Situated Resolution and Generation of Spatial Referring Expressions for Robotic Assistants*

Hendrik Zender and **Geert-Jan M. Kruijff** and **Ivana Kruijff-Korbayová**
Language Technology Lab, German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
{zender, gj, ivana.kruijff}@dfki.de

Abstract

In this paper we present an approach to the task of generating and resolving referring expressions (REs) for conversational mobile robots. It is based on a spatial knowledge base encompassing both robot- and human-centric representations. Existing algorithms for the generation of referring expressions (GRE) try to find a description that uniquely identifies the referent with respect to other entities that are in the current context. Mobile robots, however, act in large-scale space, that is, environments that are larger than what can be perceived at a glance, e.g., an office building with different floors, each containing several rooms and objects. One challenge when referring to elsewhere is thus to include enough information so that the interlocutors can extend their context appropriately. We address this challenge with a method for context construction that can be used for both generating and resolving REs – two previously disjoint aspects. Our approach is embedded in a bi-directional framework for natural language processing for robots.

1 Introduction

The past years have seen an extraordinary increase in research on robotic assistants that help the users perform their daily chores. Although the autonomous vacuum cleaner "Roomba" has already found its way into people's homes and lives, there is still a long way until fully conversational robot "gophers" will be able to assist people in more demanding everyday tasks. For example, imagine a robot that can deliver objects and give directions to visitors on a university campus. Such a robot must be able to verbalize its knowledge in a way that is understandable by humans, as illustrated in Figure 1.

A conversational robot will inevitably face situations in which it needs to refer to an entity (e.g., an object, a locality, or even an event) that is located somewhere outside the current scene. There are conceivably many ways in which a robot might refer to things in the world, but many such expressions are unsuitable in most human-robot dialogues. Consider the following set of examples:

*Supported by the EU FP7 Project "CogX" (FP7-ICT-215181).

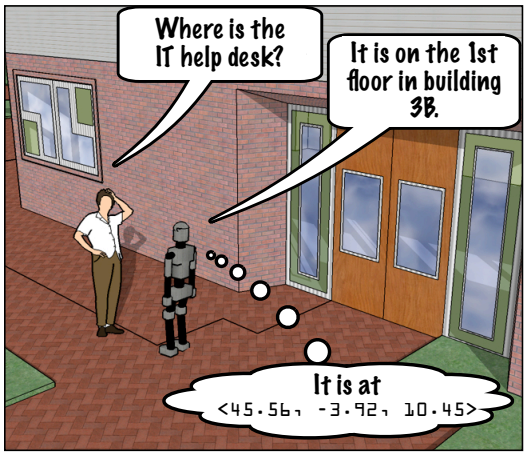


Figure 1: Situated dialogue with a campus service robot

1. "position $P = \langle 45.56, -3.92, 10.45 \rangle$"
2. "the area"
3. "Peter's office at the end of the corridor on the third floor of the Acme Corp. building 7 in the Acme Corp. complex, 47 Evergreen Terrace, Calisota, Earth, (...)"

Clearly, these REs are valid descriptions of the respective entities in the robot's world representation. Still they fail to achieve their *communicative goal*, which is to specify the right amount of information so that the hearer can easily uniquely identify what is meant. The following expressions *might* serve as more appropriate variants of the previous examples (*in certain situations!*):

1. "the IT help desk"
2. "the large hall on the first floor"
3. "Peter's office"

However, the question remains how a natural language processing (NLP) system can generate such expressions which are suitable in a given situation. In this paper we identify some of the challenges that an NLP system for situated dialogue about large-scale space needs to address. We present a situated model for generating and resolving REs that addresses these issues, with a special focus on how a conversational mobile robot can produce and interpret such expressions against an appropriate part of its acquired knowledge base (KB). One benefit of our approach is that most components, including the situated model and the linguistic resources, are bi-directional, i.e., they use the same representa-

tions for comprehension and production of utterances. This means that the proposed system is able to understand and correctly resolve all the REs that it is able to generate.

