A context-dependent algorithm for generating locative expressions in physically situated environments

John D. Kelleher & Geert-Jan M. Kruijff
Language Technology Lab
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
{kelleher,gj}@dfki.de

Abstract

This paper presents a framework for generating locative expressions. The framework addresses the issue of combinatorial explosion inherent in the construction of relational context models by: (a) contextually defining the set of objects in the context that may function as a landmark, and (b) sequencing the order in which spatial relations are considered using a cognitively motivated hierarchy of relations.

1 Introduction

Our long-term goal is to develop embodied conversational robots that are capable of natural, fluent visually situated dialog with one or more interlocutors. An inherent aspect of visually situated dialog is reference to objects located in the physical environment. In this paper, we present a computational framework for the generation of a spatial locative expressions in such contexts.

In the simplest form of locative expression, a prepositional phrase modifies a noun phrase to explicitly specify the location of the object. (1) is an example of the type of locative we focus on generating. In this example, *the book* is the subject of the expression and *the table* is the object. Following [Langacker, 1987], we use the terms **trajector** and **landmark** to respectively denote the subject and the object of a locative expression: the location of the trajector is specified relative to the landmark by the semantics of the preposition.

(1) a. the book [subject] on the table [object]

Generating locative expressions is part of the general field of generating referring expressions (GRE). Most GRE algorithms deal with the same problem: given a domain description and a **target object** generate a description of the target object that distinguishes it from the other objects in the domain. The term **distractor objects** is used to describe the objects in the context excluding the trajector that at a given point in processing fulfil the description of the target object that has been generated. The description generated is said to be **distinguishing** when the set of distractor objects is empty.

Several GRE algorithms have addressed the issue of generating locative expressions [Dale and Haddock, 1991; Horacek, 1997; Gardent, 2002; Krahmer and Theune, 2002; Varges, 2004]. However, all these algorithms assume the GRE component has access to a predefined scene model. For an embodied conversational robot functioning in dynamic partially known environments this assumption is a serious drawback. If an agent wishes to generate a contextually appropriate reference it cannot assume the availability of a domain model, rather it must dynamically construct one. However, constructing a model containing all the relationships between all the entities in the domain is prone to combinatorial explosion, both in terms of the number of objects in the context (the location of each object in the scene must be checked against all the other objects in the scene) and number of inter-object spatial relations (as a greater number of spatial relations will require a greater number of comparisons between each pair of objects.[1] Moreover, the context free *a priori* construction of such an exhaustive scene model is cognitively implausible. Psychological research indicates that spatial relations are not preattentively perceptually available [Treisman and Gormican, 1988]. Rather, their perception requires attention [Logan, 1994; 1995]. These findings point to subjects constructing contextually dependent reduced relational scene models, rather than an exhaustive context free model.

Contributions In this paper we present a framework for generating locative expressions. This framework addresses the issue of combinatorial explosion inherent in relational scene model construction by incrementally creating a series of reduced scene models. Within each scene model only one spatial relations is considered and only a subset of objects are considered as candidate landmarks. This reduces both the number of relations that must be computed over each object pair and the number of object pairs. The decision as to which relations should be included in each scene model is guided by a cognitively motivated hierarchy of spatial relations. The set of candidate landmarks in a given scene is dependent on the set of objects in the scene that fulfil the description of the

[1] In English, the vast majority of spatial locatives are binary, some notable exceptions include: *between*, *amongst* etc. However, we will not deal with these exceptions in this paper

target object and the relation that is being considered.

Overview In §2 we present some background data relevant to our discussion. In §3 we present our GRE framework. In §4 we illustrate the framework with a worked example and expand on some of the issues relevant to the framework. We end with conclusions.

2 Data

When one considers that the English lexicon of spatial prepositions numbers above eighty members (not considering compounds such as *right next to*) [Landau, 1996], the combinatorial aspect of relational scene model construction becomes apparent. It should be noted that for our purposes, the situation is somewhat ameliorated by the fact that a distinction can be made between static and dynamic prepositions: static prepositions primarily[2] denote the location of an object, dynamic prepositions primarily denote the path of an object [Jackendoff, 1983; Herskovits, 1986], see (2). However, even focusing exclusively on the set of static prepositions does not remove the combinatorial issues effecting the construction of a scene model.

