To Appear in: The Third International Conference of Industrial and Engineering Applications of Artificial Intelligence and Expert Systems
Charleston, SC July 16-18, 1990

Using Focus for Generating Felicitous Locative Expressions*

Susan M. Haller
Syed. S. Ali
Department of Computer Science
State University of New York at Buffalo
226 Bell Hall
Buffalo, NY 14260
(716) 636-3197

Abstract

We are concerned with using discourse focus to generate felicitous natural language responses to "Where is"-type queries by a user with respect to a map. In ordinary language this is typically achieved by using a locative expression whose syntax involves using a preposition (such as *in* or *at*) and its object (which serves as a reference point). The selection of an appropriate reference point is important when generating such locative expressions. We attempt to use discourse focus to model the user's mental body position in the selection of an appropriate reference point. This enables the user to use body-oriented inference strategies associated with small scale space to make better sense of the overall spatial organization of the geographic entities and the large scale space of which these geographic entities are a part.

1 Introduction

We are concerned with using discourse focus to generate felicitous natural language responses to "Where is"-type queries by a user with respect to a map. This work was conducted by extending an existing generation grammar in a multi-modality interface system called CUBRICON (the CUBRc Intelligent CONversationalist) [9, 10]. This system accepts natural language input that can include pointing to a map display via a mouse. The output of the system is natural language, most often in conjunction with pointing and blinking gestures on the map [2].

We follow Herskovits who defines a *locative expression* as any spatial expression involving a preposition, its object, and whatever the prepositional phrase modifies (noun, clause, etc.) [4]. The prototypical use of the expression is to inform the addressee of the location of the modified noun or clause. Herskovits refers to this located object as the *figure*. The object of the preposition, called the *ground*, is the reference object used to specify a spatial relationship between itself and the figure. It is assumed that the location of the ground object is either familiar to, or easily discovered by the addressee. In the locative expression, *the house on the hill*, *house* is the figure object whose location is constrained by the ground object *hill* through the spatial relationship *on*. In the following discussion, we use the term *landmark* to refer to any geographic entity deemed a suitable candidate for use as a ground object. We use the terms *reference point*, *reference object* and *ground object* interchangeably to mean a landmark that has been selected for use as a ground object.

We restrict our problem by assuming that landmarks should be stationary geographic locations of enough size and importance that are prominent on the map. In the CUBRICON system, the user can ask for information on the location of any entity in the knowledge base that has a name. Such entities are air bases and factories located throughout East-West Germany. Since they are of primary interest, their iconic representations are almost always visible on the map. Information associating these entities with appropriate icons, and other military and geographic information on the entities is included in a knowledge base represented in the Semantic Network Processing System (SNePS)[12]. A semantic network is

*The research discussed in this report represents part of Research Initiative #2, "Languages of Spatial Relations," of the National Center for Geographic Information and Analysis, and was supported in part by a grant from the National Science Foundation (SES-88-10917) and also in part by the Defense Advanced Research Agency under Contract No. F30603-87-C-0136.

a data representation usually consisting of nodes and directed arcs. SNePS is a fully intensional propositional semantic network in which nodes are used to represent all objects of thought such as rules, propositions and individual entities.

The natural language understanding and generation components are implemented in a Generalized ATN grammar [11]. The grammar is used to analyze natural language input and build SNePS representations as well as interpreting SNePS representations when generating English responses. Our work has resulted in an extension to this generation grammar for the multi-modality interface system that selects a reference point and expresses the location of an object with respect to the selected reference point. In doing so, the system attends to what is prominent on the map, and what has been used recently to orient the addressee.

2 The Task

Previous to our work, when the user asked where such an entity was (via a "Where is" query), if the site was viewable or could be made viewable on the map, the system responded by blinking, highlighting, and labeling its iconic representation. In addition, a canned phrase was printed and spoken:

`The` < *name and class of object* > `is located here.`

For example, if the user asked for the location of the Nuernburg air base, the base was gestured to as the system said:

```
The Nuernburg air base is located here.
```

If the enitity asked for was outside the region represented by the map system, a message was issued to that effect. Then information from the knowledge base corresponding to the site's location in longitude and latitude was handed to the natural language generator:

`Its location is` < *latitude and longitude of site* >.