The rest of the paper is organized as follows. We first briefly discuss relevant existing approaches to comprehending and producing REs (Section 2). We then motivate our approach to context determination for situated interaction in large-scale space (Section 3), and describe its implementation in a dialogue system for an autonomous robot (Section 4). We conclude in Section 5.

2 Background

The main purpose of an RE is to enable a hearer to correctly and uniquely identify the target entity to which the speaker is referring, the so-called *intended referent*. The GRE task is thus to produce a natural language expression for a KB entity that fulfills this purpose.

As can be seen from the examples in the previous section, an RE needs to meet a number of constraints in order to be successful. First, it needs to make use of concepts that can be understood by the hearer. This becomes an important consideration when we are dealing with a robot which acquires its own models of the environment and is to talk about the contents of these. Second, it needs to contain enough information so that the hearer can distinguish the intended referent from other entities in the world, the so-called *potential distractors*. Finally, this needs to be balanced against the third constraint: Inclusion of unnecessary information should be avoided so as not to elicit false implications on the part of the hearer.

We will only briefly mention how to address the first challenge, and refer the reader to our recent work on multi-layered conceptual spatial maps for robots that bridge the gap between robot-centric representations of space and human-centric conceptualizations [Zender *et al.*, 2008].

The focus in this paper lies on the second and third aspect, namely the problem of including the right amount of information that allows the hearer to identify the intended referent. According to the seminal work on GRE by Dale and Reiter [1995], one needs to distinguish whether the intended referent is already in the hearer's *current context* or not. This context can consist of a local visual scene (visual context) or a shared workspace (spatial context), but also contains recently mentioned entities (dialogue context). If the intended referent is already part of the current context, the GRE task merely consists of singling out the referent among the other members of the context, which act as distractors. In this case the generated RE contains *discriminatory* information, e.g., "the red ball" if several kinds of objects with different colors are in the current context. If, on the other hand, the referent is not in the hearer's focus of attention, an RE needs to contain what Dale and Reiter call *navigational*, or *attention-directing* information. The example they give is "the black power supply in the equipment rack," where "the equipment rack" is supposed to direct the hearers attention to the rack and its contents.

While most existing GRE approaches assume that the intended referent is part of a given scene model, the *context set*, very little research has investigated the nature of references to entities that are not part of the current context.

The domain of such systems is usually a small visual scene, e.g., a number of objects, such as cups and tables, located in the same room, other closed-context scenarios, including a human-robot collaborative table-top scenario [Dale and Reiter, 1995; Horacek, 1997; Krahmer and Theune, 2002; Kelleher and Kruijff, 2006]. What these scenarios have in common is that they focus on a limited part of space, which is immediately and fully observable: *small-scale space*.

In contrast, mobile robots typically act in more complex environments. They operate in *large-scale space*, i.e., space "larger than what can be perceived at once" [Kuipers, 1977]. At the same time they do need the ability to understand and produce verbal references to things that are beyond the current visual and spatial context. When talking about remote places and things outside the current focus of attention, the task of *extending the context* becomes crucial.

Paraboni et al. [2007] are among the few to address this problem. They present an algorithm for *context determination* in hierarchically ordered domains, e.g., a university campus or a document structure. Their approach is mainly targeted at producing textual references to entities in written documents (e.g., figures and tables in book chapters), and consequently they do not touch upon the challenges that arise in a physically and perceptually situated dialogue setting. Nonetheless their approach presents a number of contributions towards GRE for situated dialogue in large-scale space. An appropriate context, as a subset of the full domain, is determined through Ancestral Search. This search for the intended referent is rooted in the "position of the speaker and the hearer in the domain" (represented as d), a crucial first step towards situatedness. Their approach suffers from the shortcoming that their GRE algorithm treats spatial relationships as one-place attributes. E.g., a spatial containment relation that holds between a room entity and a building entity ("the library in the Cockroft building") is given as a property of the room entity (BUILDING NAME = COCKROFT), rather than a two-place relation (`in(library,Cockroft)`). Thereby they avoid recursive calls to the GRE algorithm, which are necessary for intended referents related to another entity that needs to be properly referred to. We claim that this imposes an unnecessary restriction onto the KB design. Moreover, it makes it hard to use their context determination algorithm as a sub-routine of any of the many existing GRE algorithms.