(2) a. the tree is behind [static] the house
b. the man walked across [dynamic] the road

In general, the set of static prepositions can be decomposed into two sets called **topological** and **projective**. Topological prepositions are the category of prepositions referring to a region that is proximal to the landmark; e.g., *at*, *near*, etc. Often, the distinctions between the semantics of the different topological prepositions is based on pragmatic contraints, for example the use of *at* licences the trajector to be in contact with the landmark, by contrast the use of *near* does not. Projective prepositions describe a region projected from the landmark in a particular direction, the specification of the direction is dependent on the frame of reference being used; e.g., *to the right of*, *to the left of*, etc.

The semantics of static prepositions exhibit both qualitative and quantitative properties. The qualitative aspect of their semantics is evident when they are used to denote an object by contrasting its location with the distractor objects location. Taking Figure 1 as a visual context the locative expression *the circle on the left of the square* exhibits the contrastive semantics of a projective preposition. Only one of the circles in the scene is located in the region *to the right of the square*. Taking Figure 2 as a visual context the locative expression *the circle near the black square* illustrates the contrastive semantics of a topological preposition. Again, of the two circles in the scene only one of them may be appropriately described as being *near the black square*, the other circle is more appropriately described as being *near the white square*. The quantitative aspect of the semantics of static prepositions is evident when they denote an object using a relative scale. In the context provided by Figure 3 the locative *the circle to the right of the square* exhibits the relative semantics of a projective preposition. Although both the circles are located *to the right of the square* it is possible to adjudicate between them based on their location in the region. The relative semantics of a topological preposition can also be illustrated using Figure 3. A description such as *the circle near the square* could be applied to either circle if the other circle was not present. However, when both are present it is possible to interpret the reference based on their relative proximity to the landmark *the square*.

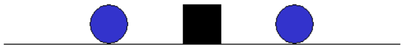

Figure 1: Visual context used to illustrate the contrastive semantics of projective prepositions.

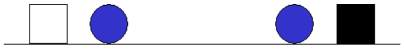

Figure 2: Visual context used to illustrate the contrastive semantics of topological prepositions.

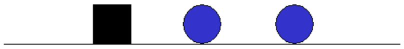

Figure 3: Visual context used to illustrate the relative semantics of topological and projective prepositions.

3 Approach

The approach we adopt to generating locative expressions involves extending the incremental algorithm [Dale and Reiter, 1995]. The motivation for this is the polynomial complexity of the incremental algorithm. The incremental algorithm iterates through the properties of the target and for each property computes the set of distractor objects for which (a) the conjunction of the properties selected so far, and (b) the current property hold. A property is added to the list of selected properties if it reduces the size of the distractor object set. The algorithm succeeds when all the distractors have been ruled out, it fails if all the properties have been processed and there are still some distractor objects. The algorithm can be refined by ordering the checking of properties according to fixed preferences, e.g. first a taxonomic description of the target, second an absolute property such as colour, third a relative property such as size. [Dale and Reiter, 1995] also stipulate that the type description of the target should be included in the description even if its inclusion does not distinguish the target from any of the distractors, see Algorithm 1.

However, before applying the incremental algorithm we must construct a context model within which we can check whether or not the description generated distinguishes the target object. In order to constrain the combinatorial issues inherent in relational scene model construction we construct a series of reduced scene models, rather than constructing one complex exhaustive model. This construction process is driven by a hierarchy of spatial relations and the partitioning of the context model into objects that may and may not function as landmarks. These two components are developed

[2] Static prepositions can be used in dynamic contexts, e.g. *the man ran behind the house*, and dynamic prepositions can be used in static ones, e.g. *the tree lay across the road*.