For example, if the user asked for the Stargard air base, (which is in Poland), the system responded with:

```
The Stargard air base is outside the region of
responsibility.
Its location is 51.300 N latitude, 16.900 E
longitude.
```

Our goal was to generate a felicitous locative phrase that expresses a site's location relative to some landmark that is viewable on the map. We attempted to express the location of the figure in terms of the map whether the figure is viewable or not. Generating such a locative expression involves three distinct steps: (1) choosing a suitable landmark for the ground, (2) building a knowledge representation to express the relationship of the figure to the ground, and (3) generating natural language to express that relationship. In this paper we are concerned with a strategy developed for performing the first step of this process.

Obvious criteria for reference point selection are that the selected point should be highly visible, important, and near the entity being located. What is missing from this formulation is any consideration of reference points just used and where these points have mentally placed the user. For example, figure 1 is a map of the Middle Atlantic United States that has the cities of Buffalo, Pittsburgh, Philadelphia and New York as major landmarks, where all other smaller cities and points of interest are to be located on the map with respect to one of these four cities. If one asks, "Where is Altoona?", an appropriate reply might well be "100 miles east of Pittsburgh." If the next query is "Where is Erie?", we feel that "130 miles north of Pittsburgh" would be more appropriate than "90 miles southwest of Buffalo". This is because the first reply takes into account that the user's current orientation to the map is from Pittsburgh. Though the second reply (using Buffalo) uses a reference point closer to the figure, it does not take the previous orientation of the user into consideration.

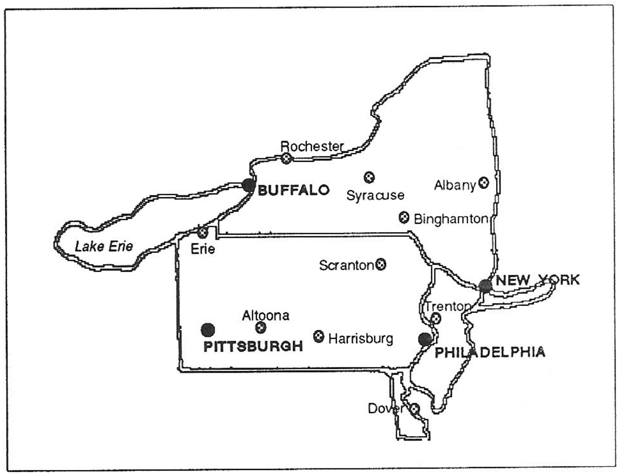

Figure 1: Map of Middle Alantic US

3 Reference Point Selection

As previously mentioned, we assume that appropriate candidate ground objects are geographic entities of the East-West Germany region of enough size and importance to be viewable on the map and familiar to the user. With this in mind, seven major cities located throughout the East-West Germany region were selected and represented in the knowledge base as landmarks of the region. Additionally, the East-West Germany border was added to the knowledge base for use as a possible reference point. It was chosen for its political relevance and because it is visually prominent on the map of East-West Germany and on the maps of two subregions that can be displayed on request. A portion of the East-West Germany border is usually in view and serves as a good landmark when no cities are in view.

3.1 Principles of Selection

The following principles are employed to select a ground object:

1. The ground object selected must be visible without expanding an existing window or creating a new one.

2. Landmarks previously used as ground objects are preferable to new ones if they are reasonably near the figure object.

3. If a new landmark must be used as ground and there is more than one, select the landmark nearest the figure object.

CUBRICON is capable of expanding or changing maps when a request is made to the system for an entity that is viewable but perhaps not currently displayed. It is our opinion that reference points are objects that are available at no effort to help inform the user about the location of the figure. Hence, the algorithm we use carefully checks to ensure that candidate ground objects are viewable within the boundaries of the existing window with the figure. This corresponds to the first principle.

If there are no appropriate landmarks in focus the third heuristic embodies the bottom line choice strategy among available landmarks. Currently, all landmarks in the window with the figure are viewed as equally good candidates for the ground, so unless one of them has been recently used as a reference point, we use the landmark closest to the figure.