3 Situated Dialogue in Large-Scale Space

Imagine the situation in Figure 1 did not take place somewhere on campus, but rather inside building 3B. It would have made little or no sense for the robot to say that "the IT help desk is on the 1st floor in building 3B." To avoid confusion, an utterance like "the IT help desk is on the 1st floor" would be appropriate. Likewise, if the IT help desk happened to be located on another site of the university, the robot would have had to identify its location as being, e.g., "on the 1st floor in building 3B on the new campus". This illustrates that the hierarchical representation of space that humans adopt [Cohn and Hazarika, 2001] reflects upon the choice of an appropriate context when producing referential descriptions that involve attention-directing information.

Thus, the physical and spatial situatedness of the dialogue participants plays an important role when determining which related parts of space come into consideration as potential distractors. Another important observation concerns the verbal behavior of humans when talking about remote objects and places in a complex dialogue (i.e., more than just a question and a reply). E.g., consider the following dialogue:

Person A: "Where is the exit?"

Person B: "First go down this corridor. Then turn right. After a few steps you'll see the big glass doors."

Person A: "And the bus station? Is it to the left?"

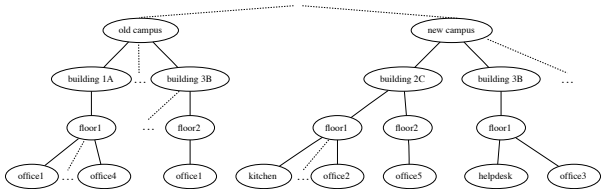
As can be seen, an utterance in such a collaborative dialogue is usually grounded in previously introduced discourse referents, both temporally and spatially. Initially, the physical surroundings of the dialogue partners form the context to which references are related. Then, as the dialogue unfolds, this point can conceptually move to other locations that have been explicitly introduced. Usually, a discourse marker denoting spatial or temporal cohesion (e.g., "then" or "there") establishes the last mentioned referent as the new anchor, creating a "mental tour" through large-scale space.

3.1 Context Determination Through Topological Abstraction

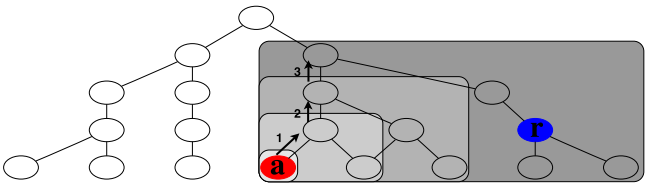
To keep track of the correct referential context in such a dialogue, we propose a general principle of *Topological Abstraction*[1] (TA) for context extension. TA is applied whenever a reference cannot be generated or resolved with respect to the current context. In such a case TA incrementally extends the context until the reference can be established. TA is designed to operate on a spatial abstraction hierarchy; i.e., a decomposition of space into parts that are related through a tree or lattice structure in which edges denote a containment relation (cf. Figure 2a). Originating in the *Referential Anchor* a, TA extends the context by incrementally ascending the spatial abstraction hierarchy until the intended referent is in the resulting sub-hierarchy (cf. Figure 2b). When no other information, e.g., from a preceding dialogue, is present, a is assumed to correspond to the spatio-visual context that is shared by the hearer and the speaker – usually their physical location and immediate surroundings. During a dialogue, however, a corresponds to the most salient discourse entity, reflecting how the *focus of attention* moves to different, even remote, places, as illustrated in the example dialogue above.

Below we describe two instantiations of the TA principle, a TA algorithm for reference generation (TAA1) and one for reference resolution (TAA2). They differ only minimally, namely in their use of an intended referent r or an RE $desc(x)$ to determine the conditions for entering and exiting the loop for topological abstraction. The way they determine a context through topological abstraction is identical.

