Chapter 1

A COMPUTATIONAL MODEL OF THE REFERENTIAL SEMANTICS OF PROJECTIVE PREPOSITIONS

John Kelleher
Media Lab Europe,
Sugar House Lane,
Dublin 8, Ireland
john.kelleher@medialabeurope.org

Josef van Genabith
School of Computing,
Dublin City University,
Dublin 9, Ireland
josef@computing.dcu.ie

Abstract In this paper we present a framework for interpreting locative expressions containing the prepositions *in front of* and *behind*. These prepositions have different semantics in the viewer-centred and intrinsic frames of reference (Vandeloise, 1991). We define a model of their semantics in each frame of reference. The basis of these models is a novel parameterized continuum function that creates a 3-D spatial template. In the intrinsic frame of reference the origin used by the continuum function is assumed to be known a priori and object occlusion does not impact on the applicability rating of a point in the spatial template. In the viewer-centred frame the location of the spatial template's origin is dependent on the user's perception of the landmark at the time of the utterance and object occlusion is integrated into the model. Where there is an ambiguity with respect to the intended frame of reference, we define an algorithm for merging the spatial templates from the competing frames of reference, based on psycholinguistic observations in (Carlson-Radvansky, 1997).

Keywords: Frames of reference, spatial templates, potential field models, object occlusion.

1. Introduction

The focus of the Situated Language Interpreter (SLI) (Kelleher, 2003) project is to develop a natural language interpretive framework to underpin the development of natural language virtual reality (NLVR) systems. An NLVR system is a computer system that allows a user to interact with simulated 3-D environments through a natural language interface. People often use locative expressions to refer to objects in a visual environment. The term locative expression describes "an expression involving a locative prepositional phrase together with whatever the phrase modifies (noun, clause, etc.)" (Herskovits, 1986, pg. 7). In the simplest form of locative expression, a prepositional phrase has an adjectival role modifying a noun phrase and locates an object. Following (Langacker, 1987) we use the terms Landmark (LM) and Trajector (TR) to describe the noun phrases in a simple locative expression, see Example (1).

Example 1 . [The book]$_{\mathrm{TR}}$ on [the table]$_{\mathrm{LM}}$.

Section 1.2 describes the challenges in modelling projective prepositions.[1] Section 1.3 reviews previous computational work. In Section 1.4 we develop the SLI model for the interpretation of projective prepositions. This model combines novel approaches to: the computation of the spatial template's origin; the gradation of a preposition's applicability across its 3-D spatial template; object occlusion and frame of reference ambiguity resolution.

2. The Challenges

2.1 Cognitive Models of Projective Prepositions' Spatial Templates

Psycholinguistic research indicates that "people decide whether a relation applies by fitting a spatial template to the object's regions of acceptability for the relation in question" (Logan and Sadler, 1996, pg. 496). A spatial template is a representation of the regions of acceptability associated with a given preposition. It is centred on the landmark, and it identifies for each point in space the acceptability of the spatial relationship between the landmark and a trajector at that point. Using a spatial template, candidate trajectors can be assessed and rank-ordered by comparing the ratings of their locations in the spatial template. The candidate object whose location has the highest acceptability rating is then selected as the trajector.

Gapp's (1995) and Logan and Sadler's (1996) experiments reveal some of the parameters that define the constituency of a projective preposition's spatial template. There are three areas of acceptability within a spatial template: good, acceptable and bad; the areas within a spatial template are symmetrical around the search axis; the good and acceptable regions blend into one another; there

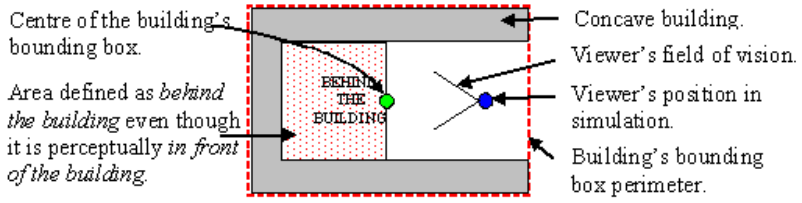


Figure 1.1. Bird's eye view of a concave building and viewer. Here, the use of the building's bounding box centre as the spatial template's origin results in an area being defined as *behind the building* even though it is perceptually *in front of the building*.

is a sharp boundary between the acceptable and bad regions; the acceptability of a projective preposition decreases linearly as the angular deviation from the search axis increases; acceptability approaches 0 as the angular deviation approaches 90°.