3.2 Focus List

The natural language component of the CUBRICON maintains a list of *discourse items* on a *focus list* [13, 7]. This list is a discourse model and consists of discourse items (concepts) evoked by a continuing dailogue. Associated with discourse items are degrees of activation corresponding to item salience at a particular point in a dialogue. Salience of a discourse item is determined by three characteristics; its syntactic/semantic role, its syntactic position (whether it occurs in a main or embedded clause), and its referential distance (how long ago, relative to the current dialogue, it was last used) [1]. When an item's activation falls below a minimal threshold it is removed from the focus list.

The intended use of the focus list is to decode and encode pronoun referents, however we put it to a further use. Landmarks previously used as ground objects are viewed as particularly salient because they are included in the current discourse model. They are put onto the focus list with a degree of activation based on their syntactic category.

3.3 Example

We illustrate the use of the focus list with the following exchange which results in the map of figure 2 being displayed.

```
User:   Display the Fulda Gap region.
```

(A map of the region is displayed)

```
User:   Where is the Nuernberg airbase?
System: The map on the color graphics screen
        is being expanded to include the
        Nuernberg air base.

        Its location is east of Nuernberg city.
```

At this point, the city of Nuernberg has been added to the focus list.

When a landmark is used as a ground object, it is added to the focus list with sufficiently low priority so as to be removed from the list within a few exchanges. However, for a short time the landmark is available for use, and if it is in the same window as the current figure object, and if it is "near"[1] the figure, it will be selected once again as the ground object. The intention here is that the system does not arbitrarily pick the closest ground object, but

[1] The intention is to use perceptually valid metrics.

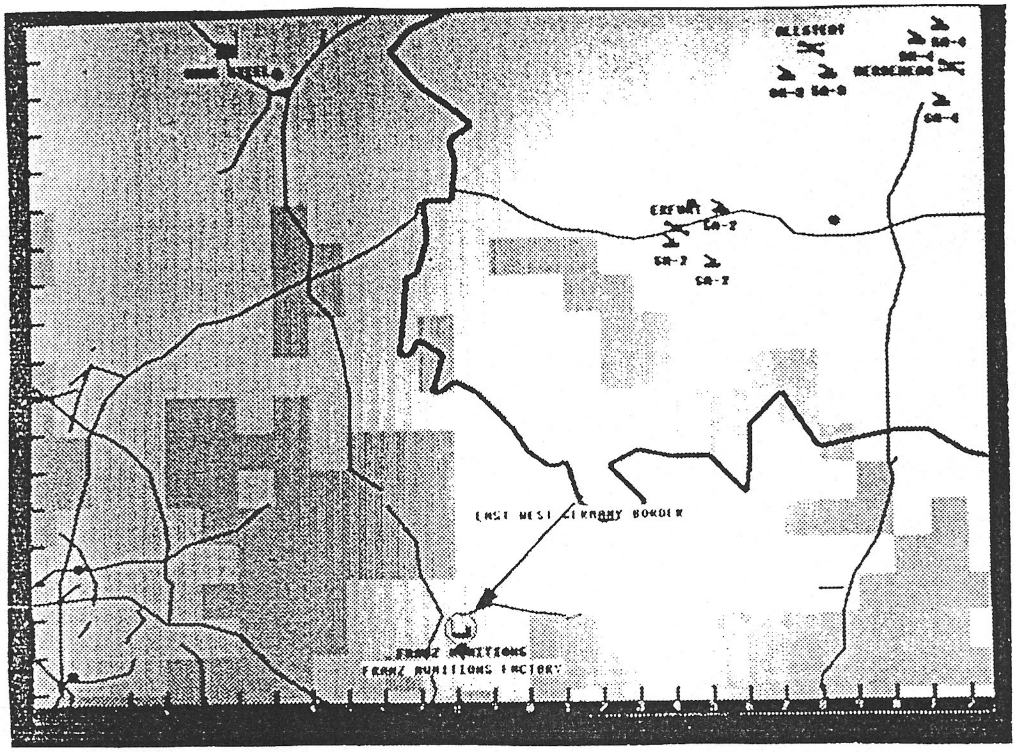

Figure 2: A CUBRICON window displaying the Fulda Gap region

rather, maintains a model of the discourse with the user and attempts to frame the response in terms of a subarea of the map currently under discussion. To illustrate the use of the focus list, we might continue this exchange as follows:

```
User:   Where is the Hans steel plant?
System: Its location is 8 miles west of
        Kassel city.
User:   Where is the Erfurt airbase?
System: Its location is 70 miles east of
        Kassel city.
User:   Where is the Merseberg airbase?
System: Its location is 20 miles west of
        Leipzig city.
```