Context Determination for GRE TAA1 (cf. Algorithm 1) constructs a set of entities dominated by the Referential Anchor a (including a itself). If this set contains the intended referent r, it is taken as the current utterance context set. Else TAA1 moves up one level of abstraction and adds the set of all child nodes to the context set. This loop continues until r is in the thus constructed set. At that point TAA1 stops and returns the constructed context set.



(a) Example for a hierarchical representation of space



(b) Illustration of the TA principle: starting from the Referential Anchor (a), the smallest sub-hierarchy containing both a and the intended referent (r) is formed incrementally

Figure 2: Topological Abstraction in a spatial hierarchy

Algorithm 1 TAA1 (for reference generation)

Require: a = referential anchor; r = intended referent
  Initialize context: $C = \{\}$
  $C = C \cup topologicalChildren(a) \cup \{a\}$
  if $r \in C$ **then**
    return C
  else
    Initialize: $SUPERNODES = \{a\}$
    for each $n \in SUPERNODES$ **do**
      for each $p \in topologicalParents(n)$ **do**
        $SUPERNODES = SUPERNODES \cup \{p\}$
        $C = C \cup topologicalChildren(p)$
      end for
      if $r \in C$ **then**
        return C
      end if
    end for
    return failure
  end if

TAA1 is formulated to be neutral to the kind of GRE algorithm that it is used for. It can be used with the original Incremental Algorithm [Dale and Reiter, 1995], augmented by a recursive call if a relation to another entity is selected as a discriminatory feature. It could in principle also be used with the standard approach to GRE involving relations [Dale and Haddock, 1991], but we agree with Paraboni et al. [2007] that the mutually qualified references that it can produce[2] are not easily resolvable if they pertain to circumstances where a confirmatory search is costly (such as in large-scale space). More recent approaches to avoiding infinite loops when using relations in GRE make use of a graph-based knowledge representation [Krahmer *et al.*, 2003; Croitoru and van Deemter, 2007]. TAA1 is compatible with these approaches, as well as with the salience based approach of Krahmer and Theune [2002].

[1] similar to Ancestral Search [Paraboni *et al.*, 2007]

[2] Stone and Webber [1998] present an approach that produces sentences like "take the rabbit from the hat" in a context with several hats and rabbits, but of which only one is in a hat. Humans find such REs natural and easy to resolve in visual scenes.

Algorithm 2 TAA2 (for reference resolution)

Require: a = ref. anchor; $desc(x)$ = description of referent
  Initialize context: $C = \{\}$
  Initialize possible referents: $R = \{\}$
  $C = C \cup topologicalChildren(a) \cup \{a\}$
  $R = desc(x) \cap C$
  if $R \neq \{\}$ **then**
    return R
  else
    Initialize: $SUPERNODES = \{a\}$
    for each $n \in SUPERNODES$ **do**
      for each $p \in topologicalParents(n)$ **do**
        $SUPERNODES = SUPERNODES \cup \{p\}$
        $C = C \cup topologicalChildren(p)$
      end for
      $R = desc(x) \cap C$
      if $R \neq \{\}$ **then**
        return R
      end if
    end for
    return failure
  end if

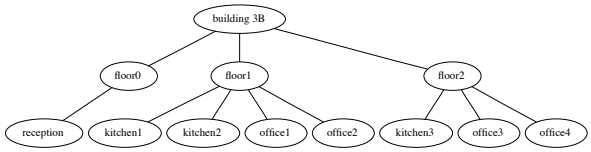
Context Determination for Reference Resolution A conversational robot must also be able to understand verbal descriptions by its users. In order to avoid overgenerating possible referents, we propose TAA2 (cf. Algorithm 2) which tries to select an appropriate referent from a relevant subset of the full KB. It is initialized with a given semantic representation of the referential expression, $desc(x)$, in a format compatible with the KB. We will show how this is accomplished in our framework in Section 4.1. Then, an appropriate entity satisfying this description is searched for in the KB. Similarly to TAA1, the description is first matched against the current *context set* C consisting of a and its child nodes. If this set does not contain any instances that match $desc(x)$, TAA2 enlarges the context set along the spatial abstraction axis until at least one possible referent can be identified within C.