In order to interpret a projective preposition in an NLVR scenario, two other factors should be integrated into the spatial template model. Firstly, the distance between each of the candidate trajectors and the landmark should be accommodated to allow the model to distinguish between candidates with the same angular deviation. Secondly, in the viewer-centred frame of reference the spatial template's origin should be located based on the user's position at the time of the utterance. The spatial template's origin is the point in space that the spatial template search axis originates from and the point from which the distances of the trajectors from the landmark are computed. Consequently, the location of the spatial template origin impacts on the acceptability ascribed to a point in the spatial template. Many previous NLVR systems (Fuhr et al., 1998; Gapp, 1994; Olivier and Tsujii, 1994; Yamada, 1993) define this origin as the centroid of the landmark's bounding box.[2] While this approach works well for simple solid objects, applying it to more complex shapes can be problematic. For example, when applied to a concave object the centroid of the bounding box may be outside the object. This can result in paradoxical classification of regions around the landmark, see Figure 1.1.

2.2 Frame of Reference Ambiguity

Intrinsic to the use of a projective preposition (e.g., *in front of*, *behind*, *etc.*) is the definition of the direction the preposition describes. This directional constraint is referred to as the search axis. The orientation of the search axis associated with projective prepositions is dependent on the frame of reference being used. A frame of reference consists of six half-line axes with their origin at the landmark; these axes are sometimes referred to as the base axes (Her-

skovits, 1986). In English, these axes are usually labelled *front*, *back*, *right*, *left*, *up* and *down*. Significantly, a frame of reference's base axes are not fixed in space, but may be rotated depending on the perspective used. Consequently, a number of frames of reference are possible. In English,[3] there are three different types of frames of reference: absolute, intrinsic and viewer-centred (Levelt, 1996; Levinson, 1996; Carlson-Radvansky and Irwin, 1993). Following (Levinson, 1996), we distinguish between the frames of reference based on the cardinality of their relations.

Absolute (extrinsic, environmental, world based) frame of reference: this is a binary reference frame that locates a trajector relative to a landmark. The labelling of the landmark's axes is dependent on salient environmental features; e.g., gravity, magnetic poles, etc.

Intrinsic (object-centred, landmark-based) frame of reference: involves binary relations that locate a trajector relative to a landmark. The axes of the coordinate system are oriented around the landmark based on its canonical position.

Viewer-centred (egocentric, relative, deictic) frame of reference: presupposes a viewpoint with ternary relations that locate an object relative to a landmark. The axes of the landmark are oriented based on a "canonical encounter" (Clark, 1973) between an observer and the landmark.

One of the difficulties for interpreting a locative expression is that many spatial expressions are common between intrinsic and viewer-centred systems. The sharing of linguistic terms across frames of reference can cause misinterpretations based on frame of reference ambiguity. Levelt (1996) uses the term coordination failure to describe such misinterpretation. In some instances, the possibility of coordination failure can be avoided by the speaker using an explicit linguistic cue. For example, the use of the determiner *the* in a noun phrase which describes a spatial region X, such as *the X*, implies that an intrinsic frame of reference is being used. The region denoted by *on top of X* could apply to any frame of reference described; in contrast, the region denoted by *on the top of X* could only apply to X's intrinsic frame of reference (Landau and Munnich, 1998). However, explicit linguistic cues are exceptional. Consequently, if an NLVR system is going to interpret locative expressions, it must define an algorithm for handling the issue of frame of reference ambiguity.

3. Previous Computational Work

3.1 Computational Models of Spatial Templates

If a computational model is going to accommodate the gradation of applicability across a preposition's spatial template it must define the semantics of

the preposition as some sort of continuum function. A potential field model is one form of continuum measure that is widely used (Gapp, 1994; Olivier and Tsujii, 1994; Yamada, 1993). Using this approach, a model of a preposition's spatial template is constructed using a set of equations that for a given origin and point computes a value that represents the cost of accepting that point as the interpretation of the preposition. Another form of continuum model is proposed by (Mukerjee et al., 2000). In this model the continuum field is created by first defining the location of the field's global minimum. Following this, a set of concentric ellipses that use the global minimum as a fixed focus are created by varying the eccentricity of the ellipse and the position of the second focus. These concentric ellipses define the different regions of applicability within the model. Fuhr *et al.* (1998) propose a hybrid approach which uses the degree of overlap of an object with discretised regions as its measure.

Although these continuum models can distinguish between different locations within a spatial template, they are not ideal. Some of these models only work in 2-D (Mukerjee et al., 2000; Olivier and Tsujii, 1994; Yamada, 1993). (Fuhr et al., 1998) has problems distinguishing between the position of trajectors that are fully enclosed within a region. Most models (Fuhr et al., 1998; Gapp, 1994; Olivier and Tsujii, 1994; Yamada, 1993) use the centre of the landmark's bounding box as the spatial template's origin (this can lead to paradoxical interpretations, see Figure 1) and those that do not (Mukerjee et al., 2000) are dependent on locating the local minimum within the continuum field of a preposition which is problematic because the location of the local minimum varies from person to person. Furthermore, they all ignore the psycholinguistic evidence which indicates that, when frames of reference are dissociated, multiple frames of reference are activated and this multiple activation alters the constituency of the preposition's spatial template, see Section 1.4.4 (Carlson-Radvansky and Irwin, 1994) and (Carlson-Radvansky, 1997).