Algorithm 1 The Basic Incremental Algorithm

```
Require: T = target object; D = set of distractor objects.
  Initialise: $P = \{type, colour, size\}$; $DESC = \{\}$
  for $i = 0$ to $|P|$ do
    if $|D| \neq 0$ then
      $D\prime = \{x : x \in D, P_i(x) = P_i(T)\}$
      if $|D\prime| < |D|$ then
        $DESC = DESC \cup P_i(T)$
        $D = \{x : x \in D, P_i(x) = P_i(T)\}$
      end if
    else
      Distinguishing description generated
      if $type(x) \notin DESC$ then
        $DESC = DESC \cup type(x)$
      end if
      return $DESC$
    end if
  end for
  Failed to generate distinguishing description
  return $DESC$
```

in the next two sections. In §3.1 we develop the hierarchy of spatial relations and in §3.2 we develop a classification of landmarks and use these groupings to create a definition of a distinguishing locative description. In §3.3 we present the generation algorithm that integrates these components.

3.1 Cognitive Ordering of Contexts

Psychological research indicates that spatial relations are not preattentively perceptually available [Treisman and Gormican, 1988]. Rather, their perception requires attention [Logan, 1994; 1995]. These findings point to subjects constructing contextually dependent reduced relational scene models, rather than an exhaustive context free model. Mimicking this, we have developed an approach to context model construction that attempts to constrain the combinatorial explosion inherent in the construction of relational context models by incrementally constructing a series of reduced context models. Each context model focuses on a different spatial relation. The ordering of the spatial relations is based on the cognitive load of interpreting the relation. In this section, we motivate and develop the ordering of relations used.

It seems reasonable to asssume that it takes less effort to describe one object than two. Consequently, following the Principle of Minimal Cooperative Effort [Clark and Wilkes-Gibbs, 1986], a speaker should only use a locative expression when they cannot create a distinguishing description of the target object using a simple feature based approach. Moreover, the Principle of Sensitivity [Dale and Reiter, 1995] states that when producing a referring expression, the speaker should prefer features which the hearer is known to be able to interpret and perceive. This points to a preference, due to cognitive load, towards descriptions that distinguish an object using purely physical and easily perceivable features over descriptions that use spatial expressions. Psycholinguistic results support this preference [van der Sluis and Krahmer, 2004].

Similarly, we can distinguish between the cognitive loads of processing different forms of spatial relations. In comparing the cognitive load associated with different spatial relations it is important to recognize that they are represented and processed at several levels of abstraction. For example, the **geometric level**, where metric properties are dealt with, the **functional level**, where the specific properties of spatial entities deriving from their functions in space are considered, and the **pragmatic level**, which gathers the underlying principles that people use in order to discard wrong relations or to deduce more information [Edwards and Moulin, 1998]. Our discussion is grounded at the geometric level of representation and processing.

Focusing on static prepositions, it is reasonable to propose that topological prepositions have a lower perceptual load than projective prepositions, due to the relative ease of perceiving two objects that are close to each other and the complex processing required to handle frame of reference ambiguity [Carlson-Radvansky and Irwin, 1994; Carlson-Radvansky and Logan, 1997]. Figure 4 lists these preferences, with further distinctions among features: objects type is the easiest to process, before absolute gradable predicates (e.g. color), which is still easier than relative gradable predicates (e.g. size) [Dale and Reiter, 1995].

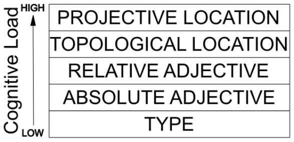

Figure 4: Cognitive load of reference forms

This topological versus projective preference can be further refined if we consider the contrastive and relative uses of these relations noted in §2. Perceiving and interpreting a constrastive use of a spatial relation is computationally easier than judging a relative use. Finally, within the set of projective prepositions, psycholinguistic data indicates a perceptually based ordering of the relations: *above*/*below* are easier to percieve and interpret than *in front of*/*behind* which in turn are easier than *to the right of*/*to the left of* [Bryant *et al.*, 1992; Gapp, 1995].