4 Implementation

Our approach for resolving and generating spatial referring expressions has been fully integrated with the dialogue functionality in a cognitive system for a mobile robot [Zender *et al.*, 2008; Kruijff *et al.*, 2009]. The robot is endowed with a *conceptual spatial map* [Zender and Kruijff, 2007], which represents knowledge about places, objects and their relations in an OWL-DL[3] ontology. We use the Jena reasoning framework[4] with its built-in OWL reasoning and rule inference facilities. Internally, Jena stores the facts of the *conceptual map* as RDF[5] triples, which can be queried through SPARQL[6] queries. Figure 3 shows a subset of such a KB.

Below, we use this example scenario to illustrate our approach to generating and resolving spatial referring expressions in the robot's dialogue system. We assume that the interaction takes place at the reception on the ground floor ("floor0"), so that for TAA1 and TAA2 a =reception.

[3] http://www.w3.org/TR/owl-guide/
[4] http://jena.sourceforge.net
[5] http://www.w3.org/RDF
[6] http://www.w3.org/TR/rdf-sparql-query



(a) Topological abstraction hierarchy

```
(kitchen1 rdf:type Kitchen), (...)
(office1 rdf:type Office), (...)
(kitchen2 size big), (...)
(bob rdf:type Person), (bob name Bob),
(bob owns office1), (...)
(floor1 contains kitchen1), (...)
(floor2 contains office3), (...)
(floor1 ordNum 1), (floor2 ordNum 2), (...)
```
(b) RDF triples in the conceptual map (namespace URIs omitted)

Figure 3: Part of a representation of an office environment

4.1 The Comprehension Side

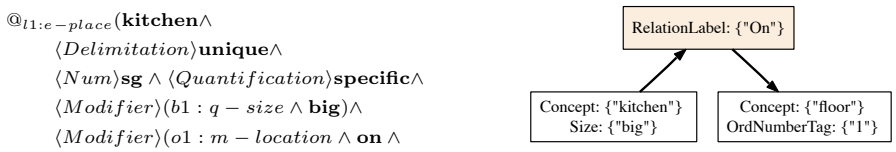
In situated dialogue processing, the robot needs to build up an interpretation for an utterance which is linked both to the dialogue context and to the (referenced) situated context. Here, we focus on the meaning representations.

We represent meaning as a logical form (LF) in a description logic [Blackburn, 2000]. An LF is a directed acyclic graph (DAG), with labeled edges, and nodes representing propositions. Each proposition has an ontological sort, and a unique index. We write the resulting ontologically sorted, relational structure as a conjunction of elementary predications (EPs): $@_{idx:sort}(\mathbf{prop})$ to represent a proposition **prop** with ontological sort $sort$ and index idx, $@_{idx1:sort1}\langle Rel\rangle(idx2 : srt2)$ to represent a relation Rel from index $idx1$ to index $idx2$, and $@_{idx:sort}\langle Feat\rangle(\mathbf{val})$ to represent a feature $Feat$ with value **val** at index idx. Representations are built compositionally, parsing the word lattices provided by speech recognition with a Combinatory Categorial Grammar [Lison and Kruijff, 2008]. Reversely, we use the same grammar to realize strings (cf. Section 4.2) from these meaning representations [White and Baldridge, 2003].

An example is the meaning we obtain for "the big kitchen on the first floor," (folding EPs under a single scope of @). It illustrates how each propositional meaning gets an index, similar to situation theory. "kitchen" gets one, and also modifiers like "big," "on" and "one." This enables us to single out every aspect for possible contextual reference (Figure 4a).

Next, we resolve contextual references, and determine the possible dialogue move(s) the utterance may express. Contextual reference resolution determines how we can relate the content in the utterance meaning, to the preceding dialogue context. If part of the meaning refers to previously mentioned content, we associate the identifiers of these content representations; else, we generate a new identifier. Consequently, each identifier is considered a dialogue referent.