3.2 Computational Approaches to Frame of Reference Ambiguity

In Section 1.2.2 we noted that if an NLVR system is going to interpret locative expressions it must define an algorithm for handling frame of reference ambiguity. In general, previous NLVR systems have adopted one of four approaches to this issue: (1) situate the discourse in domains where only simple objects with no intrinsic reference frame are modelled, e.g., the SHRDLU system (Winograd, 1973); (2) assume a default frame of reference and force the user to adopt this for input, e.g., the Virtual Director system (Mukerjee et al., 2000) defaults to the intrinsic frame of reference if the landmark has one associated with it; (3) allow the user to switch between frames of reference if they use an explicit mark in the input, e.g., the CITYTOUR system (Andre

et al., 1988); (4) assume that the frame of reference is supplied to the system a priori, e.g., the Situated Artificial Communicator (Fuhr et al., 1998). All of these approaches, however, either restrict the domain of the discourse or impose restrictions on the user.

4. The SLI Model

In this section we describe the SLI semantic model for the projective prepositions *in front of* and *behind*. Vandeloise (1991) observes that the prepositions *devant/derriere* are bisemic, because the relationships they describe between the trajector and the landmark in the intrinsic frame of reference are different from the ones they describe in the viewer-centred frame of reference. He defines a topological semantics for these prepositions in the intrinsic frame of reference and argues that the primary factor in the viewer-centred usages is object occlusion. While we agree with Vandeloise in his assertion that the prepositions *in front of* and *behind* are bisemic, we do not claim that object occlusion is the primary factor in the semantics of *in front of* and *behind*; rather the approach we adopt is more aligned with that of Jackendoff and Landau, who argue that while object occlusion impacts of the semantics of these prepositions, it plays "a secondary role, possibly forming a preference rule system with the directional criteria" (1992, pg. 114). Following this, we define two spatial templates for *in front of* and *behind*: one for the intrinsic frame of reference which does not consider object occlusion, and one for the viewer-centred frame of reference which does.

4.1 Locating the Spatial Template's Origin

Most previous continuum models (Gapp, 1994; Olivier and Tsujii, 1994; Yamada, 1993) use the centre of the landmark's bounding box as the spatial template origin, irrespective of which frame of reference is being used. As noted in Section 1.2.1, for landmarks with complex geometries this can result in a paradoxical parsing of space (see Figure 1). In contrast with previous approaches, we define a different spatial template origin for each frame of reference.

In the intrinsic frame of reference, the spatial template origin is known to the system through a priori knowledge. The motivation for this is that if a person associates an intrinsic frame of reference with an object, they must have learned this intrinsic orientation based on prior experience with the object or objects of that type.

In contrast, the viewer-centred frame of reference may be applied to an object without prior knowledge of the object. From this, we argue that it is cognitively implausible to assume that a person uses a point in space whose location they do not know (i.e., the centre of the bounding box of an unfamiliar land-

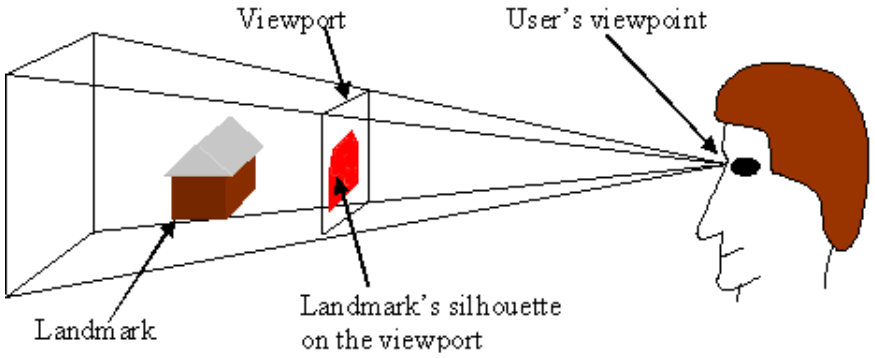


Figure 1.2. The relationships between the user's viewpoint, the viewport, a landmark's 3-D mesh and the landmark's silhouette on the viewport.

