

# Directing the Generation of Living Space Descriptions

**Penelope SIBUN**

Department of Computer & Information Science  
University of Massachusetts  
Amherst, MA 01003 USA

**Alison K. HUETTNER**

Department of Comparative Literature & Languages  
Hofstra University  
Hempstead, NY 11550 USA

**David D. MCDONALD**

Brattle Research Corporation  
55 Wheeler Street  
Boston, MA 02138 USA

## Abstract

We have developed a computational model of the process of describing the layout of an apartment or house, a much-studied discourse task first characterized linguistically by Linde (1974). The model is embodied in a program, APT, that can reproduce segments of actual tape-recorded descriptions, using organizational and discourse strategies derived through analysis of our corpus.

## 1. Introduction

At this point in research on natural language generation, it is important to select problems that will clarify what is at issue in the larger phenomena under study, while at the same time being small enough to yield principled results in a reasonable amount of time. To build on what the field has already accomplished, the problem must involve the generation of motivated discourses--rather than isolated test sentences--and should be based on a corpus of real text. Furthermore, since a computational treatment of a generation problem should include a fully programmed underlying conceptual model to facilitate experiments, and since the representation used in that model will invariably play a crucial role in any theory, part of the research is building the model and designing the representation. This means that to be tractable the problem should not require expert knowledge or be overly large.

Support for this work was provided in part by the Defense Advanced Research Projects Agency under contract number N00014-87-K0238, and by Rome Air Development Center under contract number AF30602-81-C-0169, task number I74398, both at the University of Massachusetts.

Living space descriptions fit these demands neatly. They are single-speaker monologues, allowing us to ignore issues of turn-taking strategies or interpreting an interlocutor's intentions. The task is something everyone seems to be able to do, but it is not such an everyday occurrence that it has become formulaic: it is likely that people are actively constructing what they are saying. Affective and abstract information is minimal and, to a first approximation, can be safely factored out of a corpus, reducing the potential complexity of the conceptual model in living space descriptions. This simplicity has allowed us to concentrate on our primary concerns:

- (a) understanding the relationship between the organization of a conceptual model and descriptive strategies,
- (b) determining the influence of these strategies on the discourse structure of a text, and
- (c) taking an initial look at issues in lexical choice in a familiar domain.

At the time this paper is written, we have finished the first phase of our research. We have collected and carefully transcribed a corpus of seven different people's descriptions of the same, single-story house (the residence of one of the authors). A program model of this house, as these people appear to view it, has been developed, along with a set of strategies and meta-strategies for generating some of the living space descriptions that emerged from our analysis of the corpus. This paper presents our representation, some of the strategies we have identified and their application in mimicking<sup>1</sup> a segment from our corpus, and our treatment of some linguistic issues in choosing words and constructions.

## 2. Background and approach

The seminal work in living space descriptions is Charlotte Linde's 1974 dissertation. Linde's data consisted of 72 descriptions of apartments elicited in interviews on the quality of life in New York City. She found that the great majority of speakers organize their descriptions as an imaginary "tour" of the apartment. The spatial relationships among the rooms can be expressed by describing how one might make one's way from each room to the next. Such a tour is of course constrained by the position of the actual routes through the apartment. Linde proposed a model in terms of a phrase structure network in which the terminal nodes were rooms and vectors of various categories.

Veronika Ullmer-Ehrich (1982) extended the discussion to descriptions of individual dormitory rooms, again embedded in longer interviews. The descriptions she collected focused on the spatial relationships among the rooms' furnishings. She found, as one might expect, that imaginary movement was less usual here, since the speaker can typically "see" everything to be mentioned from a single point of view. As in Linde's apartment descriptions, physical proximity has a strong influence on the order in which objects are introduced; Ullmer-Ehrich refers to the result as an imaginary "gaze tour" around the walls. (Our own informants tended to give the contents of the rooms as well as their spatial relations to each other, letting us see both kinds of strategies in action.)

Linde's and Ullmer-Ehrich's treatments were descriptive. Ours attempts to model the motivations behind the texts. Our aim is to construct a computer program that can reproduce our data and, further, produce variations on it. If simple variations on the parameters of our model still produce realistic texts, then we will have a basis for claiming that it could be a candidate explanatory model of the processing that underlies human behavior in this task.

Our implementation, APT, is composed of a **knowledge base** consisting of interconnected first-class objects that reconstruct the living space, **strategies** which traverse the knowledge base constructing descriptions, **meta-strategies** which choose among the strategies each time a new strategy is needed, and **mapping rules** between APT's knowledge structures and the realization component's knowledge structures.

