolving answers they generate.

For example, consider a question about the color of a ball: $?c1.@_{b1:entity}(\mathbf{ball} \wedge \langle Property \rangle (c1 : color))$. Let us assume the robot holds several beliefs with regard to $b1$, and the variable $c1$. A robot learning more about visual properties of objects through interaction with a human tutor (Jacobsson et al. 2007) typically holds at least beliefs about what the tutor is capable of telling it. Thus, assume the robot believes the tutor can tell it about material properties, colors, and shapes. Using `tell-val` (*tell value* action) we can model these beliefs as $(K\ \{a_1\}\ tell-val(a_2, m : material-property)$, $(K\ \{a_1\}\ tell-val(a_2, c : color)$. The variables m, b are existentially bound in these beliefs. Using the inference that $material - property \ \sqsupset \ color$ and introducing bound variables m', c' for m and c respectively, the beliefs can be combined with the question to yield the potentially resolving propositions $c' : color, m' : material - property$. Furthermore, subsumption yields $m' : material - property \sqsupset c' : color$. Thus, by combining the beliefs with what the agent already knows, it can expect to know something it doesn't yet know by asking a question. And by making use of the way its knowledge is ontologically structured, it can determine how precise that answer is likely to be.

Algorithm 2 provides a first sketch of the algorithm for establishing AS. (In the current version, propositional content and additional relational structure pertaining to m in the context of b is not yet included into AS.)

Content determination

Finally, once the beliefs about q and the potentially resolving answers for q have been established, we can turn to determining the exact content for communicating q. The purpose of content determination is to establish what, how much, should be communicated for the agent to get an appropriate answer – how much content it needs to communicate to ensure proper scaffolding and transparency. For example,

Algorithm 2 : Determine(AS) (*sketch*)

Require: BL is a set of beliefs relative to r, q is a question about r, and ONT is a collection of ontologies supporting subsumption inferences on sorts used in BL and q.

AS = $\emptyset$ (empty subsumption)
for b $\in$ BLs **do**
  $\phi = \top$
  for MVSV $m \in r$ existentially bound in b **do**
    introduce a bound variable m'
    $\phi = \phi \wedge m' : sort(MVSV)$
  end for
  AS = AS $\sqcup$ ϕ, under $\sqsupset$
end for

return AS

consider again the question about the color of the ball. How the question should be phrased, depends on whether e.g. the ball has already been talked about, what goals are involved (are we learning how this ball looks like, or how objects roll?), etc. Example 3 provides some illustrations.

(3) Asking about the color of a single ball on a table ...

a. If the robot is not sure whether the other agent knows about colors:
"Could you tell me about the color of this ball?"

b. If the robot believes the other agent knows about colors:
" Could you tell me what color this ball is?"

c. If the robot is not sure whether asking about color is relevant to the current goal:
"I would like to know more about the color of this ball. Could you tell me what it is?"

d. If the ball is under discussion, and asking for color is relevant:
"What's the color?"

Example 3 particularly illustrates how scaffolding and transparency come into play. We connect these terms explicitly to the question nucleus. We see scaffolding primarily as appropriately embedding a question into an intentional setting, relating to AS and the extent to which available beliefs lead to specific (potentially resolving) answers. Transparency relates to the referential setting of the question nucleus, relating r to BL in the sense of what the agent can already assume to be mutually known about the referent under discussion. Planning the question as a dialogue, then, means determining relevant beliefs, and the information status of relevant content. Relevant beliefs are those which are associated with maximally specific, potentially resolving answer(s). A distinction needs to be made between private and mutual believes, particularly as beliefs about competences are first and foremost private beliefs. Furthermore, it should be determined whether these beliefs fit into the current intentional context. (For the purposes of the current paper, we will consider learning goals only, and consider them to specify what ontological sorts the agent is trying to learn.) Information status regards whether content, pertaining to r, can be assumed to be mutually known – most notably, whether r is mutually known (i.e. mutually identifiable in context).