1 Resolve the landmark reference.

2 Calculate the landmark's silhouette on the viewport.

3 Calculate the point at the center of the landmark's silhouette on the vieport.

4 Calculate the point on the landmark's 3-D mesh that maps to the center of the landmark's silhouette on the viewport.

Figure 1.3. The SLI algorithm for locating the spatial template origin.

mark) as the origin for their spatial orientation. One of the insights guiding the SLI project is the grounding of the semantics of spatial language in visual perception. Following this, we argue that the most natural location for the origin of a projective preposition's spatial template in the viewer-centred frame or reference is the point on the landmark at the center of the landmark's silhouette as it is perceived by the user at the time of the utterance. In the terminolgy of 3-D graphics this point is defined as the point on the landmark's 3-D mesh that maps to the center of the landmark's silhouette on the viewport[4] at the time of the utterance. Figure 1.2 illustrates the relationships between the user's viewpoint, the viewport, a landmark's 3-D mesh and the landmark's silhouette on the viewport. Figure 1.3 lists the four step algorithm used to locate the point on the landmark's mesh that maps to the point at the center of its silhouette on the viewport.

Figure 1.4. The image on the left is the rendered visual context. The image on the right is the false colour rendering of the landmark.

The first step in the algorithm is to resolve the landmark reference. In the SLI system, the landmark reference is resolved using the SLI system's general algorithm for reference resolution (see (Kelleher, 2003) for details).

The second step in the algorithm is to calculate the landmark's silhouette on the viewport. We calculate the landmark's silhouette on the viewport by adapting a graphics technique called false colouring. False colouring was initially proposed by (Noser et al., 1995) as part of a navigation system for animated characters. Using a false colouring technique a system can extract information relating to the user's perception of the simulation at a given point in time. Implementing the technique involves assigning each object in the simulation a unique ID that differs from the normal colours used to render the object in the world; hence the term false colouring. An object's false colour is only used when rendering the object in the false colour rendering, and does not affect the renderings of the object seen by the user, which may be multi-coloured and fully textured. Once each object in the simulation has been assigned a false colour, whenever the system needs to examine what the user is currently seeing, a model of the user's view of the world using the false colours is rendered and the resulting image is scanned. By extracting the RGB[5] values found in the image, a list of objects in the image can be created. For the SLI system we adapted and extended the false colouring technique to create a dynamic real-time model of visual salience for 3-D rendered environments; the SLI system uses the resulting visual salience information to ground its reference resolution algorithm, see (Kelleher and van Genabit, 2004) for details. We calculate the silhouette of the landmark on the viewport by rendering the landmark by itself using its false colour (Figure 1.4 depicts the false colour silhouette of the house).

The next step is to calculate the coordinates of the center of the landmark's silhouette. To do this we first scan the false colour rendering of the landmark and record the maximum and minimum x and y coordinates of pixels rendered

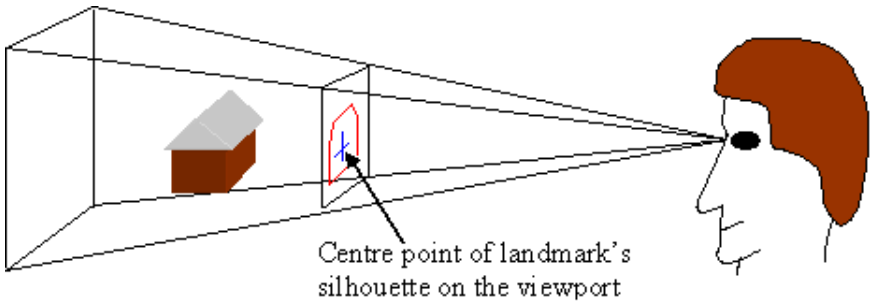


Figure 1.5. The relationships between the user's viewpoint, a landmark and the centre point of the landmark's silhouette on the viewport.

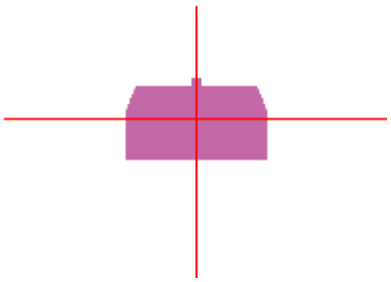


Figure 1.6. The point of intersection of the vertical and horizontal lines marks the location of the calculated center of the object's silhouette.

using the landmark's false colour. The coordinates of the center of the landmark's silhouette can then be calculated using Equation 1.1.

$$center(x,y) = (\frac{(x_{max} - x_{min})}{2}, \frac{(y_{max} - y_{min})}{2}) \tag{1.1}$$

Figure 1.5 illustrates the relationships between the user's viewpoint, a landmark and the centre point of the landmark's silhouette on the viewport. Figure 1.6 illustrates the point calculated as the the center of the landmark's silhouette on the viewport in our example.

