

From route descriptions to sketches: a model for a text-to-image translator

Lidia Fraczak

LIMSI-CNRS, bât. 508, BP 133
91403 Orsay cedex, France
fraczak@limsi.fr

Abstract

This paper deals with the automatic translation of route descriptions into graphic sketches. We discuss some general problems implied by such inter-mode transcription. We propose a model for an automatic text-to-image translator with a two-stage intermediate representation in which the linguistic representation of a route description precedes the creation of its conceptual representation.

1 Introduction

Computer text-image transcription has lately become a subject of interest, prompting research on relations between these two modes of representation and on possibilities of transition from one to the other. Different types of text and of images have been considered, for example: narrative text and motion pictures (Kahn, 1979; Abraham and Desclés, 1992), spatial descriptions and 3-dimensional sketches (Yamada et al., 1992; Arnold and Lebrun, 1992), 2-dimensional spatial scenes and linguistic descriptions (André et al., 1987), 2-dimensional image sequences and linguistic reports (André et al., 1988).

Linguistic and pictorial modes may be considered as complementary since they are capable of conveying different kinds of content (Arnold, 1990). This complementarity of expression is explored in order to be used in multi-modal systems for human-computer interaction such as computer assisted architectural conception (Arnold and Lebrun, 1992). Such systems should not only use different modes to ensure better communication, but should also be able to pass from one to the other. Given the differences in capacities of these two means of expression, one may expect some problems in trying to encode into a picture the information contained in a linguistic description.

The present research is concerned with route descriptions (RDs) and their translation into 2-dimensional graphic sketches. We deal with a type of discourse whose informational content may seem quite easy to represent in a graphic mode. In everyday communication situations, verbal RDs are often accompanied by sketches, thus participating in a 2-mode representation. A sketch can also function as a route representation by itself.

We will first outline some problems that may appear while translating descriptions into graphics. Then we will describe our general model for an automatic translator and some aspects of the underlying knowledge representation.

2 Some translation problems

Our first approach to translate RDs into graphic maps consisted in manually transcribing linguistic descriptions into sketches. By doing this, we encountered several problems, some of which we will try to illustrate through the following example, taken from the French corpus of (Gryl, 1992).

Example 2.1 *A la sortie des tourniquets du RER tu prends sur ta gauche. Il y a une magnifique descente à prendre. Puis tu tournes à droite, tu tombes sur une série de panneaux d'informations. Tu continues tout droit en longeant les terrains de tennis et tu tombes sur le bâtiment A.*¹

In the description here above we can observe some ambiguities, or incompleteness of information, which may be a problem for a graphic depiction. The most striking case is the information about the tennis courts: we do not know on which side of the path, right or left, they are located.

¹At the turnstiles of the RER station you turn left. There is a steep (a magnificent) downgrade to take. Then you turn right, you come across a series of sign posts. You continue straight on, passing alongside the tennis courts, and you come to building A.

There is also another kind of ambiguity due to the fact that in a RD the whole path does not have to be “linguistically covered”. Consider the fragment about turning to the left (“tu prends sur ta gauche”) and the downgrade (“descente”). It is difficult to judge whether the downgrade is located right after the turn, or “a little further”. The same question holds for the right turn (“puis tu tournes à droite”) and the sign posts (“panneaux d’informations”): should the posts be represented as immediately following the turning point (as expressed in the text) or should there be a path between them? This kind of ambiguity is not really perceived unless we want to derive a graphic representation of the route. The information is complete enough for a real life situation of finding one’s way.

Another kind of problem concerns the “magnifique descente”. It would not be easy to represent a slope in a simple sketch and, even less so, its characteristic of being steep, which the French word “magnifique” suggests in this context. The incompleteness of information will occur on the graphic side this time, not all properties of the described element being possible to express in this mode.

Such transcription constraints, once defined and analyzed, should be taken into account in order to obtain a “faithful” graphic representation. It seems that, in some cases, verbal-side incompleteness problems might be solved thanks to some relevant linguistic markers, as well as to the knowledge included in the conceptual model of the route. We think here in particular of the questions whether there is a significant stretch of path between two elements of environment (*landmarks*), or a turn and a landmark, mentioned in the text immediately one after the other. Concerning the ambiguity related to the location of landmarks, one can either choose an arbitrary value or try to find a way of preserving the ambiguity in the graphic mode itself.

We have mentioned here only some of the problems concerning the translation of RDs into graphic sketches. We have not considered those parts of linguistic description contents which are not representable by images, such as comments or evaluations (e.g. “you can’t miss it”; “it’s very simple”).

3 Steps of the translation process

Translating linguistic utterances into a pictorial code cannot be done without an intermediate representation, that is, a conceptual structure that bridges the gap between these two expression modes (Arnold, 1990). Abraham and Desclés (1992) talk about the necessity of creating a common semantics for the two modes.

In our case, the purpose of the intermediate representation is to extract from the linguistic description the information concerning the route with the aim of representing it in the form of a sketch. However, instead of trying to create a unique “super-structure”, we envisage a dual representation, with the linguistic and the conceptual levels. The core of the process of translating RDs into graphic maps will thus consist in the transition from the linguistic representation to the conceptual one.