Algorithm 3 : Content determination (*sketch*)

Require: BL is a set of beliefs relative to r, q is a question about r, ONT is a collection of ontologies supporting subsumption inferences on sorts used in BL and q, AS is a structure over potentially resolving answers

RelBL = $\emptyset$
for a $\in$ AS **do**
  if a is maximally specific, i.e. there is no a′ s.t. a $\sqsupset$ a′ **then**
    RelBL = RelBL $\cup$ { b }, for b yielding a
  end if
end for
MutualRelBL = mutual beliefs in RelBL
ScaffoldingBL = $\emptyset$
TransparencyBL = $\emptyset$
for MVSV m in q **do**
  if there is a $b \in$ MutualRelBL associated to m **then**
    TransparencyBL = TransparencyBL $\cup$ { b }
  else
    ScaffoldingBL = ScaffoldingBL $\cup$ { beliefs associated to most specific answers for m }
  end if
end for
return ScaffoldingBL, TransparencyBL

Algorithm 3 first determines what beliefs are relevant to achieve a maximally specific answer, and which of these beliefs are mutual. How much scaffolding needs to be done depends on whether these mutual beliefs imply all potentially resolving answers to the questioned MVSVs in r. If not, the algorithm backs off by constructing a belief set which needs to be communicated for appropriate scaffolding. The basis for transparency is formed by the mutual beliefs about r.

On the basis of these sets of beliefs, and q itself, the communication of q can be planned. We do not provide an in-depth discussion of dialogue- and content-planning here, for space (and time) reasons. We refer the interested reader to (Brenner and Kruijff-Korbayová 2008; Kruijff et al. 2009). In brief, beliefs in the scaffolding set are specified as assertions (Brenner and Nebel 2008). The plan for communicating the question starts by verifying these assertions, and then raises the question itself. It is a matter for content fusion whether such verification can be done in conjunction with the question itself (Example 3, a–b) or as preceding utterances (Example 3, c). For the realization of the question, the transparency beliefs are used to determine information status. Content planning then turns information status into decisions about how to refer to r and the asked-after properties – e.g. using pronominal reference (Example 3, c) or even omitting explicit reference, by eliding any mention of r (Example 3, d).

Conclusions

The approach presented in this paper is still under development. The key technologies it is based on (planning, motivation, dialogue processing, and ontological inferencing) are already available in the system architecture the approach will be integrated into. We will describe the full integration, with working examples, in a full version of this paper. We will then also consider how this approach can be applied in related settings, such as performance requests.

We are currently considering various alternative ways to evaluate the approach. User experiments are just one option here. The problem is that an approach as presented here, and the overall architecture it will be integrated into, present a large parameter space. Consequently, it is difficult to ensure a controlled setting for a user experiment – and, only a very limited part of the parameter space can be effectively explored. An alternative way we are therefore currently considering is to use techniques from language evolution. In simulations we would like to explore what the effects of different parameter settings would be on how agents are able to communicate, and what this consequently means for measurable parameters such as learning performance. Examples of such experiments can be found in (Ginzburg and Macura 2006).

There remain for the moment plenty of open issues to be investigated further – this paper really only provides a first description of the approach we are developing. It does aim to make clear how notions such as scaffolding and transparency can be folded into a characterization of how a system can phrase a question – seeing a question, in fact, as a subdialogue to be planned, not just a single utterance paired with a possible answer. Basic issues remain in the construction of the various belief sets, and the associated structures over potentially resolving answers. Although an "unweighted" approach as followed here will work for most simple scenarios, it remains to be seen whether associating *costs* with beliefs (and assuming them, in a plan for communicating a dialogue) could provide a more adaptive, scalable approach in the long run. Furthermore, the current formulation of the construction of the answer structure AS (Algorithm 2) does not cover polar questions (though this is an easy extension).

Acknowledgments

This work was supported by the EU FP7 IST Project "CogX" (FP7-IST-215181).

References

Baldridge, J., and Kruijff, G. 2002. Coupling CCG and hybrid logic dependency semantics. In *Proc. ACL 2002*, 319–326.

Blackburn, P. 2000. Representation, reasoning, and relational structures: a hybrid logic manifesto. *Logic Journal of the IGPL* 8(3):339–625.

Brenner, M., and Kruijff-Korbayová, I. 2008. A continual multiagent planning approach to situated dialogue. In *Proceedings of the LONDIAL (The 12th SEMDIAL Workshop on Semantics and Pragmatics of Dialogue)*.

Brenner, M., and Nebel, B. 2008. Continual planning and acting in dynamic multiagent environments. *Journal of Autonomous Agents and Multiagent Systems*.

Brenner, M.; Hawes, N.; Kelleher, J.; and Wyatt, J. 2007. Mediating between qualitative and quantitative representations for task-orientated human-robot interaction. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07)*.