Kassel city is chosen as the next reference point despite the fact that Nuernberg city is more highly activated by virtue of being used in the previous exchange. Nuernberg city (as displayed on the map of figure 2) did not meet the "nearness" constraint, resulting in the choice of a new reference point. The new reference point, Kassel city, is then also added to the focus list, resulting in its use in the next exchange to locate the Erfurt airbase. Note that it is not used to locate the Merseberg airbase because it fails the "nearness" constraint.

As previously noted, some figure objects are not viewable. In this case, the algorithm attempts to find a region (one of the East-West Germany, Fulda Gap or the Leipzig regions) that has been previously displayed and, hence, is on the focus list. Unlike other landmarks, regions added to the focus list remain there for a long time, since they are added in the subject syntactic-semantic category as a direct result of the "display < *region* >" request by the user. We consider such areas to be valid landmarks, since they can be, and have been, asked for by name and are therefore familiar. Hence, if the algorithm succeeds, the region itself is used as the ground to orient a figure that cannot be viewed. For example if we continue the previous dialogue, as follows:

```
User:   Where is the Stargard airbase?
System: Its location is 180 miles east of
        the Fulda Gap region.
```

the system uses the Fulda Gap region, because it was previously put on the focus list by the initial display request.

4 Discussion

The development of our extension has clarified several research issues related to generating natural language consistent with, and appropriate to, visual information displayed on a map. Mark and Frank [8] discuss two distinct scales of perceptual space. *Small scale* space is defined to be that area of space that people experience in the area immediately surrounding the body. Hence, an object present in small scale space is viewed from a single perspective, that of the individual, and oriented with respect to him. In contrast, *large scale* or geographic space is space that cannot be cognitively structured from a single, bodily perspective [5, 8]. Knowledge of large scale space is usually assembled internally by integrating observations over time. Kuipers [6] defines knowledge of a given large scale environment as a *cognitive map*.

Maps create a curious state of affairs since they represent large scale space in terms of small scale space. This representation allows people to apply the orientation strategies they are familiar with from direct experience in small scale space to build the cognitive map. For example, our extension locates geographic entities by specifying them as "east" or "west" of other objects on the map. In the context of a map, these terms are similar to terms associated exclusively with small scale space that orient objects with respect to the body, such as "right" or "left."

Hart and Moore [3] distinguish *internal* and *external* representations of space and note that we can only infer internal, cognitive representations of space from external representations (of which maps are a prime example). As long as geographic objects are located with respect to other geographic objects on the map, the representation in terms of small scale space helps the map-user to make spatial sense of the large scale space that is difficult to comprehend and model. To make this point clear, we note that an alternative is to use the map itself (an object in small scale space) to locate geographic entities, applying relations such as "at the top" or "at the bottom" of the map. However, these orientations give the user no insight into the spatial organization of the large scale, geographic space that the map model represents.

Our work has emphasized another aspect of small scale space that is not captured by the map at all, the viewer-centered or single perspective aspect. We feel that the use of a primitive discourse model (the focus list) to record and reuse ground objects is tantamount to planting the map-user somewhere on the face of the map and directing his attention to objects from that vantage point. The focus list logs what has been discussed, hence, reference points (ground objects) on the focus list represent where the user has mentally been placed for viewing or locating an object.

Maintaining the point of view of the user by reusing reference points on the map helps him to build an appropriate internal, cognitive map of the large scale space represented. A user at the same vantage point is more able to use the body-oriented spatial inference rules he is familiar with from small scale space to infer an internal representation from the external representation provided by the map. When we describe a room to someone, we locate the objects in it by taking account of where the body is. We may say things like:

> *As you walk through the door, the bookcase is on your right.*

If our goal is to provide an accurate account of where everything is in the room, it is helpful to keep the same viewpoint.