<sup>1</sup> Living space descriptions are a small enough subject area that it is worth trying to develop a treatment with enough articulations in its mechanisms to potentially account for every detail of what people actually say—hence "mimic". There may well be a vast amount of arbitrariness in the decisions people make; but the pressure to explain the fine structure of their utterances, not just to gloss over it by producing something "comparable" but more regular, should lead to stronger, more interesting theories.

## 3. The representation

Practically any familiar representation language that one might "take off the shelf" to use in modeling the information needed for the description of a house will be technically deficient in several ways when one comes to use it as a source for generation: it may not supply first-class objects for the information units a natural language can reference; its taxonomic hierarchy may provide the wrong generalizations, and so on. To avoid these problems, we developed our own representation system, essentially a system for building a classic semantic net.<sup>2</sup> Every minimal fact and item to which a text can refer is its own first-class object, as are the relationships among them. We refer to these objects as **noumena**,<sup>3</sup> and presently break them down into three basic types, reflecting differences in how they are mapped to the realization component.

objects, such as kitchen-window and sink  
relations, such as has-property and next-to  
properties, such as large and picture-window-like

Noumena have links to selected other noumena. These are the basis of the connectivity that (tacitly) makes a given knowledge base into a coherent whole, and allows the descriptive strategies to navigate it. Connections are introduced on an empirical basis wherever noumena are related in such a way that they can be combined by a strategy in some description as determined by our analysis of the corpus. The knowledge base for a given living space consists of all the noumena that might reasonably be mentioned, given our analysis.

While deliberate connections between noumena may seem to be begging the question, they account for a phenomenon that cannot be neglected, namely why it is that it never occurs to anyone to say, e.g., *the toilet is next to the stove*. If all aspects of describing a living space are data-directed, i.e. following or choosing among already established connections, then a speaker will never even think about infeasible possibilities. One can easily imagine other architectures, such as simply lumping all objects into a common heap organized by their salience, where one would have to actively search for interesting relations by methods like generate and test. Such a design would make different predictions

<sup>2</sup> At this point we do not include any sort of part-whole hierarchy such as house dominating room dominating furniture. If such a structure eventually emerges as a generalization that, say, simplifies the statement of our strategies, then this will suggest that it is inherent in the conceptualization of the task. On the other hand if we build in a hierarchy *a priori*, we will never know whether the structure is there only because we put it there.

<sup>3</sup> Singular: *noumenon*; a Greek word used by Kant to mean a thing-in-itself, independent of sensuous or intellectual perception of it.

about resource demands and processing effort than ours would.<sup>4</sup>

#### 4. Strategies and meta-strategies

A description is a controlled traversal of the knowledge base.<sup>5</sup> No component of this traversal is precomputed; that is, there are no "plans" which dictate *a priori* the structure of the description. Instead, control is handled by **strategies**, which are dynamically selected and linked together by **meta-strategies**. A strategy, when chosen, operates in a context which determines how it will traverse (its part of) the knowledge base. This context is composed of the most recently visited noun, all of the untraversed links emanating from it, and the most recently used strategy. There are other factors which feed into the context, many of which can be conceptualized as parameters which bias the choices of strategies within a particular house description. One such parameter is level of detail: a description may or may not include the more detailed descriptions of objects within it.

We believe that this implementation of the context is sufficient to account for most of the choices of strategies that could be made. However there are clearly cases in which a richer context is required, for example, *And the door, again, is in the same relationship to the windows as it is in Penni's room*. Here we need to model some awareness of previous patterns and the ability to refer to them in constructing new descriptions.

We have so far identified approximately 15 strategies in our corpus, each grouping nouns together and ordering their presentation. Typical strategies include moves such as **circular sweep**, in which the speaker picks an anchor point in a room, and describes the room's features or contents in an order determined by their placement along the circumference of the room, **look right - look left**, in which the speaker describes features to either side of a mental reference point, or **follow a hallway**, one of the strategies by which a speaker shifts to a new vantage point.

To understand this better, let us look at how the strategies and meta-strategies come into play in

<sup>4</sup> An arguably equivalent and perhaps preferable representation might be a non-propositional geometrical model after the fashion of an architect's drawing. However, we have never seen any evidence of the precision that such a representation would bring with it (quite the contrary), and have found many conventional aspects to the descriptions in our corpus that would be quite at odds with a model that captured the actual visual appearance of the house.

<sup>5</sup> Our observations agree with Linde's that a minimal description mentions all of the rooms (except possibly the bathroom) and their spatial relation to each other. APT keeps track of rooms (and other nouns) mentioned, and simply stops when all the rooms have been mentioned.

generating this excerpt from a description by a subject named Lisa. (This is an implemented example that APT has actually produced.)