The final step of the algorithm is to locate the point on the landmark at the center of its silhouette. We use a graphics technique called ray casting to locate this point. Ray casting can be functionally described as casting a ray (i.e., drawing an invisible line) from one point in a 3-D simulation in a certain direction, and then reporting back all the intersections with 3-D object meshes and the coordinates of these intersections. To locate the point on the

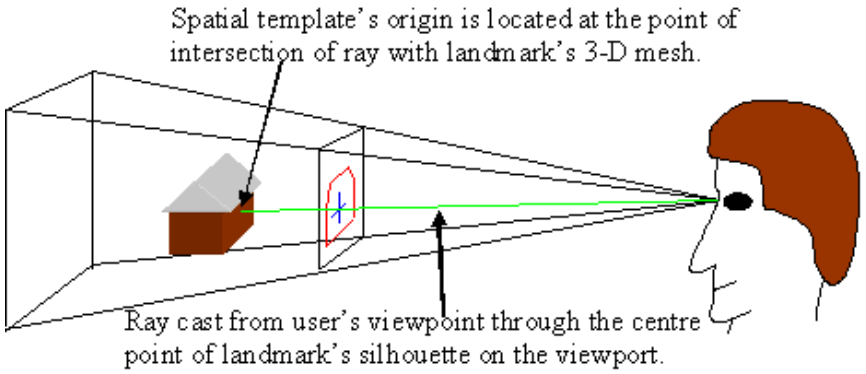


Figure 1.7. The preposition's spatial template's origin is located at the point of intersection of the ray cast from the user's viewpoint through the center point of the landmark's silhouette on the viewport and the landmark's 3-D mesh.

landmark's 3-D mesh that maps to the point at the center of its silhouette on the viewport we cast a ray from the user's viewpoint through the center point of the landmark's silhouette on the viewpoint and take the first point of intersection of this ray with the landmark's 3-D mesh as the origin of the preposition's spatial template. Figure 1.7 illustrates the casting of a ray from the user's viewpoint through the center point of the landmark's silhouette on the viewport and the intersection of this ray with the landmark's 3-D mesh. The preposition's spatial template origin is located at the point of intersection of the ray and the 3-D mesh.

4.2 Modelling the Gradation of a Preposition's Applicability

The two main factors that impact on the applicability of a projective preposition at a point relative to a landmark are: the angular deviation of the point from the canonical direction of the preposition's search axis and the distance of the point from the origin of the spatial template. Modelling these is further complicated by the requirement that the model should be scalable in order to accommodate different sizes of spatial configurations; e.g., the size of area described by *in front of the building* is larger than the area described by *in front of the door* (of the same building).

To model the directional constraint of a projective preposition, an algorithm for calculating the deviation of a point from a preposition's search axis must be defined. The first stage of this process is to assign a canonical direction to each of the prepositions. We assume that the search axes for the preposi-

tions in the intrinsic frame of reference are defined through prior knowledge. However, orienting the search axes in the viewer-centred frame of reference is dependent on the location of the user relative to the landmark at the time of the utterance. The vector originating from the spatial template's origin to the user's location describes the search axis for *in front of* in the viewer-centred frame of reference. One way of computing this vector is to convert the user's world coordinates into a set of coordinates in the local coordinate system centred on the spatial template's origin. The translated coordinates of the user's location then defines the search axis for *in front of* in the viewer-centred frame of reference. Rotating this vector by 180° gives us the search axis for *behind* in the viewer-centred frame of reference.

Having assigned a direction to each preposition, the next step in the modeling process is to devise a method for calculating the angular deviation of a candidate trajector from the search axes. θ, the angle between two vectors ν and ω can be calculated using Equation 1.2:

$$\theta = \cos^{-1}\left(\frac{\nu \bullet \omega}{|\nu|\,|\omega|}\right) \tag{1.2}$$

where $\nu = [x_1, x_2, x_3]$, $\omega = [y_1, y_2, y_3]$, $\nu \bullet \omega = (x_1y_1, x_2y_2, x_3y_3)$, $|\nu| = \sqrt{x_1^2 + x_2^2 + x_3^2}$, and $|\omega| = \sqrt{y_1^2 + y_2^2 + y_3^2}$. However, in order to use this equation to measure the angular deviation of a point from the search axis, the point must be converted into a vector that shares a common origin with the search axis. Applying this process to the coordinates of each of the candidate trajectors assigns each candidate an angular deviation from the preposition's canonical direction.