For the sake of the linguistic representation, we thought it necessary to carry out an analysis of real examples and elaborate a linguistic model of this particular type of discourse. We have worked on a corpus of 60 route descriptions in French. The analysis has been performed at two levels: the global level and the local level. Global analysis consisted in dividing descriptions into global units, defined as *sequences* and *connections*, and in categorizing these units on a functional and thematic basis. We have thus specified several categories of route description sequences, the main ones being *action prescriptions* (e.g. “tu continues tout droit”) and *landmark indications* (e.g. “tu tombes sur le bâtiment A.”).² The inter-sequence connections (e.g. “puis”, “quand”, “ou”: “then”, “when”, “or”), which mark the relationships between sequences or groups of sequences, have been categorized according to their functions (e.g. *succession*, *anchorage*, *alternative*). Local analysis consisted in the determination of semantic sub-units of descriptions and in the definition of the content of different sequences with respect to these sub-units. These latter will enable, during the processing of a RD, to extract and represent information concerning *actions* and *landmarks*, and their attributes. Thus, one of the objectives of local analysis has been to determine which types of verbs in the RD express travel actions and which ones serve to introduce landmarks. The sub-units have been further analyzed and divided into types (e.g. different types of *actions*).

For the purpose of the conceptual representation of RDs, we need a prototypical model of their referent which is the *route*. We have decomposed it into a *path* and *landmarks*. A *path* is made up of *transfers* and *relays*. *Relays* are abstract points initiating *transfers* and may be “covered” by a *turn*. Landmarks can be either associated with relays or with transfers. More formally, a *route* is structured into a list of *segments*, each segment consisting of a *relay* and of a *transfer*. *Landmarks* are represented as possible attributes (among others) of these two ele-

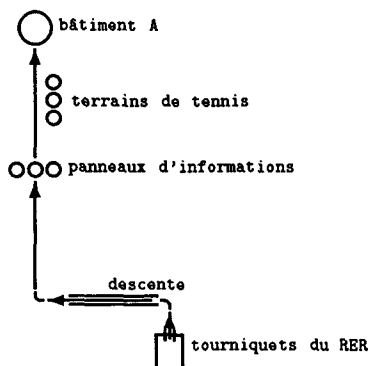
²Cf. Example 2.1

ments. Having such a prototype for routes, with all elements defined in terms of attribute-value pairs, it is relatively easy to re-construct the route described by the linguistic input: the reconstruction consists in recognizing the relevant elements and in assigning values to their attributes. Using the route model, some elements missing in the text can be inferred. For example, since every *route segment* contains one *relay* (which may be a turn) and one *transfer*, the information concerning the fragment of the route expressed by: “tournez à gauche et puis à droite” (“turn to the left and then to the right”), must be completed by adding a *transfer* between the two *turns*.

Apart from models for linguistic and conceptual representations, the rules of transition have to be defined. For this purpose, it is necessary to establish relationships between different linguistic and conceptual entities. For example, the action of the type “progression” (e.g. “continuer”, “aller”) corresponds to a *transfer* and the actions of the type “change of direction” (e.g. “tourner”) or “taking a way” (e.g. “prendre la rue”) to a *relay* (which will coincide with a *turn* or with the beginning of a *way-landmark*, e.g. a street, respectively).

Another aspect of modeling consists in specifying graphic objects corresponding to the entities in the route model. For the time being, we decided to do with simple symbolic elements, without a fine distinction between landmarks. The graphic symbols have been created on the basis of the information accessible from the context rather than the one contained in the “names” of landmarks. These latter are included in sketches in the form of verbal labels.

Once the whole route has been reconstructed at the conceptual level, we start to generate the corresponding graphic map, like the one here below.



4 Conclusion

Computer translation of route descriptions into sketches raises some interesting issues. Firstly, one

has to investigate the relationships between the linguistic and the graphic modes, the constraints and possibilities which appear while generating images from linguistic descriptions.

Secondly, a thorough linguistic analysis of route descriptions is necessary. We have used a discourse based approach and analyze “local” linguistic elements by filtering them through the discourse structure, described at the “global” level. Our goal is to build a linguistic model for the text type “route description”.

Another interesting problem is the form and the derivation of the conceptual representation of the described route. We believe that it cannot be directly obtained from the linguistic material itself. During the understanding process, the linguistic meaning has to be represented before the conceptual representation can be created. That is why we need a two-stage internal representation, based on specific linguistic and conceptual models.

References

- M. Abraham and J-P. Desclés. 1992. Interaction between lexicon and image: Linguistic specifications of animation. In *Proc. of COLING-92*, pages 1043–1047, Nantes.
- E. André, G. Bosch, G. Herzog, and T. Rist. 1987. Coping with the intrinsic and the deictic uses of spatial prepositions. In K. Jorrand and L. Sgurev, editors, *Artificial Intelligence II: Methodology, Systems, Applications*, pages 375–382. North-Holland, Amsterdam.
- E. André, G. Herzog, and T. Rist. 1988. On the simultaneous interpretation of real world image sequences and their natural language description: The system SOCCER. In *Proc. of the 8th ECAI*, pages 449–454, Munich.
- M. Arnold and C. Lebrun. 1992. Utilisation d’une langue pour la création de scènes architecturales en image de synthèse. Expérience et réflexions. *Intellectica*, 3(15):151–186.
- M. Arnold. 1990. Transcription automatique verbal-image et vice versa. Contribution à une revue de la question. In *Proc. of EuropIA-90*, pages 30–37, Paris.
- A. Gryl. 1992. Opérations cognitives mises en oeuvre dans la description d’itinéraires. Mémoire de DEA, Université Paris 11, France.
- K.M. Kahn. 1979. Creation of computer animation from story descriptions. A.I. Technical report 540, M.I.T. Artificial Intelligence Laboratory, Cambridge, MA.
- A. Yamada, T. Yamamoto, H. Ikeda, T. Nishida, and S. Doshita. 1992. Reconstructing spatial image from natural language texts. In *Proc. of COLING-92*, pages 1279–1283, Nantes.