*Then, in the kitchen,<sup>(1)</sup> there's a large window which faces the backyard, with two flanking windows.<sup>(2)</sup> And, if we're facing the backyard,<sup>(3)</sup> on the righthand side is a sliding glass door, and then a small window. If we're again facing the backyard,<sup>(4)</sup> on the lefthand side is the stove, then a refrigerator. And, beneath that large window is the sink,<sup>(5)</sup> and on the righthand side is the dishwasher.*

This segment starts with a preposed adverbial to mark a shift of vantage point.<sup>(1)</sup> Upon entering a major room a meta-strategy preferring any especially salient objects over object sequences applies, giving us the matrix clause of the first sentence.<sup>(2)</sup> That window is connected to three sets of objects, each of which is organized by a sweep strategy. This pattern (i.e. a salient object that is the nexus of several sweeps) triggers a room-sweep meta-strategy that anchors them all to the same object (the window), expressing the sweeps as displacements from this anchor using deictic terms (*righthand side, and then*) and reorienting to the salient focal point between sweeps.<sup>(3,4,5)</sup> A meta-strategy, probably specific to Lisa, prefers starting with "righthand" alternatives, thus giving the sweeps their order.

#### 5. Linguistic choices

Thus far we have been talking about issues of what we would call "orchestration": planning the text structure that provides the order of presentation, segmentation into sentences, and the textual function and salience of a body of information that has been selected for inclusion in an utterance. We must also look at issues in "mapping": selecting the specific wording and choice of construction that will realize a given noun.<sup>6</sup>

Lexical selection is in most respects a non-issue in living space descriptions. Nearly every physical object has an obvious and very over-learned name (e.g. *kitchen, bathroom, sink, refrigerator*), making the process one of simple retrieval rather than judgment and planning. The exceptions are, as one would predict, the objects whose associated common nouns do not pick them out uniquely, such as "hallway", "closet", or "window". For these APT will have to explicitly construct descriptions by folding in restrictive modifiers as they are needed. In the corpus, such descriptions were most often constructed

<sup>6</sup> The other principal activities of generation (as we see it) are "selection", which is in most respects trivial in this domain since we stipulate that all of the nouns in the knowledge base are to be mentioned, and "realization", which is carried out by the program Mumble-86 in the fashion described in Meteer et al. 1987.

from the same sort of spatial information used in clauses. Thus we have references to *a large hallway that leads into the kitchen, or the smaller hallway that leads to the bedrooms*. After it has been mentioned a few times, a description will be abbreviated and canonicalized: *that wide hallway, that smaller hallway*, with or without further (non-restrictive) modification.

Choosing syntactic constructions is a constrained problem in this task, since our corpus contains surprisingly few construction types. For example, once affective comments and digressions have been removed, more than half of all clauses fall within the class **locative relation**:

*there is* <object> <at location>  
<at location> *there is* <object>  
<at location> *is* <object>  
<object> *is* <at location>  
<object1> *has* <object2> *at* <location>

Which construction is selected is determined by a set of discourse-level heuristics. For example within a sweep the "<at location> *is* <object>" choice is natural because it facilitates chaining. Breaks between discourse segments can be flagged with a marked construction like There-Insertion (*Then there's Sabine's room on the right, as opposed to Sabine's room is on the right*).

## 6. Future Work

A major goal of this work is to construct a library of meta-strategies, strategies, and mapping rules that is comprehensive enough to allow APT to produce a variety

of new descriptions (in addition to mimicking the ones actually in the corpus) from the same knowledge base by varying library selections in what appear to be natural ways. Towards the end of establishing and strengthening our theory, we are also planning to collect and model a larger set of descriptions of another living space. In a parallel effort, we are working on a computationally feasible model of spatial deixis. We would ultimately like to use the architecture we develop to reanalyze prior work in related domains, such as the scene analysis done by Conklin's GENARO (1983).

## 7. References

- E. Conklin. Data-Driven Indelible Planning of Discourse Generation Using Saliency. COINS Technical Report 83-13, University of Massachusetts, 1983.
- C. Linde. The Linguistic Encoding of Spatial Information. Doctoral Dissertation, Columbia University, 1974.
- M. Meter, D. McDonald, S. Anderson, D. Forster, L. Gay, A. Huettner & P. Sibun. Mumble-86: Design and Implementation. COINS Technical Report 87-87, University of Massachusetts, September 1987.
- P. Sibun. APT: A System to Direct and Control Natural Language Generation. COINS Technical Report 87-42, University of Massachusetts, 1987.
- V. Ullmer-Ehrich. The Structure of Living Space Descriptions. In *Speech, Place, and Action*, R. Jarvella and W. Klein, ed., John Wiley & Sons, Ltd. 1982.