In sum, we would like to propose the following ordering of spatial relations:

1. topological contrastive
2. topological relative
3. projective constrastive [above/below, front/back/, right/left]
4. projective relative [above/below, front/back, right/left]

For each level of this hierarchy we require a computational model of the semantics of the relation at that level that accomodates both contrastive and relative representations. In §2 we noted that the distinctions between the semantics of the different topological prepositions is often based on functional and pragmatic issues.[3] Currently, however, more psycholinguistic data is required to distinguish the cognitive load associated with the different topological prepositions. We use the

[3] See *inter alia* [Talmy, 1983; Herskovits, 1986; Vandeloise, 1991; Fillmore, 1997; Garrod *et al.*, 1999] for more discussion on these differences

model of topological proximity developed in [Kelleher and Kruijff, 2005] to model all the relations at this level. Using this model we can define the extent of a region proximal to an object. If the trajector or one of the distractor objects is the only object within the region of proximity around a given landmark this is taken to model a contrastive use of a topological relation relative to that landmark. If the landmark's region of proximity contains more than one object from the trajector and distractor object set then it is a relative use of a topological relation. We handle the issue of frame of reference ambiguity and model the semantics of projective prepostions using the framework developed in [Kelleher and van Genabith, 2005]. Here again, the contrastive-relative distinction is dependent on the number of objects within the region of space defined by the preposition.

3.2 Landmarks and Distinguishing Descriptions

In order to use a locative expression an object in the context must be selected to function as the landmark. An implicit assumption in selecting an object to function as a landmark is that the hearer can easily identify and locate the object within the context. A landmark can be: the speaker (3)a, the hearer (3)b, the scene (3)c, an object in the scene (3)d, or a group of objects in the scene (3)e.[4]

(3) a. the ball on *my* right [speaker]
b. the ball to *your* left [hearer]
c. the ball on the right [scene]
d. the ball to the left of *the box* [an object in the scene]
e. the ball in the middle [group of objects]

Currently, new empirical research is required to see if there is a preference order between these landmark categories. Intuitively, in most situations, either of the interlocutors are ideal landmarks because the speaker can naturally assume that the hearer is aware of the speaker's location and their own. Focusing on instances where an object in the scene is used as a landmark, several authors [Talmy, 1983; Landau, 1996; Gapp, 1995] have noted a trajector-landmark asymmetry: generally, the landmark object is more permanently located, larger, and taken to have greater geometric complexity. These characteristics are indicative of salient objects and empirical results support this correlation between object salience and landmark selection [Beun and Cremers, 1998]. However, the salience of an object is intrinsically linked to the context it is embedded in. For example, in the context provided by Figure 5 the ball has a relatively high salience, because it is a singleton, despite the fact that it is smaller and geometrically less complex than the other figures. Moreover, in this context, the ball is the only object in the scene that can function as a landmark without recourse to using the scene itself or a grouping of objects in the scene.

Clearly, deciding which objects in a given context are suitable to function as landmarks is a complex and contextually dependent process. Some of the factors effecting this decision are object salience and the functional relationships between objects. However, one basic constraint on landmark selection is that the landmark should be distinguishable from the trajector. For example, given the context in Figure 5 and all other factors being equal, using a locative such as *the man to the left of the man* would be much less helpful than using *the man to the right of the ball*. Following this observation, we treat an object as a **candidate landmark** if the trajector object can be distinguished from it using the basic incremental algorithm, Algorithm 1.[5] Furthermore, a **trajector landmark** is a member of the candidate landmark set that stands in relation to the trajector and a **distractor landmark** is a member of the candidate landmark set that stands in relation to a distractor object under the relation being considered. Using these categories of landmark we can define a **distinguishing locative description** as a locative description where there is trajector landmark that can be distinguished from all the members of the set of distractor landmarks under the relation used in the locative.

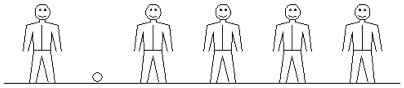

Figure 5: Visual context used to illustrate the relative semantics of topological and projective prepositions.