The distance applicability of a candidate trajector can be computed using the standard coordinate geometry distance formula for the distance between two points $[x_1, y_1, z_1]$and $[x_2, y_2, z_2]$, given in Equation 1.3:

$$Dist = \sqrt{\left((x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2\right)} \tag{1.3}$$

To create the topological spatial template for a projective preposition, the angular applicability ratings must be combined with the distance applicability ratings. This is done using the algorithm listed in Figure 1.8. This algorithm requires the definition of a maximum allowable angle of deviation β and a maximum distance γ. Following the findings of (Gapp, 1995; Logan and Sadler, 1996), the maximum angle of acceptability, β, should be set to 90°. To date, no ratio of the maximum distance γ to landmark size has been identified in the research literature. We propose that γ be set to the distance of the candidate trajector (simply satisfying the linguistic description of the trajector NP

Input A set of candidate trajectors $\{ct_1, ct_2, \ldots, ct_n\}$ each with an angular deviation α and distance rating δ; a maximum angle of deviation for the spatial template β; a maximum distance for the spatial template γ; and ρ the computed scaling factor.

Output A set of candidate trajectors $\{ct_1, ct_2, \ldots, ct_n\}$ each with an applicability rating λ within the preposition's spatial template.

1 **let** $\rho = 0$

2 **foreach** ct_i

(a) **if** $ct_i.\alpha \geq \beta$ **then** $ct_i.\alpha = 0$ **else** $ct_i.\alpha = 1 - (\frac{ct_i.\alpha}{\beta})$

(b) **if** $ct_i.\delta \geq \gamma$ **then** $ct_i.\delta = 0$ **else** $ct_i.\delta = 1 - (\frac{ct_i.\delta}{\gamma})$

(c) $ct_i.\lambda = ct_i.\alpha \times ct_i.\delta$

(d) **if** $ct_i.\lambda > \rho$ **then** $\rho = ct_i.\lambda$

3 **foreach** ct_i

(a) $ct_i.\lambda = \frac{ct_i.\lambda}{\rho}$

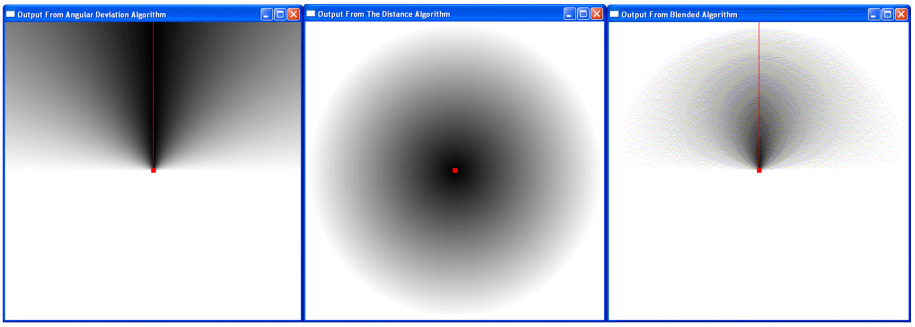
Figure 1.8. Algorithm for combining the angular deviation and distance scores.

and within the maximum allowable angular deviation) farthest from the spatial template origin. This means that the distance from the spatial template origin does not preclude a candidate trajector from being considered as the locative expression's referent; however, it does affect its rating within the process for selecting the referent. Moreover, by allowing the spatial template's maximum distance to vary depending on the context, the spatial template is scalable to different situations. This process results in each candidate trajector being assigned a rating within the spatial template. Figure 1.9 illustrates the continuum created using the algorithm listed in Figure 1.8.

4.3 Perceptual Cues in the Viewer-Centred Frame of Reference

At the beginning of Section 1.4 we proposed that the perceptual phenomenon of object occlusion impacts on the spatial templates of the prepositions *in front of* and *behind* in the viewer-centred frame of reference. We use two rules to integrate object occlusion with the continuum model:

1 If we are interpreting a locative containing the preposition *in front of* and there is a candidate trajector which partly occludes the landmark,



(a) (b) (c)

Figure 1.9. Diagrams illustrating 2-D slices of angle, distance, and amalgamated spatial template continuums. In these diagrams the darker the pixel the higher the applicability assigned to the point. The landmark is located at the centre of each image. (a) illustrates applicability gradation computed using Equation 1.2 (with search axis as the vertical axis and $\beta = 90°$). (b) illustrates the gradation computed using Equation 1.3 (γ set to a distance just inside the border of the image). (c) highlights the search axis used and illustrates the continuum created by merging the angular and distance applicabilities using the algorithm listed in Figure 1.8.

it is ascribed a maximum applicability rating within the viewer-centred spatial template irrespective of the rating based on the continuum model.

2 If we are interpreting a locative containing the preposition *behind* and there is a candidate trajector which is partly or wholly occluded by the landmark, it is ascribed a maximum applicability rating within the viewer-centred spatial template irrespective of the rating based on the continuum model.

If there is more than one candidate trajector with the maximum applicability rating we distinguish between them using a visual salience algorithm (based on size and location within the view volume). Moreover, if the visual salience is inconclusive (i.e., the differences in the saliences ascribed to the candidates is not sufficient to distinguish between them) we treat the locative as ambiguous and the system asks the user for clarification.