Once we have a representation of utterance meaning in dialogue context, we build a further level of representation to facilitate connecting dialogue content with models of the robot's situation awareness. This next level of representation is essentially an a-modal abstraction over the linguistic aspects of meaning, to provide an a-modal conceptual structure



$@_{l1:e-place}(\mathbf{kitchen} \wedge$
$\langle Delimitation \rangle \mathbf{unique} \wedge$
$\langle Num \rangle \mathbf{sg} \wedge \langle Quantification \rangle \mathbf{specific} \wedge$
$\langle Modifier \rangle (b1 : q - size \wedge \mathbf{big}) \wedge$
$\langle Modifier \rangle (o1 : m - location \wedge \mathbf{on} \wedge$
$\langle Anchor \rangle (f1 : thing \wedge \mathbf{floor} \wedge$
$\langle Delimitation \rangle \mathbf{unique} \wedge$
$\langle Num \rangle \mathbf{sg} \wedge \langle Quantification \rangle \mathbf{specific} \wedge$
$\langle Modifier \rangle (n1 : number - ordinal \wedge \mathbf{1}))))$

(a) Logical form

(b) A-modal directed acyclic graph

```
SELECT ?x0 ?x1 WHERE {
  ?x0 rdf:type Kitchen.
  ?x0 size big.
  ?x1 rdf:type Floor.
  ?x1 ordNum 1.
  ?x0 containedIn ?x1.
```

(c) SPARQL query
In the previous example this would resolve `?x0` to `kitchen2`

Figure 4: Logical form, a-modal DAG and corresponding SPARQL query for "the big kitchen on the first floor"

[Jacobsson *et al.*, 2008]. Abstraction is a recursive translation of DAGs into DAGs, whereby the latter (conceptual) DAGs are typically flatter than the linguistic DAGs (Figure 4b).

The final step in resolving an RE is to construct a query to the robot's KB. In our implementation we construct a SPARQL query from the a-modal DAG representations (Figure 4c). This query corresponds to the logical description of the referent $desc(r)$ in TAA2. TAA2 then incrementally extends the context until at least one element of the result set of $desc(r)$ is contained within the context.

4.2 The Production Side

Production covers the entire path from handling dialogue goals to speech synthesis. The dialogue system can itself produce goals (e.g., to handle communicative phenomena like greetings), and it accepts goals from a higher level planner. Once there is a goal, an utterance content planner produces a content representation for achieving that goal, which the realizer then turns into one or more surface forms to be synthesized. Below we focus on utterance content planning.

A dialogue goal specifies a goal to be achieved, and any content that is associated with it. A typical example is to convey an answer to a user: the goal is to tell, the content is the answer. Content is given as a conceptual structure, *proto LF*, abstracting away from linguistic specifics, similar to the a-modal structures we produce for comprehension.

Content planning turns this proto LF into an LF which matches the specific linguistic structures defined in the grammar we use to realize it. "Turning into" means extending the proto LF with further semantic structure. This may be non-monotonic in that parts of the proto LF may be rewritten, expanding into locally connected graph structures.

Planning is agenda-based, and uses a planning domain defined as a (systemic) grammar network alike [Bateman, 1997; Kruijff, 2005]. A grammar network is a collection of systems that define possible sequences of operations to be performed on a node with characteristics matching the applicability conditions for the system. A system's decision tree determines which operations are to be applied. Decisions are typically context-sensitive, based on information about the shape of the (entire) LF, or on information in context models (dialogue or otherwise). While constructing an LF, the planner cycles over its nodes, and proposes new agenda items for nodes which have not yet been visited. An agenda item consists of the node, and a system which can be applied to that node.

A system can explicitly trigger the generation of an RE for the node on which it operates. It then provides the dialogue system with a request for an RE, with a pointer to the node in the (provided) LF. The dialogue system resolves this request by submitting it to GRE modules which have been registered with the system. (Registration allows us to plug-and-play with content-specific GRE algorithms.) Assuming a GRE module produces an LF with the content for the RE, the planner gets this LF and integrates it into the overall LF.