We can illustrate these different categories of landmark using Figure 6 as the visual context. In this context, if W1 is taken as the target object, the distractor set equals {T1,B1,W2,B2}. Running the basic incremental algorithm would generate the description *white block*. This distinguishes W1 from T1, B1 and B2 but not from W2. Consequently, the set of candidate landmarks equals {T1,B1,B2}. If we now create a context model for the relation *near* the set of trajector landmarks would be {T1,B1} and the set of distractor landmarks would be {B1,B2}. Obviously, B1 cannot be distinguished from all the distractor landmarks as it cannot be distinguished from itself. As a result, B1 cannot function as the landmark for a distinguishing locative description for W1 using the relation *near*. However, T1 can be distinguished from the distractor landmarks B1 and B2 by its type, *triangle*. So *the white block near the triangle* would be considered a distinguishing description.

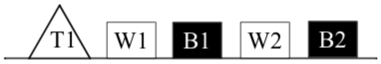

Figure 6: Visual context used to illustrate the different categories of landmark.

[4] See [Gorniak and Roy, 2004] for further discussion on the use of spatial extrema of the scene and groups of objects in the scene as landmarks

[5] As noted by one of our reviewers, one unwanted effect of this definition of a landmark is that it precludes the generation of descriptions that use a landmark that are themselves distinguished using a locative expression. For example, *the block to the right of the block which has a ball on it*.

3.3 Algorithm

The basic approach is to try to generate a distinguishing description using the standard incremental algorithm. If this fails, we divide the context into three components:

the trajector: the target object,

the distractor objects: the objects that match the description generated for the target object by the standard incremental algorithm,

the set of candidate landmarks: the objects that do **not** match the description generated for the target object by the standard incremental algorithm.

We then begin to iterate through the hierarchy of relations and for each relation we create a context model that defines the set of trajector and distractor landmarks. Once a context model has been created we iterate through the trajector landmarks (using a salience ordering if there is more than one)[6] and try to create a distinguishing locative description. A distinguishing locative description is created by using the basic incremental algorithm to distinguish the trajector landmark from the distractor landmarks. If we succeed in generating a distinguishing locative description we return the description and stop processing. Algorithm 2 lists the steps in the algorithm.

Algorithm 2 The Locative Algorithm

```
Require: T = target object; D = set of distractor objects; R = hierarchy of relations.
  DESC = Basic-Incremental-Algorithm(T,D)
  if $DESC \neq Distinguishing$ then
    create $CL$ the set of candidate landmarks
    $CL = \{x : x \neq T, DESC(x) = false\}$
    for $i = 0$ to $|R|$ do
      create a context model for relation $R_i$ consisting of $TL$ the set of trajector landmarks and $DL$ the set of distractor landmarks
      $DL = \{z : z \in CL, R_i(D, z) = true\}$
      $TL = \{y : y \in CL, y \notin DL, R_i(T, y) = true\}$
      for $j = 0$ to $|TL|$ by salience(TL) do
        LANDDESC = Basic-Incremental-Algorithm(TL$_j$, DL)
        if $LANDDESC = Distinguishing$ then
          Distinguishing locative generated
          return {DESC,R$_i$,LANDDESC}
        end if
      end for
    end for
  end if
  FAIL
```

If we cannot create a distinguishing locative description we are faced with a choice of: (1) iterate on to the next relation in the hierarchy, (2) create an embedded locative description that distinguishes the landmark. Currently, we prefer option (1) over (2), preferring *the dog to the right of the car* over *the dog near the car to the right of the house*. However, the algorithm can generate these longer embedded descriptions if needed. This is done by replacing the call to the basic incremental algorithm for the trajector landmark object with a call to the whole locative expression generation algorithm, with the trajector landmark as the target object and the set of distractor landmarks as the distractor objects. Algorithm 3 lists the steps in the recursive version of the algorithm.