4.4 Resolving Frame of Reference Ambiguity

To date there have been several sets of psycholinguistic experiments on frames of reference selection in spatial language. Carlson-Radvansky and Irwin's (1994)(Carlson-Radvansky and Irwin, 1994) work revealed that when frames of reference are dissociated, more than one reference frame is initially activated and these active frames compete. Carslon-Radvansky and Logan

(1997) investigated the influence of frame of reference selection on the construction of a preposition's spatial template. Their findings indicate that, if there is a competition between reference frames, the construction of a preposition's spatial template in one frame of reference interferes with the construction of the spatial template in the other frame of reference. This interference between reference frames results in an amalgamated spatial template which extends over the areas covered by both of the individual spatial templates. Furthermore, the constituency of this amalgamated spatial template differs from a spatial template constructed when there is no competition: there is no good region; the acceptable regions are bigger and the bad regions are smaller; the regions that are rated as acceptable in both the viewer-centred and intrinsic frame of reference have a higher acceptability rating in the amalgamated frame of reference than those in the regions which are acceptable in only one of the individual spatial templates. Carlson-Radvansky and Logan (1997) concluded that when frames of reference are dissociated, the spatial templates constructed for each of the competing reference frames should be amalgamated using a weighting that reflects the bias towards a particular reference frame for a given preposition. With respect to the bias in this competition, Carslon-Radvansky and Irwin (1993) showed the where a preposition is canonically aligned with the vertical axis, the absolute frame of reference dominates its use, and findings in (Taylor et al., 2000) indicate that, in contrast with the vertically aligned prepositions, there is a slight bias toward the intrinsic frame of reference for the horizontally aligned prepositions. Based on these psycholinguistic findings we present an algorithm (Figure 1.10) for resolving frame of reference ambiguity. Figure 1.11 illustrates the template resulting from this process.

The weighting of 2:1 towards the viewer-centred frame of reference for the vertically aligned prepositions is derived from an analysis of Carlson-Radvansky and Irwin's (1993) results. Although the work of Taylor *et al.* (2000) does not quantify the bias toward the intrinsic frame of reference for the horizontally aligned prepositions, a ratio of 1.1:1 in favour of the intrinsic frame of reference for horizontally aligned prepositions is assumed. While there is a marginal difference across this ratio, it is sufficient to prefer the intrinsic frame of reference in the event of a tie.

4.5 Selecting the Referent

The semantic model described in the preceding sections allows us to model the applicability of a preposition across a region. Using this model, a projective locative expression can be resolved by selecting a referent from the set of candidate trajectors based on their location within a region and object occlusion effects. However, the abstraction used to represent the candidate trajectors impacts on this process as it affects the applicability ratings assigned to them

1 **if** the frames of reference in a scene are dissociated **then**

(a) construct a spatial template for the preposition in both frames of reference

(b) **if** preposition = *above* or *below* **then**

i multiply the ratings in the viewer-centred spatial template by 2

(c) **elseif** preposition = *in front of* or *behind* **then**

i multiply the ratings in the intrinsic spatial template by 1.1

(d) assign each point an overall applicability equal to the sum of its applicability ratings in both spatial templates

(e) select the candidate with the highest overall applicability as the referent.

Figure 1.10. The SLI frame of reference competition resolution algorithm.

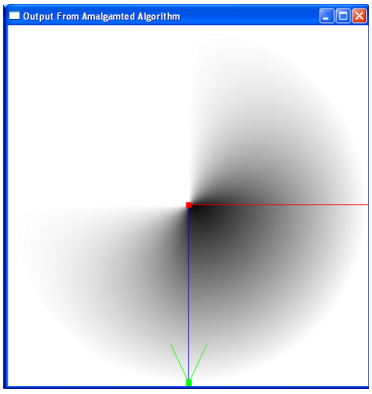


Figure 1.11. Bird's eye view of a 2-D slice of the spatial template for *in front of* created using the algorithm listed in Figure 1.10. The landmark is located at the centre, the viewer at the bottom, the search vector used to create the intrinsic spatial template is illustrated by the line going from the landmark to the right of the image, the search vector used to create the viewer-centred spatial template is illustrated by the line going from the landmark to the viewer's location.

by the model. Most previous systems have used the centre of the candidate trajector's bounding box. There are, however, problems with this abstraction for elongated objects. To account for this, we use the vertex in the candidate's 3-D mesh which has the highest applicability rating to represent each candidate. This ensures that the candidate with a point at the highest applicability will be selected as the referent.