Dale, R., and Reiter, E. 1995. Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science* 19(2):233–263.

Ginzburg, J., and Macura, Z. 2006. Lexical acquisition with and without metacommunication. In Lyon, C.; Nehaniv, C.; and Cangelosi, A., eds., *The Emergence of Communication and Language*. Springer Verlag. 287–301.

Ginzburg, J. 1995a. Resolving questions, I. *Linguistics and Philosophy* 18(5):459–527.

Ginzburg, J. 1995b. The semantics of interrogatives. In Lappin, S., ed., *Handbook of Contemporary Semantic Theory*. Blackwell.

Grosz, B., and Kraus, S. 1999. The evolution of shared plans. In Rao, A., and Wooldridge, M., eds., *Foundations and Theories of Rational Agency*. Springer. 227–262.

Jacobsson, H.; Hawes, N.; Skocaj, D.; and Kruijff, G. 2007. Interactive learning and cross-modal binding – a combined approach. In *Language and Robots: Proceedings of the Symposium*, 1pp–1pp.

Kelleher, J., and Kruijff, G. 2006. Incremental generation of spatial referring expressions in situated dialog. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 1041–1048.

Krahmer, E.; van Erk, S.; and Verleg, A. 2003. Graph-based generation of referring expressions. *Computational Linguistics* 29(1):53–72.

Kruijff, G.; Kelleher, J.; Berginc, G.; and Leonardis, A. 2006a. Structural descriptions in human-assisted robot visual learning. In *Proceedings of the 1st Annual Conference on Human-Robot Interaction (HRI'06)*.

Kruijff, G.; Zender, H.; Jensfelt, P.; and Christensen, H. 2006b. Clarification dialogues in human-augmented mapping. In *Proceedings of the 1st Annual Conference on Human-Robot Interaction (HRI'06)*.

Kruijff, G.; Lison, P.; Benjamin, T.; Jacobsson, H.; and Hawes, N. 2007a. Incremental, multi-level processing for comprehending situated dialogue in human-robot interaction. In *Language and Robots: Proceedings from the Symposium (LangRo'2007)*.

Kruijff, G.; Zender, H.; Jensfelt, P.; and Christensen, H. 2007b. Situated dialogue and spatial organization: What, where... and why? *International Journal of Advanced Robotic Systems* 4(2).

Kruijff, G.; Lison, P.; Benjamin, T.; Jacobsson, H.; Zender, H.; and Kruijff-Korbayová, I. 2009. Situated dialogue processing for human-robot interaction. In Christensen, H.; Kruijff, G.; and

Wyatt, J., eds., *Cognitive Systems*. Available at http://www.cognitivesystems.org/cosybook.

Kruijff, G.; Brenner, M.; and Hawes, N. 2008. Continual planning for cross-modal situated clarification in human-robot interaction. In *Proceedings of the 17th International Symposium on Robot and Human Interactive Communication (RO-MAN 2008)*.

Larsson, S. 2002. *Issue-Based Dialogue Management*. Phd thesis, Department of Linguistics, Göteborg University, Göteborg, Sweden.

Li, S.; Wrede, B.; and Sagerer, G. 2006. A computational model of multi-modal grounding. In *Proc. ACL SIGdial workshop on discourse and dialog, in conjunction with COLING/ACL 2006*, 153–160.

Lochbaum, K.; Grosz, B.; and Sidner, C. 1999. Discourse structure and intention recognition. In Dale, R.; Moisl, H.; ; and Somers, H., eds., *A Handbook of Natural Language Processing: Techniques and Applications for the Processing of Language as Text*. New York: Marcel Dekker.

Paraboni, I.; van Deemter, K.; and Masthoff, J. 2007. Generating referring expressions: Making referents easy to identify. *Computational Linguistics* 33(2):229–254.

Purver, M.; Ginzburg, J.; and Healey, P. 2003. On the means for clarification in dialogue. In Smith, R., and van Kuppevelt, J., eds., *Current and New Directions in Discourse and Dialogue*, volume 22 of *Text, Speech and Language Technology*. Kluwer Academic Publishers. 235–255.

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

Thomaz, A. L. 2006. *Socially Guided Machine Learning*. Ph.D. Dissertation, Massachusetts Institute of Technology.

Zender, H.; Kruijff, G.; and Kruijff-Korbayová, I. 2009. A situated context model for resolution and generation of referring expressions. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, 126–129.