Algorithm 3 The Recursive Locative Algorithm

```
Require: T = target object; D = set of distractor objects; R = hierarchy of relations.
  DESC = Basic-Incremental-Algorithm(T,D)
  if $DESC \neq Distinguishing$ then
    create $CL$ the set of candidate landmarks
    $CL = \{x : x \neq T, DESC(x) = false\}$
    for $i = 0$ to $|R|$ do
      create a context model for relation $R_i$ consisting of $TL$ the set of trajector landmarks and $DL$ the set of distractor landmarks
      $DL = \{z : z \in CL, R_i(D, z) = true\}$
      $TL = \{y : y \in CL, y \notin DL, R_i(T, y) = true\}$
      for $j = 0$ to $|TL|$ by salience(TL) do
        LANDDESC =
        Recursive-Locative-Algorithm(T=TL$_j$,D=DL,R)
        if $LANDDESC = Distinguishing$ then
          Distinguishing locative generated
          return {DESC,R$_i$,LANDDESC}
        end if
      end for
    end for
  end if
  FAIL
```

For both versions of the locative algorithm an important consideration is the issue of infinite regression. As noted by [Dale and Haddock, 1991] a compositional GRE system may, in certain contexts, generate an infinite description by trying to distinguish the landmark in terms of the trajector and the trajector in terms of the landmark, see (4). However, this infinite recursion can only occur if the context is not modified between calls to the algorithm. This issue does not effect Algorithm 2 because each call to the algorithm results in the domain being partitioned into those objects that can and cannot be used as landmarks. One effect of this partitioning is a reduction in the number of object pairs that relations must be computed for. However, and more importantly for this discussion, another consequence of this partitioning is that the process of creating a distinguishing description for a landmark is carried out in a context that is a subset of the context the trajector description was generated in. The distractor set used during the generation of a landmark description is the set of distractor landmarks. This minimally excludes the trajector object, since by definition the landmark objects cannot fulfill the description of the trajector generated by the basic incremental algorithm. This naturally removes the possibility for the algorithm to distinguish a landmark using its trajector.

[6] We model both visual and linguistic salience. Visual salience is computed using a modified version of the visual saliency algorithm described in [Kelleher and van Genabith, 2004]. Discourse salience is computed based on recency of mention as defined in [Hajicová, 1993] except we represent the maximum overall salience in the scene as 1, and use 0 to indicate object is not salient. We integrate these two components by summing them and dividing the result by 2.

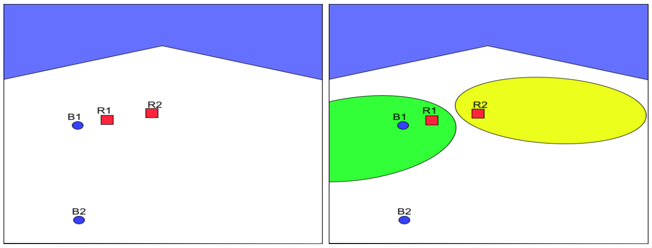

Figure 7: A visual scene and the topological analsis of R1 and R2

(4) the bowl on the table supporting the bowl on the table supporting the bowl ...

4 Discussion

We can illustrate the framework using the visual context provided by the scene on the left of Figure 7. This context consists of two red boxes R1 and R2 and two blue balls B1 and B2. Imagine that we want to refer to B1. We begin by calling the locative incremental algorithm, Algorithm 2. This in turn calls the basic incremental algorithm, Algorithm 1, which will return the property *ball*. However, this is not sufficient to create a distinguishing description as B2 is also a ball. In this context the set of candidate landmarks equals {R1,R2} and the first relation in the hierarchy is topological proximity, which we model using the algorithm developed in [Kelleher and Kruijff, 2005]. The image on the right of Figure 7 illustrates the analysis of the scene using this framework: the green region on the left defines the area deemed to be proximal to R1, and the yellow region on the right defines the area deemed to be proximal to R2. It is evident that B1 is in the area proximal to R1, consequently R1 is classified as a trajector landmark. As none of the distractors (i.e., B2) are located in a region that is proximal to a candidate landmark there are no distractor landmarks. As a result when the basic incremental algorithm is called to create a distinguishing description for the trajector landmark R1 it will return *box* and this will be deemed to be a distinguishing locative description. The overall algorithm will then return the vector {*ball, proximal, box*} which would result in the realiser generating a reference of the form: *the ball near the box*.