5. Conclusions

In summary, the advantages of the SLI interpretive framework described in this paper are: it avoids the problems associated with using the landmark's bounding box centre as the spatial template origin in the viewer-centred frame of reference; it offers a new model for the gradation of the preposition's applicability across a 3-D volume; it is scalable and consequently it is able to accommodate different size landmarks; it accommodates the impact of frame of reference ambiguity on the construction of a spatial template model in terms of amalgamated spatial template models and it accommodates the perceptual cue of object occlusion.

Notes

1. For a model-theoretic analysis of locative expressions see (Zwarts and Winter, 2000).
2. An object's bounding box is the minimal rectangle that encompasses the geometry of the object.
3. Although the use of a tripartite system is common in European languages, this is not universal with many languages taking different approaches, see (Levinson, 1996) and (Levelt, 1996)
4. A viewport is the rectangular area of the display window. It can be conceptualised as a window onto the 3-D simulation.
5. RGB: Red, green and blue; the primary colours that are mixed to display the color of picels on a computer monitor.

References

Andre, E., Herzog, G., and Rist, T. (1988). On the simultaneouse interpretation of real world image sequences and their natural language description: The system soccer. In *In Proceedings of the 8th European Conference on Artificial Intelligence (ECAI-88)*, pages 449–454. Pitmann.

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

Carlson-Radvansky, L. and Irwin, D. (1993). Frames of reference in vision and language: Where is above? *Cognition*, 46:223–224.

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

Clark, H. (1973). Space, time, semantics, and the child. In Moore, T., editor, *Cognitive development and the acquisition of language*, pages 65–110. Academic Press, New York.

Fuhr, T., Socher, G., Scheering, C., and Sagerer, G. (1998). A three-dimensional spatial model for the interpretation of image data. In Olivier, P. and Gapp, K., editors, *Representation and Processing of Spatial Expressions*, pages 103–118. Lawrence Erlbaum Associates.

Gapp, K. (1994). Basic meanings of spatial relations: Computation and evaluation in 3d space. In *National Conference on Artifical Intelligence (AAAI-94)*, pages 1393–1398.

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

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

Jackendoff, R. and Landau, B. (1992). Spatial language and spatial cognition. In Jackendoff, R., editor, *Languages of the Mind*, pages 99–125. MIT Press.

Kelleher, J. (2003). *A Perceptually Based Computational Framework for the Interpretation of Spatial Language*. PhD thesis, Dublin City University.

Kelleher, J. and van Genabit, J. (2004). A false colouring real time visual salency algorithm for reference resolution in simulated 3d environments. *AI Review (Forthcoming)*.

Landau, B. and Munnich, E. (1998). The representation of space and spatial language: Challenges for cognitive science. In Olivier, P. and Gapp, K., editors, *Represetation and Processing of Spatial Expressions*, pages 262–272. Lawrence Erlbaum Associates.

Langacker, R. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*, volume 1. Standford University Press.

Levelt, W. (1996). Perspective taking and ellipsis in spatial descriptions. In Bloom, P.and Peterson, M., Nadell, L., and Garrett, M., editors, *Language and Space*, pages 77–108. MIT Press.

Levinson, S. (1996). Frame of reference and molyneux's question: Crosslinguistic evidence. In Bloom, P.and Peterson, M., Nadell, L., and Garrett, M., editors, *Language and Space*, pages 109–170. MIT Press.

Logan, G. and Sadler, D. (1996). A computational analysis of the apprehension of spatial relations. In Bloom, P.and Peterson, M., Nadell, L., and Garrett, M., editors, *Language and Space*, pages 493–529. MIT Press.

Mukerjee, A., Gupta, K., Nauityal, S., Mukesh, P., Singh, M., and Mishra, N. (2000). Conceptual description of visual scenes from linguistic models. *Journal of Image and Vision Computing*, 18.

Noser, H., Renault, O., Thalmann, D., and Magnenat-Thalmann, N. (1995). Navigation for digital actors based on synthetic vision, memory and learning. *Computer Graphics*, 19(1):7–9.

Olivier, P. and Tsujii, J. (1994). Quantitative perceptual representation of prepositional semantics. *Artificial Intelligence Review*, 8(147-158).

Taylor, H., Naylor, S., Faust, R., and Holcomb, P. (2000). Could you hand me those keys on the right? disentangling spatial reference frames using different methodologies. *Spatial Cognition and Computation*, 1(14):381–397.

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

Winograd, T. (1973). A procedural model of language understanding. In Schank, R. and Colby, K., editors, *Computer Models of Thought and Language*, pages 152–186. W. H. Freeman and Company.

Yamada, A. (1993). *Studies in Spatial Descriptions Understanding based on Geometric Constraints Satisfaction*. PhD thesis, University of Kyoto.

Zwarts, J. and Winter, Y. (2000). Vector space semantics: A model-thoeretic analysis of locative prepositions. *Journal of Logic, Language and Information*, 9(2):169–211.