> *As you walk through the door, the bookcase is on your right. To the left is the desk, in front of you is the closet door.*

Not only does this organize the spatial information with respect to the viewer's body, but it allows him to use knowledge about how objects oriented at different positions to his body spatially relate to each other.

Similarly, we feel that taking the user's mental body position into account (as we have by placing him with a reference point) enhances his understanding of the overall organization of the large scale space under discussion. In our previous example, if the system tells the user that the Hans steel plant is 8 miles west of Kassel, it has mentally placed him at Kassel. If the system later informs the user that the Erfurt air base is 70 miles east of Kassel, this gives the him significantly more information to build a cognitive map of the subarea than if the system had told him that the Erfurt air base is 50 miles west of Leipzig. The reason is that the use of discourse focus to select the reference point has taken into account the user's mental position, hence, both air bases are oriented with respect to a user "in Kassel." As a result, the user can use his knowledge of small scale space to orient the air bases with respect to each other and enhance his understanding of the large scale space this portion of the map represents.

5 Summary

We feel that attending to discourse focus to select reference points, when generating locative expressions, attends to the user's mental body position. Thus, he is able to use body-oriented inference strategies associated with small scale space to make better sense of the overall spatial organization of the geographic entities and the large scale space these geographic entities are a part of. We feel that any natural language interface to systems that use locative expressions to orient the user must take this into account.

6 Acknowledgments

Our thanks to Joseph M. Piazza for his help in the preparation of this paper.

References

[1] Givon, T. *Topic Continuity in Discourse: A Quantitative Cross-Language Study.* John Benjamins, Amsterdam, 1983, ch. Introduction, pp. 1–41. T. Givon (ed.).

[2] Haller, S. M. Spatial relations and locative phrase generation in a map context. Tech. rep., State University of New York at Buffalo, 1989.

[3] Hart, R. A., and Moore, G. T. The development of spatial cognition: A review. In *Image and Environment*, R. M. Downs and D. Stea, Eds. Aldine Publishing Company, Chicago, 1973, pp. 246–288.

[4] Herskovits, A. Semantics and pragmatics of locative expressions. *Cognitive Science 9* (1985), 341–378.

[5] Kuipers, B. Modeling spatial knowledge. *Cognitive Science 2* (1978), 129–153.

[6] Kuipers, B. The cognitive map: Could it have been any other way? In *Spatial Orientation Theory, Research, and Application*, H. L. Pick and L. Acredolo, Eds. Plenum Press, New York, 1983, pp. 345–359.

[7] Li, N. Pronoun resolution in sneps. Tech. rep., State University of New York at Buffalo, 1986. SNeRG Technical Note 18.

[8] Mark, D. M., and Frank, A. U. Concepts of space and spatial knowledge. In *Proceedings, Auto Carto 9* (1989).

[9] Neal, J. G., Shapiro, S. C., Thielman, C. Y., Lammens, J. M., Funke, D. J., Byoun, J. S., Dobes, Z., Glanowski, S., Summers, M. S., Gucwa, J. R., and Paul, R. Final report for the intelligent multi-media interfaces project. Tech. rep., Calspan-UB Research Center, 1990.

[10] Neal, J. G., Thielman, C., Z.Dobes, Haller, S., Glanowski, S., and Shapiro, S. Cubricon: A multi-modal user interface. In *Proceedings, GIS/LIS Conference* (1989). To appear.

[11] Shapiro, S. C. Generalized augmented transition network grammars for generation from semantic networks. *American Journal of Computational Linguistics 8* (1982), 12–25.

[12] Shapiro, S. C., and Rapaport, W. J. Sneps considered as a fully intensional propositional semantic network. In *The Knowledge Frontier, Essays in the Representation of Knowledge*, N. Cercone and G. McCalla, Eds. Springer-Verlag, New York, 1987, pp. 263–315.

[13] Sidner, C. L. Focusing in the comprehension of definite anaphora. In *Computational Models of Discourse*, M. Brady and R. C. Berwick, Eds. MIT Press, Cambridge, 1983.