For example, say the robot in our previous example is to answer the question "Where is Bob?". We receive a communicative goal (see below) to inform the user, specifying the goal as an assertion related to the previous dialogue context as an answer. The content is specified as an ascription e of a property to a target entity. The target entity is t which is specified as a person called "Bob" already available in the dialogue context, and thus familiar to the hearer. The property is specified as topological inclusion (TopIn) within the entity k, the reference to which is to be produced by the GRE algorithm (hence the type "rfx" and the "RefIndex" which is the address of the entity).

$@_{d:dvp}(c - goal \wedge$
$\langle SpeechAct \rangle \mathbf{assertion} \wedge$
$\langle Relation \rangle \mathbf{answer} \wedge$
$\langle Content \rangle (e : ascription \wedge$
$\langle Target \rangle (t : person \wedge Bob \wedge$
$\langle InfoStatus \rangle \mathbf{familiar}) \wedge$
$\langle TopIn \rangle (p : rfx \wedge RefIndex)))$

The content planner makes a series of decisions about the type and structure of the utterance to be produced. As it is an assertion of a property ascription, it decides to plan a sentence in indicative mood and present tense with "be" as the main verb. The reference to the target entity makes up the copula restriction, and a reference to the ascribed property is in the copula scope. This yields an expansion of the goal content:

$@_{e:ascription}(\mathbf{be} \wedge$
$\langle Tense \rangle \mathbf{pres} \wedge$
$\langle Mood \rangle \mathbf{ind} \wedge$
$\langle Cop - Restr \rangle (t : entity \wedge$
$Bob \wedge \langle InfoStatus \rangle \mathbf{familiar}) \wedge$
$\langle Subject \rangle (t : entity) \wedge$
$\langle Cop - Scope \rangle (prop : m - location \wedge$
$\mathbf{in} \wedge \langle Anchor \rangle (p : rfx \wedge RefIndex)))$

The next step consists in calling the GRE algorithm to produce an RE for the entity p. In our NLP system we use a slightly modified implementation of the Incremental Algorithm [Dale and Reiter, 1995]. The context set C is determined using TAA1. Let's assume that Bob is currently in

kitchen3. In our example (a =reception) the GRE algorithm hence produces the following result, which is then returned to the planner and inserted into the proto LF created so far:

$$@_{p:entity}(\mathbf{kitchen} \wedge \langle TopOn\rangle(f : entity \wedge floor \wedge \langle Unique\rangle\mathbf{true} \wedge \langle Number\rangle(n : quality \wedge 2)))$$

The planner then makes further decisions about the realization, expanding this part of the LF to the following result:

$$@_{p:entity}(\mathbf{kitchen} \wedge \langle Delimitation\rangle\mathbf{unique} \wedge \langle Num\rangle\mathbf{sg} \wedge Quantification\rangle\mathbf{specific} \wedge \langle Modifier\rangle(o1 : m - location \wedge on \wedge \langle Anchor\rangle(f : thing \wedge floor \wedge \langle Delimitation\rangle\mathbf{unique} \wedge \langle Num\rangle\mathbf{sg} \wedge \langle Quantification\rangle\mathbf{specific} \wedge \langle Modifier\rangle(t1 : number - ordinal \wedge 2))))$$

Once the planner is finished, the resulting overall LF is provided to a CCG realizer [White and Baldridge, 2003], turning it into a surface form ("Bob is in the kitchen on the second floor"). This string is synthesized to speech using the MARY TTS software [Schröder and Trouvain, 2003].

5 Conclusions and Future Work

We have presented an algorithm for context determination that can be used both for resolving and generating referring expressions in a large-scale space domain. We have presented an implementation of this approach in a dialogue system for an autonomous mobile robot.

Since there exists no suitable evaluation benchmark for situated human-robot dialogue to compare our results against, we are currently planning a user study to evaluate the performance of the TA algorithm. Another important item for future work is the exact nature of the spatial progression in situated dialogue, modeled by "moving" the referential anchor.