The relational hierarchy used by the framework has some commonalities with the relational subsumption hierarchy proposed in [Krahmer and Theune, 2002]. However, there are two important differences between them. First, an implication of the subsumption hierarchy proposed in [Krahmer and Theune, 2002] is that the semantics of the relations at lower levels in the hierarchy are subsumed by the semantics of their parent relations. For example, in the portion of the subsumption hierarchy illustrated in [Krahmer and Theune, 2002] the relation *next to* subsumes the relations *left of* and *right of*. By contrast, the relational hierarchy developed here is based solely on the relative cognitive load associated with the semantics of the spatial relations and makes no claims as to the semantic relationships between the semantics of the spatial relations. Secondly, [Krahmer and Theune, 2002] do not use their relational hierarchy to guide the construction of domain models.

By providing a basic contextual definition of a landmark we are able to partition the context in an appropriate manner. This partitioning has two advantages:

1. it reduces the complexity of the context model construction, as the relationships between the trajector and the distractor objects or between the distractor objects themselves do not need to be computed;
2. the context used during the generation of a landmark description is always a subset of the context used for a trajector (as the trajector, its distractors and the other objects in the domain that do not stand in relation to the trajector or distractors under the relation being considered are excluded). As a result the framework avoids the issue of infinite recusion. Furthermore, the trajector-landmark relationship is automatically included as a property of the landmark as its feature based description need only distinguish it from objects that stand in relation to one of the distractor objects under the same spatial relationship.

.

In future work we will focus on extending the framework to handle some of the issues effecting the incremental algorithm, see [van Deemter, 2001]. For example, generating locative descriptions containing negated relations, conjunctions of relations and involving sets of objects (sets of trajectors and landmarks).

5 Conclusions

In this paper we have argued that an if an embodied conversational agent functioning in dynamic partially known environments wishes to generate contextually appropriate locative expressions it must be able to construct a context model that explicitly marks the spatial relations between objects in the scene. However, the construction of such a model is prone to the issue of combinatorial explosion both in terms of the number of objects in the context (the location of each object in the scene must be checked against all the other objects in the scene) and number of inter-object spatial relations (as a greater number of spatial relations will require a greater number of comparisons between each pair of objects.

We have presented a framework that address this issue by: (a) contextually defining the set of objects in the context that may function as a landmark, and (b) sequencing the order in which spatial relations are considered using a cognitively motivated hierarchy of relations. Defining the set of objects in the scene that may function as a landmark reduces the number of object pairs that a spatial relation must be computed over. Sequencing the consideration of spatial relations means that in each context model only one relation needs to be checked and in some instances the agent need not compute some of the spatial relations, as it may have succeeded in generating a distinguishing locative using a relation earlier in the sequence.

A further advantage of our approach stems from the partitioning of the context into those objects that may function as a landmark and those that may not. As a result of this partitioning the algorithm avoids the issue of infinite recursion, as the partitioning of the context stops the algorithm from distinguishing a landmark using its trajector.

References

[Beun and Cremers, 1998] R.J. Beun and A. Cremers. Object reference in a shared domain of conversation. *Pragmatics and Cognition*, 6(1/2):121–152, 1998.

[Bryant *et al.*, 1992] D.J. Bryant, B. Tversky, and N. Franklin. Internal and external spatial frameworks representing described scenes. *Journal of Memory and Language*, 31:74–98, 1992.

[Carlson-Radvansky and Irwin, 1994] L.A. Carlson-Radvansky and D. Irwin. Reference frame activation during spatial term assignment. *Journal of Memory and Language*, 33:646–671, 1994.

[Carlson-Radvansky and Logan, 1997] L.A. Carlson-Radvansky and G.D. Logan. The influence of reference frame selection on spatial template construction. *Journal of Memory and Language*, 37:411–437, 1997.

[Clark and Wilkes-Gibbs, 1986] H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986.

[Dale and Haddock, 1991] R. Dale and N. Haddock. Generating referring expressions involving relations. In *Proceeding of the Fifth Conference of the European ACL*, pages 161–166, Berlin, 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.

[Edwards and Moulin, 1998] G. Edwards and B. Moulin. Towards the simulation of spatial mental images using the voronoï model. In P. Oliver and K.P. Gapp, editors, *Representation and processing of spatial expressions*, pages 163–184. Lawrence Erlbaum Associates., 1998.