References

[Bateman, 1997] J. A. Bateman. Enabling technology for multilingual natural language generation: the KPML development environment. *Journal of Natural Language Engineering*, 3(1):15–55, 1997.

[Blackburn, 2000] P. Blackburn. Representation, reasoning, and relational structures: a hybrid logic manifesto. *Journal of the Interest Group in Pure Logic*, 8(3):339–365, 2000.

[Cohn and Hazarika, 2001] A. G. Cohn and S. M. Hazarika. Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 46:1–29, 2001.

[Croitoru and van Deemter, 2007] M. Croitoru and K. van Deemter. A conceptual graph approach to the generation of referring expressions. In *Proc. IJCAI-2007*, Hyderabad, India, 2007.

[Dale and Haddock, 1991] R. Dale and N. Haddock. Generating referring expressions involving relations. In *Proc. EACL-1991*, Berlin, Germany, April 1991.

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

[Horacek, 1997] H. Horacek. An algorithm for generating referential descriptions with flexible interfaces. In *Proc. ACL/EACL-1997*, Madrid, Spain, 1997.

[Jacobsson *et al.*, 2008] H. Jacobsson, N. Hawes, G. J. Kruijff, and J. Wyatt. Crossmodal content binding in information-processing architectures. In *Proc. HRI-2008*, Amsterdam, The Netherlands, 2008.

[Kelleher and Kruijff, 2006] J. Kelleher and G. J. Kruijff. Incremental generation of spatial referring expressions in situated dialogue. In *In Proc. Coling-ACL-2006*, Sydney, Australia, 2006.

[Krahmer and Theune, 2002] E. Krahmer and M. Theune. Efficient context-sensitive generation of referring expressions. In K. van Deemter and R.Kibble, editors, *Information Sharing: Givenness and Newness in Language Processing*. CSLI Publications, Stanford, CA, USA, 2002.

[Krahmer *et al.*, 2003] E. Krahmer, S. van Erk, and A. Verleg. Graph-based generation of referring expressions. *Computational Linguistics*, 29(1), 2003.

[Kruijff *et al.*, 2009] G. J. Kruijff, P. Lison, T. Benjamin, H. Jacobsson, H. Zender, I. Kruijff-Korbayová, and N. Hawes. Situated dialogue processing for human-robot interaction. In H. I. Christensen, G. J. Kruijff, and J. Wyatt, editors, *Cognitive Systems*. Springer, 2009. to appear.

[Kruijff, 2005] G. J. Kruijff. Context-sensitive utterance planning for CCG. In *Proc. ENLG-2005*, Aberdeen, Scotland, 2005.

[Kuipers, 1977] B. Kuipers. *Representing Knowledge of Large-scale Space*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1977.

[Lison and Kruijff, 2008] P. Lison and G. J. Kruijff. Salience-driven contextual priming of speech recognition for human-robot interaction. In *ECAI 2008*, 2008.

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

[Schröder and Trouvain, 2003] M. Schröder and J. Trouvain. The german text-to-speech synthesis system MARY: A tool for research, development and teaching. *Int. Journal of Speech Technology*, 6:365–377, 2003.

[Stone and Webber, 1998] M. Stone and B. Webber. Textual economy through close coupling of syntax and semantics. In *Proc. INLG-1998*, pages 178–187, Niagara-on-the-Lake, ON, Canada, 1998.

[White and Baldridge, 2003] M. White and J. Baldridge. Adapting chart realization to CCG. In *Proc. ENLG-2003*, Budapest, Hungary, 2003.

[Zender and Kruijff, 2007] H. Zender and G. J. Kruijff. Multi-layered conceptual spatial mapping for autonomous mobile robots. In *Control Mechanisms for Spatial Knowledge Processing in Cognitive / Intelligent Systems*, AAAI Spring Symposium 2007, March 2007.

[Zender *et al.*, 2008] H. Zender, O. Martínez Mozos, P. Jensfelt, G. J. Kruijff, and W. Burgard. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6):493–502, June 2008.