[Fillmore, 1997] C. Fillmore. *Lecture on Deixis*. CSLI Publications, 1997.

[Gapp, 1995] K.P. Gapp. Angle, distance, shape, and their relationship to projective relations. In *Proceedings of the 17th Conference of the Cognitive Science Society*, 1995.

[Gardent, 2002] C Gardent. Generating minimal definite descriptions. In *Proceedings of the 40th International Confernce of the Association of Computational Linguistics (ACL-02)*, pages 96–103, 2002.

[Garrod *et al.*, 1999] S. Garrod, G. Ferrier, and S. Campbell. In and on: investigating the functional geometry of spatial prepositions. *Cognition*, 72:167–189, 1999.

[Gorniak and Roy, 2004] P. Gorniak and D. Roy. Grounded semantic composition for visual scenes. *Journal of Artificial Intelligence Research*, 21:429–470, 2004.

[Hajicová, 1993] E. Hajicová. Issues of sentence structure and discourse patterns. In *Theoretical and Computational Linguistics*, volume 2, Charles University, Prague, 1993.

[Herskovits, 1986] A Herskovits. *Language and spatial cognition: An interdisciplinary study of prepositions in English*. Studies in Natural Language Processing. Cambridge University Press, 1986.

[Horacek, 1997] H. Horacek. An algorithm for generating referential descriptions with flexible interfaces. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid, 1997.

[Jackendoff, 1983] R. Jackendoff. *Semantics and Cognition*. Current Studies in Linguistics. The MIT Press, 1983.

[Kelleher and Kruijff, 2005] J. Kelleher and G.J. Kruijff. A context-dependent model of proximity is physically situated environments. In *Proceedings of the 2nd ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, 2005.

[Kelleher and van Genabith, 2004] J. Kelleher and J. van Genabith. A false colouring real time visual salency algorithm for reference resolution in simulated 3d environments. *AI Review*, 21(3-4):253–267, 2004.

[Kelleher and van Genabith, 2005] J. Kelleher and J. van Genabith. In press: A computational model of the referential semantics of projective prepositions. In P. Saint-Dizier, editor, *Dimensions of the Syntax and Semantics of Prepositions*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2005.

[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: Reference and Presupposition in Language Generation and Interpretation*. CLSI Publications, Standford, 2002.

[Landau, 1996] B Landau. Multiple geometric representations of objects in language and language learners. In P Bloom, M. Peterson, L Nadel, and M. Garrett, editors, *Language and Space*, pages 317–363. MIT Press, Cambridge, 1996.

[Langacker, 1987] R.W. Langacker. *Foundations of Cognitive Grammar: Theoretical Prerequisites*, volume 1. Standford University Press, 1987.

[Logan, 1994] Gordon D. Logan. Spatial attention and the apprehension of spatial realtions. *Journal of Experimental Psychology: Human Perception and Performance*, 20:1015–1036, 1994.

[Logan, 1995] G.D. Logan. Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, 12:523–533, 1995.

[Talmy, 1983] L. Talmy. How language structures space. In H.L. Pick, editor, *Spatial orientation. Theory, research and application*, pages 225–282. Plenum Press, 1983.

[Treisman and Gormican, 1988] A. Treisman and S. Gormican. Feature analysis in early vision: Evidence from search assymetries. *Psychological Review*, 95:15–48, 1988.

[van Deemter, 2001] K. van Deemter. Generating referring expressions: Beyond the incremental algorithm. In *4th Int. Conf. on Computational Semantics (IWCS-4)*, Tilburg, 2001.

[van der Sluis and Krahmer, 2004] I van der Sluis and E Krahmer. The influence of target size and distance on the production of speech and gesture in multimodal referring expressions. In *Proceedings of International Conference on Spoken Language Processing (ICSLP04)*, 2004.

[Vandeloise, 1991] C. Vandeloise. *Spatial Prepositions: A Case Study From French*. The University of Chicago Press, 1991.

[Varges, 2004] S. Varges. Overgenerating referring expressions involving relations and booleans. In *Proceedings of the 3rd International Conference on Natural Language Generation*, University of Brighton, 2004.